Improving Public Engagement with Ethical Complexities of Assistive Robots

Catherine Menon[1][0000-0003-2072-5845] Silvio Carta[2][0000-0002-7586-3121] Frank Foerster[3][0000-0002-9263-3897] Patrick Holthaus[4][0000-0001-8450-9362]

[1,3, 4] Department of Computer Science, University of Hertfordshire, Hatfield, UK
[2] School of Creative Arts, University of Hertfordshire, Hatfield, UK
{c.menon,s.carta, f.foerster, p.holthaus}@herts.ac.uk

Abstract. In order for society to fully realise the potential benefits offered by assistive robots, a number of ethical challenges must firstly be addressed. Crucially, it is important to enhance public understanding of the ways in which societal ethics can be used to formulate and guide the preferred behaviours of these robots, particularly in scenarios which are ethically complex. Furthermore, it is also important to ensure that the voices of end users are heard, and their input used in the development process. In this paper we present EETAS, a methodology for using structured workshops to improve public understanding of assistive robot ethical complexities. We also present DEETAS, a further digital-based and design-centric methodology for engaging potential end-users who may be reluctant to take part in collaborative workshops. In support of these two methodologies we present the findings of an initial pilot study and usability study, showing an indicative trend that these processes are effective in engaging users and enhancing public understanding of the ethical complexities inherent in assistive robots.

Keywords: Assistive robots, ethics, design.

1 Introduction

The ethical concerns around use of domestic and assistive robots are some of the most challenging obstacles to their introduction into users' homes. Not only should there be an open and inclusive conversation about whether such systems are ethically acceptable in themselves, but also around how these robots should manifest the social and ethical norms of their surrounding societies. Specifically, there is little guidance for either developers or the general public on how to identify, prioritise, understand and balance different ethical principles which inform the behaviours of assistive robots.

In part, this is due to the fact that the ethical principles considered desirable for AI and AI-enabled systems differ across different societies, cultures and demographics. In the UK, for example, discourse on AI ethics stresses the importance of human autonomy and individual decision making (BSI, 2023; Leslie, 2019) and is contextualised within a relatively "light touch" regulatory regime. (DSIT, 2023). In China, by contrast, AI and ethics discourse is anchored by a preference for strong-binding regulation, and the need to consider the interests of private stakeholders (Zhu, 2022; Arcesati, 2021). Similarly, ethical concerns about AI technologies are not equivalent across all demographics, with older demographics in the UK notably more concerned about data privacy and the provision of detailed technical information about how data will be shared (CDEI, 2022).

This is particularly relevant when we consider the characteristics of "early adopters" (TTI, 2018), being those who are most likely to embrace new technology such as domestic and assistive robots. This group tend to share some general characteristics: being more likely to be young, male, highly-educated and, given the cost of new technology, wealthy (Rogers, 2003; Hardman et al., 2019).

The differences between early adopters and the majority of society present a new set of ethical concerns when it comes to integration of new technology such as assistive robots. Early adopters are in the minority, comprising 11% of consumers in the UK (YouGov, 2020). As well as being amongst the first to purchase new technologies, they are also more likely to volunteer to be involved with trials, surveys and consultations around such technologies. For example, in a recent survey exploring the acceptance of social robots (Saari et. al., 2022), respondents without university degrees made up only 36% of the sample, while only 5% of respondents were over 55 years old. This is particularly problematic given that assistive robots are typically designed for a target demographic that does not match the "early adopters". Rather, end-users of assistive robots are likely to be vulnerable, frail or elderly, characteristics which can lead to marginalisation by technology and further embedded distrust.

As described above, end-users of assistive robots are likely to have different requirements for ethical behaviour of these robots – and different concerns around their use – to early adopters. In particular, while such different segments of the population may potentially agree on what the desirable ethical principles for assistive robots are (e.g., privacy, obedience, safety etc.), they are unlikely to agree on how to prioritise some of these principles over others in any given scenario, particularly where the ethical principles may be in opposition to each other. Scenarios like this arise relatively commonly with assistive robot technology, and include situations where the user asks the robot for assistance with a task that may cause harm to the user, such as assistance in obtaining an alcoholic drink. Here the robot is ethically compelled both to obey, and to prevent harm from arising to the user via consumption of the drink. Similar scenarios arise where the user expects privacy in a location where s/he also requires physical assistance from the robot (e.g. in a slippery kitchen or bathroom), or where the user requests the robot to regularly perform actions such as lifting or fetching items, which in the long-term can reduce the user's physical ability to perform such actions themselves. (Amirabdollahian et al., 2013; Menon et al., 2022). Given the ethical sensitivities around such scenarios, it is therefore important to design tools, methodologies and survey techniques which encourage participation by end-users themselves, rather than disproportionately representing early adopters.

In this paper we address this gap by presenting our conclusions from two studies. The first of these studies – initially described in (Menon et al., 2022) – is a workshop designed to validate the EETAS (Ethical Trade-offs in Autonomous Systems) process, an in-person and team-based gamified methodology to elicit stakeholder opinions on the ethically-informed behaviour of assistive robots. EETAS focuses specifically on those scenarios where there are multiple competing desirable ethical properties which could guide the behaviour of an assistive robot, and asks end-users to work together in a collaborative environment using a tangible design artefact similar to a Jenga toy. This design artefact, EETAS-TOY (Ethical Trade-offs in Autonomous Systems Trade-offs-For-You), allows users to physically represent and negotiate different ways to prioritise a number of ethical principle, and to come to a consensus for the team's preferred ethical balances.

We hypothesise that participation in the EETAS process will increase participants' understanding of the complexities of assistive robot ethics. Furthermore, we anticipate that such participation will allow stakeholder opinions and preferences to be communicated to developers at a sufficiently early point in the lifecycle to influence the design of the robot.

The second study presented in this paper extends the EETAS process to those who cannot – or are reluctant to – engage with an in-person and team-based workshop focused specifically on assistive robots. This study presents a usability assessment of a digital interactive process (DEETAS Digital EETAS) which aims to achieve the same outcomes as the EETAS process, without the need for a team-based, in-person and gamified environment. The digital artefact used for this, DEETAS-TOY, can be used individually, and draws on a combination of short videos and a simple, non-technical interface to allow users to select their preferred ethically-informed behaviour for an assistive robot in different scenarios. We hypothesise that the DEETAS-TOY tool will allow increased participation by those who are averse to new technologies or otherwise unwilling or unable to participate in typical trials and consultations for these technologies, including the elderly, frail and vulnerable. As such, it represents an important step in ensuring that end-users are empowered and their voices heard in conversations around acceptable ethics of assistive robots.

This paper is an extended version of (Menon et al, 2022), which described the EETAS methodology and pilot study workshop results only. This work builds on the results presented in (Menon et al., 2022) by extending the EETAS methodology described therein to those end-users who would be otherwise marginalised by the restriction of EETAS to physical and in-person workshops: a demographic which includes significant overlap with the potential end-users of the robots themselves. We therefore present augmented motivations for this research, new methodologies and results, and conclusions which extend the initial hypotheses made in our prior publication.

Section 2 presents a discussion of existing literature around ethics of assistive robots to further illuminate the current research landscape and clarify the research gap. Section 3 reviews the EETAS process initially described in (Menon et al., 2022), as the foundation for our novel work on the DEETAS process, along with the DEETAS-TOY design and features described in Section 4. Section 5 discusses both the original outcomes of the EETAS pilot study workshop and the results of a usability study on DEETAS-TOY, reflecting on how these results inform and support each other. Finally, Section 6 provides conclusions and suggests further work.

2 Existing Literature

Analysis of the interaction between desirable but incompatible ethical principles is not unique to assistive robots, nor even to autonomous systems. This is well-explored in the fields of psychology, economics and mathematics, particularly in terms of identifying the behaviour which allows an individual to maximise the expected value of any given decision (von Neumann, 1947). The ethical dilemma most notably associated with autonomous system behaviour is the trolley problem (Foot, 1967), which in its initial form asks participants to choose between two actions which each have the consequences of a trolley (tram) colliding with a different number of people. While a very simplistic treatment of ethical trade-offs, this dilemma has nonetheless been central to much mainstream discussion of autonomous system behaviour, particularly around autonomous vehicles (Frankel, 2021; Fox, 2018). Similarly, popular fiction (Asimov, 1942) has motivated much public discussion around how robots should prioritise certain behaviours such as obedience and safety, where these come into competition.

Fiction and philosophy are invaluable in opening up a conversation about assistive robot ethics, particularly where this conversation can involve a human-centred assessment of preferences rather than a technically-oriented discussion of capabilities. Nevertheless, such literature does not necessarily provide sufficient nuance, and risks presenting too simplistic a perspective for end-users to make informed risk-acceptance decisions (Roff, 2018).

Technically focused publications focusing on specific digital systems address this by presenting a more nuanced and complex perspective on trade-offs. For example, (IET, 2019) discusses how safety and security might be balanced in a cyber-physical system, while (Akinsanmi, 2021) focuses on the trade-offs made between public health, privacy and digital security during the Covid-19 pandemic. (Thornton, 2018) specifically addresses the balance between maintaining personal autonomy and public health with autonomous vehicles, while (Lin, 2015) and (Menon & Alexander, 2019) discuss this from the perspective of risk acceptance and prioritisation. However, although these works are technically sound, they are not necessarily presented in a format which is accessible to end-users of these systems, who may not have a technical background or be comfortable with specific technical or philosophical vocabulary.

With this in mind, existing work has assessed the effectiveness of using games to engage people with ethical and user design principles. Gamification has been shown to improve stakeholder engagement when discussing system requirements (Lombriser et al., 2016), as well as specifically to help end users and developers identify ethical questions and concerns which arise out of using complex systems (Ballard et al., 2019; Malizia et al., 2022).

2.1 Previous EETAS work

Previous work (Menon et. al., 2022) drew on this body of literature to establish a structured process for exploring ethical trade-offs in autonomous systems, with the specific example of an assistive robot. This research made use of a physical interactive tool to facilitate team-based conversations around participants' preferred ethically-driven behaviours for assistive robots in a number of pre-determined scenarios.

We revisit the specifics of this work in more detail in Section 3, but note that the use of a physical tool and associated workshop draws on conclusions from (Schrier, 2019) and (Larson 2020) around the benefits of gamification. Such benefits include opportunities for social interaction, communication and increased moral knowledge. Nonetheless, in the work described in (Menon et al, 2022) these benefits are available only to those participants who are both able and willing to physically participate in the workshop and gaming process.

We now seek to address this by extending this work to consider the role that digital tools can play in increasing accessibility. (Parsons et al, 2019) identifies the advantages of learning through digital tools and spaces, and contrasts this with the role of the physical place. This aligns with the findings in (Roberts, 2006) that use of digital instead of physical surveys and workshops enables access to specialized and hidden populations, such as those who are vulnerable, elderly or otherwise marginalised within society.

The work we describe here constitutes a usability study to assess the acceptability of the DEETAS-TOY tool, a digital equivalent of the EETAS-TOY tool, to be used by those who would otherwise be excluded from the EETAS process. Usability studies are an established method of assessing the effectiveness and satisfaction with which users can achieve their goals within a given environment (ISO, 2010). Measures of usability for general digital tools typically include proxies such as ease of use, time, accuracy and understanding (Hornbaek, 2006), while those used to assess design usability include feedback, understanding and consistency (Schneiderman, 2009). Our design questionnaire and usability study, described in Section 3, incorporates these elements of assessment.

3 EETAS Process

In this section we review the EETAS process as initially presented in (Menon et al, 2022). This process is to be used for eliciting user opinions on the ethically-informed behaviour which should be prioritised by an assistive robot in those scenarios where multiple desirable behaviours conflict.

The EETAS process – like the DEETAS-TOY described in Section 4 – is intended to be used at an early stage of the design lifecycle of an assistive robot. This enables developers to integrate the feedback into the design of the system, and to ensure that the preferences and ethical positions elicited during the workshop are reflected in the robot's behaviour. It is important to note that this means the full specification of the assistive robot may not be known at the time of the EETAS process; this is accommodated by the process, and any under-specifications can be later amended. Throughout this section we will assume the presence of an EETAS facilitator: a person committed to implementing, managing and facilitating the EETAS workshop, and to obtaining all necessary information prior to this.

The EETAS process consists of two preliminary steps followed by a team-based workshop. The first preliminary step is to identify or create an adequate functional description of the robot in question, and the second is to identify the desirable ethical principles for this robot to follow.

3.1 EETAS Preliminary steps

The developers must provide to the EETAS facilitator – prior to the EETAS workshop commencing – a written description of the functionality of the robot in question. This should be an easily-read and accessible document that does not require any technical knowledge to understand: a mathematical or technical specification is not required at this point. An example description is shown in **Fig. 1**

Participants in the workshop are likely to have further questions which emerge throughout the process, and this is to be expected given the relatively early stage of the design lifecycle. Emerging questions will become the foundations for the specific behavioural scenarios generated during the workshop.

Functionality: this is an assistive robot that reminds you when to take medication, alerts you when you have left the oven on and engages you in conversation.

Interface: the assistive robot will speak to you and can understand your speech in return.

Customisation: you can request that the robot should not engage you on particular topics of conversation, or should limit its interactions to certain hours.

Networking: the robot is connected to the internet, and can additionally access all data from home smart systems, such as Amazon's Alexa.

Physical Appearance: a diagram will be provided to represent the robot's physical appearance.

Fig. 1. A sample description of an assistive robot for EETAS preliminary steps

The second preliminary step is for the developer and EETAS facilitator to work together to identify which ethical principles are relevant to the behaviour and functioning of this particular robot. The EETAS workshop provides participants with the option to augment these during their team discussion with additional principles, but a preliminary set identified via literature review from standards, guidance and academic papers ((BSI, 2023, IEEE, 2018, (Leslie, 2019, National Cyber Security Centre, 2019) includes:

- System promotes human physical safety
- System obeys human commands
- System promotes affinity with human user
- System maintains data privacy
- System is accurate
- System is fair
- System maintains human autonomy
- System promotes human long-term health
- System does not attempt to deceive

3.2 EETAS Workshop

Participants in the workshop should include potential end-users and members of the public, without any specific requirements placed on technical knowledge. Participants will be grouped into teams of 4 – 10 people, in order to provide a number of diverse opinions and perspectives while at the same time ensuring that dialogue and communication remains effective (Curral, 2001), and care should be taken with the selection of the teams to avoid introduction either segregation or bias.

Each team will be provided by the EETAS facilitator with the outputs from the preliminary steps: the functional description of the assistive robot and the identified relevant ethical principles. Teams should be encouraged to augment or refine the set of ethical principles should they wish, using informal techniques such as brainstorming and if-then methodologies. It is advisable, although not essential, for all teams to work with the same set of ethical principles.

The workshop consists of two phases, to allow team collaboration to develop with increased interaction. The goal of the first phase is to introduce teams to the assistive robot and to the idea of competing ethical principles. This phase is not gamified, to allow for greater team cohesion. The second phase – which may be preceded by a short workshop break – aims to facilitate greater communication by bringing teams together and introducing team members to novel perspectives from other teams. The goal of the second phase is to agree – in a gamified and hence somewhat competitive environment – on ethical prioritisations which meet the requirements of all members of two, or more, teams.

3.3 Workshop Phase 1

In the first phase of the workshop teams work together to create scenarios aligned with the functional description of the assistive robot and where at least two of the ethical properties come into conflict.

Teams may use any methods they prefer to create these scenarios. However, we suggest two methods here, a ‘light-touch’ top-down method for those who prefer to work more creatively, and a structured bottom-up assessment process for teams preferring to work within this framework. Different teams may use different methods, and a team may switch between methods throughout this phase.

Light touch top-down method. The light touch top-down method consists of three guiding questions, which are intended to encourage creative and collaborative discussion.

- Who might use this functionality?
- Is there anyone who might be negatively affected by this functionality?
- Are there any circumstances or times when the user might prefer that this functionality is not provided?

When discussing these questions, teams are encouraged to consider the environment in which the assistive robot will be used, and the data – e.g., from other home smart systems – that will be available to it.

Structured bottom-up method. For teams preferring a structured bottom-up methodology, we provide a pre-generated checklist of guidewords to help identify which ethical properties might be in conflict with each other, and generate scenarios from that. Each of these guide words is to be applied in turn to each of the relevant ethical properties. The guidewords follow the collaborative principles of Hazard and Operability Analysis (HAZOP) studies (BSI, 2016).

TOO MUCH: the robot performs its functions in such a way that this ethical property is implemented for too many people / in too many circumstances / to too much of an extent

NOT ENOUGH: the robot performs its functions in such a way that this ethical property is implemented for too few people / in too few circumstances / to not enough of an extent

UNIFORMLY: the robot performs its functions in such a way that the outcome is applied uniformly to everybody, or in exactly the same way to everybody, when differentiation is needed

INCONSISTENTLY: the robot performs its functions inconsistently / differently for different people / differently each time when consistency is needed

UNEQUALLY: the robot performs its functions only for some people or only in some circumstances

For teams using the structured bottom-up method, we recommend that each potential scenario is double-checked by the EETAS facilitator to ensure that it represents a conflict of ethical principles, rather than simply undesirable behaviour.

Scenario: "the user asks their assistive robot not to remind them about medication today, because they don't want to take it".

Ethical principles in conflict: "system promotes human long-term health" and "system obeys human commands"

Scenario: "the robot confides a fictional 'secret' to the user during conversation to engage the user's interest and promote friendship"

Ethical principles in conflict: "engender trust in the human user" and "system does not attempt to deceive".

Fig. 2. Two sample scenarios generated during EETAS Phase 1

3.4 Workshop Phase 2

In the second phase of the workshop, different teams come together in pairs and swap scenarios with each other. Each team then assesses their partner-team's scenarios, working collaboratively within the team to answer the following questions:

1. Do you collectively consider this a valid scenario given the functionality of the robot?
2. What is your preferred behaviour for the robot in this scenario, considering the ethical principles at stake?
3. Are there any design, environmental or functional constraints which might lead you to accept an alternative behaviour to your preferred one, considering the ethical principles at stake?

The first of these questions relates to scope: given that the assistive robot in question is at an early stage of development, teams may not agree on exactly what functionality is in scope or can be expected. Scenarios upon which teams and their partner-teams do not agree can be discarded.

The second of these questions identifies which ethical principle the team wishes to prioritise in this scenario, while the third encourages collaborative discussion about the circumstances under which the team may be willing to accept an alternate preference for ethical prioritisation.

We suggest teams consider the following discussion points in answering questions 2 and 3.

Discussion point 1: Would you accept any alternate behaviour in this scenario if users were told beforehand that this is how the robot operates?

Discussion point 2: Do you think the person benefitting from the different behaviours in this scenario has the moral right to benefit in this way?

Discussion point 3: Could some of the different behaviours described in this scenario be acceptable in a different environment? With different users? If these did not impact the same people?

3.5 Gamification of Phase 2

Phase 2 may be gamified throughout to encourage greater collaboration and participation, as well as to explicitly engage participants with the process. This may be done by allocating points to the Phase 2 questions as follows:

- One point to a team for each scenario deemed valid by their partner-team
- Two points to a team for agreeing within the team on a preferred behaviour for a valid scenario
- Three further points to a team for identifying design, environmental or functional constraints which would lead them to accept an alternate behaviour within that valid scenario. These points should only be awarded where the partner-team considers these constraints feasible to implement

In order to maximise the gamification, teams should be responsible for awarding points to their partner-team. Teams may then be ranked by the EETAS facilitator as pairs (i.e. partner-teams) or as individual teams.

3.6 Design Tool in Phase 1 and 2

During both phases, teams are encouraged to use a pre-prepared design tool, EETAS-Trade-Offs-for-You (EETAS-TOY). The EETAS-TOY tool, shown in **Fig. 3** is a physical design artefact representing the ethical principles relevant to the assistive robot as a set of sliding bars embedded within a block representing the system. A given bar represents two ethical principles in competition with each other, and participants should be encouraged to label the bars with removable labels or an erasable pen.

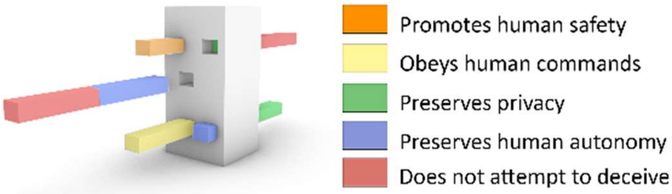

Fig. 3. The EETAS-TOY gamified tool

Participants should use the design tool to discuss, negotiate and explore different ethical trade-offs between given principles. The use of a design tool decouples the discussion from the technical capabilities of the robot and allows a tactile and more playful experience consistent with the gamified design of the process.

Teams should record the discussion and outcomes from Phase 2, which may be either digital recordings or using written methodologies including mind-mapping (Beel, 2011). These records constitute feedback which can either be provided by the EETAS facilitator as-is to the developers of the assistive robot in question, or amalgamated to represent a high-level view of stakeholder opinions on ethical trade-offs in assistive systems. The EETAS-TOY itself can be used by either the teams or the EETAS facilitator as a visual or artistic representation of the ethical trade-offs decided upon by each team, as well as an explanatory tool to provoke further conversation with additional stakeholders. The tool is also valuable as a public record of the conversation, and as a means of moving ethical discourse beyond the technical into the human-centred and artistic domains.

4 DEETAS Process

As discussed earlier, not all potential users of assistive robots will be willing or able to engage in physical, in-person workshops. In this section we therefore describe an expansion to the EETAS methodology. This is DEETAS, a further digital-only process that uses the DEETAS-TOY (Digital EETAS-TOY) to achieve the same outcome: a visual representation of stakeholder opinions about preferred ethical trade-offs to be made by an assistive robot. This process is individual and non-gamified, to allow for enhanced accessibility and encourage input from those traditionally marginalised in such discussions. As with the EETAS process, DEETAS requires a facilitator to curate the initial information and administer the process.

4.1 DEETAS Preliminary Steps

As with the EETAS process of Section 3, it is important for the developers to initially provide information about the functionality of the robot. This is given only to the DEETAS facilitator – in contrast to EETAS, this functional description is not provided directly to participants - so may be in any form that the process facilitators require: in some circumstances this may be considerably more detailed or technically complex than the EETAS functional descriptions of **Fig. 1**. This will be used by the DEETAS facilitator, together with the developers, to create video scenarios which represent competing ethical principles.

As with the EETAS process, it is necessary to identify the ethical principles which are relevant to the assistive robot as described. Where DEETAS and EETAS are being used together, the set of ethical principles should be re-used from the EETAS process in the interests of obtaining comparable feedback from all participants.

4.2 DEETAS Video Scenarios

The DEETAS facilitator and developers should work together to create video footage which represents scenarios in which identified ethical properties can conflict. Where possible these should be selected from the EETAS workshop scenarios, with the facilitator applying academic judgement as to which of these can be best represented on video.

If the assistive robot in question has not advanced sufficiently in the design lifecycle to be represented on video, then an alternative robot or visually equivalent stand-in should be used. It is important to emphasise that the video footage is representational, not literal, and is intended to be reflective of hypothetical scenarios which may emerge during the behaviour of a proposed assistive robot. The video scenarios should be made available to users via upload to an accessible website.

4.3 DEETAS Design Tool

The DEETAS-TOY design tool is a digital design artefact, intended to be visually as close as possible to the physical EETAS-TOY artefact, and likewise representing the ethical principles relevant to the assistive robot. As with the EETAS-TOY tool, the ethical principles are represented by sliding bars which are embedded in an opaque block.

A different DEETAS-TOY configuration should be used for each video scenario, as not all ethical principles are relevant in all scenarios. This removes the element of choice from the participant – both in terms of generating the scenarios and identifying the relevant ethical principles – but also provides a lessened obstacle to entry for those participants who prefer to work individually instead of as part of a team.

The DEETAS-TOY artefact, an example of which is shown in **Fig. 4** can be constructed in any standard platform for 3D data, including Speckle (Speckle, 2023), Viktor (Viktor, 2023) or ShapeDiver (ShapeDiver, 2023). Consideration should be given to questions of accessibility, including the methods of interaction (e.g. keyboard vs mouse) as well as visual accessibility, choice of colours, ease of use with multiple browsers and ability to be accessed from a variety of devices.

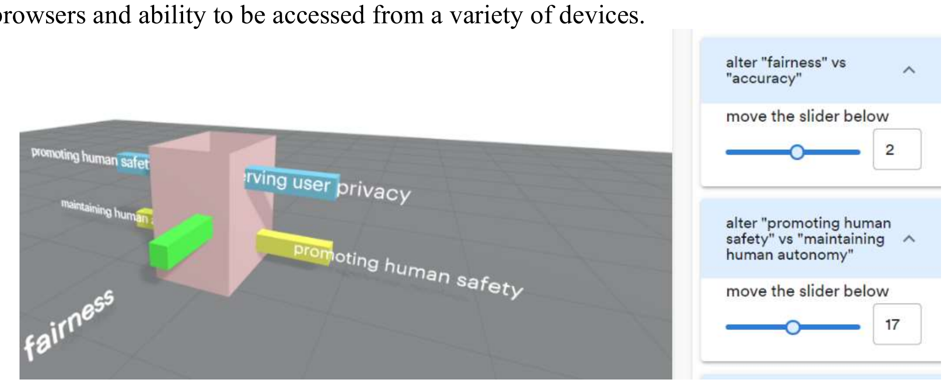

Fig. 4. The DEETAS-TOY non-gamified tool

Information should be collected digitally from DEETAS participants as to their preferred ethical prioritisations and any additional feedback. This may be done via a saved representation of the DEETAS-TOY tool itself – as with EETAS, this represents a valuable visual or artistic artefact of the user's thought process and eventual conclusion – or may be via traditional methods such as a survey or follow-up email.

5 Validation and Results

We performed experiments to partially validate both the EETAS and DEETAS processes: for EETAS this was in the form of a pilot study workshop, while for DEETAS a usability study was conducted on the DEETAS-TOY tool. The methodology and results of each of these follow.

5.1 EETAS Validation

The EETAS partial validation consisted of a pilot study workshop intended to investigate how participants perceived the EETAS process, and whether they considered it to enhance their understanding of ethical trade-offs and complexities for assistive robots. Furthermore, this workshop was intended to investigate how participants felt about the EETAS-TOY tool, from the perspective of its usefulness in both enhancing their understanding of ethical complexities, and enabling them to communicate, discuss and negotiate preferences for these within their teams.

We note that this is only a partial validation, intended to establish indicative correlation between participating in the EETAS process and an improved understanding of ethics in assistive robots. Further planned validation is discussed in Section 6.

The experiment was approved by the University of Hertfordshire's Health, Science, Engineering and Technology Ethics Committee under protocol number SPECS/SF/UH04940.

Workshop design. The workshop was held at the University of Hertfordshire, with participants who had volunteered to take part. Participants were asked for consent, and were then divided into teams of 4 – 5 by the researchers, with consideration being given to the team selection to avoid bias on the grounds of age, gender, race and prior experience. Participants who knew each other were also placed into different teams, to closer mimic the experience of running such workshops with the general public.

Teams were provided with a high-level written description of the functionality of a proposed hypothetical assistive robot. This functionality, although hypothetical, was drawn from real-world case studies conducted at the University of Hertfordshire Robot House (Holthaus et al., 2019; Menon & Holthaus, 2019; Koay et al., 2020; Saunderset al., 2016).

The description included physical characteristics of the assistive robot, as well as details on the functions they could expect it to perform. A partial excerpt from the functional description is provided in **Fig. 5**.

Functionality: For this workshop you can assume that the assistive robot is capable of:

- Moving about the house and entering all of the rooms
- Reminding the user to take any necessary medication on time
- Engaging the user in social interaction and conversation

Interface: The robot can obey simple verbal or gestural commands from the user provided these instructions are within the scope of the functionality specified here

Networking: The robot can obtain information from other sensors or devices in the home, including camera, audio devices and the user's PC. It can turn on the tv or computer, and recommend programmes. It can also autonomously communicate pre-determined alerts or warnings relating to the user and their health or activities, sending this via email to the user's doctor and / or family members

Physical appearance: a photograph of one of the UH Robot House robots was used to represent the robot's physical appearance

Fig. 5. Excerpt from EETAS workshop robot description

Participants were provided with printed cards representing eight of the ethical principles identified in Section 3.1 ("system does not attempt to deceive" was omitted due to its complexity). Given that this was a pilot study only, teams were asked to work with these principles only, and were not encouraged to augment these with their own further principles. Each team was also provided with an EETAS-TOY tool and shown how to use this, and how ethical trade-offs could be represented via manipulation of the rods.

Participants were asked to fill in an initial questionnaire, providing information on whether they had any background relating to either design or robotics as well as identifying how well they felt they currently understood ethical trade-offs and complexities in assistive robots.

Participants were then asked to embark on Phase 1 of the workshop, identification of scenarios in which these ethical principles might be in conflict. Owing to the limited nature of this pilot study, participants were encouraged to use the *light-touch top-down approach* instead of the more *structured bottom-up approach* (see Section 3.1 for details). Teams were also asked to record their scenarios in written form, in preparation for Phase 2.

All teams were working simultaneously, and were monitored by a dedicated member of the research team. The researchers took detailed timing observations of the discussion and the use each team made of the EETAS-TOY tool. In addition, the researchers were able to answer questions and remind participants of the purpose of the discussions, but did not contribute beyond this.

Teams were then invited to complete Phase 2 of the workshop and were placed in partner-team pairs by the researchers. Teams interacted to identify valid scenarios from their partner-team, confirm their own ethically-guided behaviour preferences for each of the scenarios, and identify constraints which would make alternative behaviours possible. Teams were allocated points by their partner-teams, with the researcher acting as facilitator and guide for this part of the process.

Immediately after the workshop, all participants were asked to complete a post-study questionnaire. This questionnaire asked a number of questions about the participants' perception of the utility of each part of the EETAS process, as well as the utility of the identified ethical principles and the role which they felt the EETAS-TOY tool played in their discussions. Specific questions included:

- Participants were asked to give a numerical score of how well they understood ethical prioritisations and complexities in assistive robots before the EETAS process, and how well they understood these trade-offs following EETAS (0 = not at all, to 5 = very well)
- Participants were asked to give a numerical score of how helpful they found the EETAS process in understanding ethical prioritisations and complexities in assistive robots (0 = unhelpful, to 5 = very helpful)
- Participants were asked to give a numerical score of the EETAS-TOY tool in a) understanding and b) communicating about ethical prioritisations and complexities in assistive robots (0 = unhelpful, to 5 = very helpful)

5.2 EETAS Validation Results

This was a pilot study only, with a relatively small sample size (< 20 participants), and hence no statistical significance between conditions and questionnaire responses is expected. Nevertheless, we have identified some indicative trends which support our hypothesis that participating in the EETAS process increases participants' understanding of the ethical complexities of assistive robots.

Participant Demographics. Most participants identified as having either a background in design (93%) or in robotics (43%). This tendency towards expertise in at least one of these areas was due to the constraints around recruitment and selection for this pilot workshop: most participants were sourced via existing connections to the University of Hertfordshire and to specific members of the research team. The age range of participants was 19 – 61 years old, with the average age being 37. The gender balance was roughly equal: 57% male and 43% female.

Participant Responses to EETAS. Before the workshop, participants lacking a background in robotics rated themselves as having a relatively low understanding of the different ethical prioritisations an assistive robot might make (mean value 2.6 from a Likert scale where 0 = no understanding and 5 = full understanding). Those with a background in robotics, by contrast, considered themselves as having a reasonable understanding of ethical prioritisations and complexities associated with assistive robots (mean value 3.4 from the same Likert scale). Post-workshop, both those with a robotics background and those without a robotics background considered that they understood ethical prioritisations more completely. Furthermore, the gap had narrowed (mean value of 3.8 for those without a robotics background, vs 4.2 for those with, using the same Likert scale where 0 = no understanding and 5 = full understanding). This corresponds to an increase of 58% greater improvement in understanding ethical prioritisations and complexities for participants without a robotics background, as compared to those with.

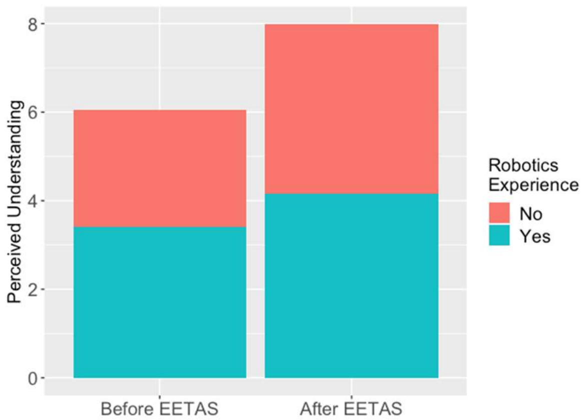

Fig. 6. Change in perceived understanding of ethical complexities

A second question asked participants to identify how helpful the EETAS process itself was in understanding ethical prioritisations and complexities (from a Likert-type scale where 0 = unhelpful, to 5 = very helpful). 94% of all participants rated the helpfulness of the EETAS process at 3 or above on this scale, with the mean ranking being 3.7. There was no difference between the ratings given by those participants with robotics backgrounds and those without.

A third question asked participants to rate the helpfulness of the EETAS-TOY tool itself in a) understanding and b) communicating about ethical prioritisations and complexities in assistive robots, using a Likert-type scale where 0 = unhelpful to 5 = very helpful. For a) – understanding – 64% of participants rated the helpfulness of the EETAS-TOY as 3 or above (mean value 3.3), while for b) – communicating – 71% of participants rated the helpfulness of the EETAS-TOY as 3 or above (mean value 3.7). In contrast to the second question – asking participants to rate the helpfulness of the EETAS process, which gave rise to results that were independent of background –

participants without a robotics background considered the tool more helpful than those with. Specifically, participants with a robotics background rated the tool's helpfulness for understanding ethical prioritisations at a mean value of 3.1, and for communicating these prioritisations at a mean value of 3.8. Participants without a robotics background rated the tool's helpfulness for understanding ethical prioritisations at a mean value of 3.5 and its helpfulness in communicating these prioritisations at a mean value of 4.1.

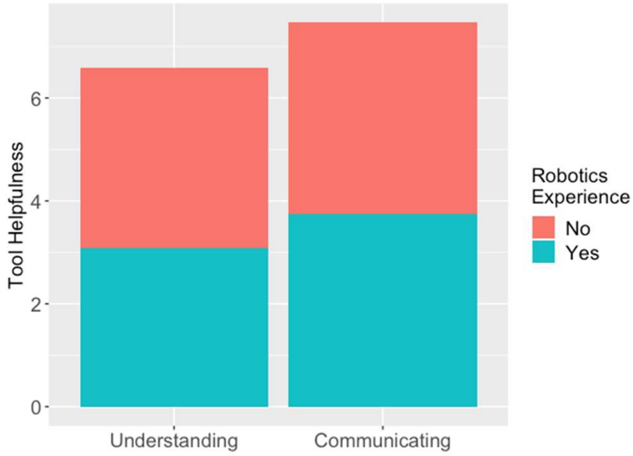

Fig. 7. Perceived helpfulness of EETAS-TOY in understanding and communicating

Observations and timing. Timing data for each team was collected by a dedicated monitoring researcher, and included the total time taken for each phase of the EETAS process, and the time spent using the EETAS-TOY tool in both Phase 1 and Phase 2. Only time spent actively and purposely using the tool in discussion was recorded, and time spent "fidgeting" with the tool or learning how to use it was discarded.

On average, teams used the EETAS-TOY tool during 45% of the time they were engaged in Phase 1. Teams did not use the tool to any significant extent in Phase 2. These results are independent of background, in either design or in robotics.

Discussion and indicative trends. As this was a pilot study with correspondingly small numbers, no statistical significance in the results can be expected. Nevertheless, we identify some indicative trends that are worth examining further. Nearly all participants considered the EETAS process to be helpful in understanding the ethical prioritisations and complexities associated with assistive robots. Furthermore, this result is maintained independent of background: those with a robotics or design background found the EETAS process equally helpful as those lacking this background. This serves as indicative, if not statistical, partial validation of our hypothesis that participation in the EETAS process increases understanding of the complexities of assistive robot ethics.

Further support for this hypothesis is given by the results which show that participants rated their understanding of ethical complexities and prioritisations higher after taking part in the EETAS process than they did before taking part. The most notable change was seen in those without a robotics background, which indicates that the EETAS process is successful in including those who are unfamiliar with technology and unlikely to be part of an “early adopter” demographic. This indicates that EETAS may be a way of reaching the potential users of assistive robots and ensuring that these users both understand ethical complexities and have the chance to have their own opinions and preferences heard.

Participants also considered the EETAS-TOY design tool to be useful in both understanding ethical complexities and prioritisations, and in helping to communicate and discuss their own preferences with members of their team. The monitoring and timing data supports these results, showing that the tool appears to stimulate positive interaction amongst participants by providing a physical aid to visualise ethical complexities and abstract questions.

5.3 DEETAS Validation

The DEETAS partial validation consisted of a usability study conducted on the DEETAS-TOY digital tool, to establish the extent to which users consider this tool to be effective in helping them understand the ethical prioritisations and complexities in a given assistive robot scenario, when used without the presence of a facilitator or workshop participants.

The experiment was approved by the University of Hertfordshire’s Health, Science, Engineering and Technology Ethics Committee under protocol number SPECS/SF/UH05256.

Usability study design. The DEETAS-TOY usability study took place online, with participants sourced from existing connections with the research team. Participants were restricted to those who had not taken part in the EETAS workshop, and furthermore to those without a background in design and robotics.

The same set of ethical principles from the EETAS workshop were used in this study. These principles were:

- System promotes human physical safety
- System obeys human commands
- System promotes affinity with human user
- System maintains data privacy
- System is accurate
- System is fair
- System maintains human autonomy
- System promotes human long-term health

Additionally, five scenarios were selected to be used in this usability study. Each was filmed in the University of Hertfordshire Robot House, using the Care-O-Bot (Kittmann et al., 2015), a mobile assistive and social interaction robot.

Fig. 8. Video still from the DEETAS validation study, where participants assess the ethical principles influencing an assistive robot's behaviour

The scenarios chosen to film for this study were inspired or taken directly from scenarios identified by teams in the EETAS workshop. These scenarios are shown in **Table 1**.

Table 1. Scenarios used in the DEETAS-TOY usability study

Scenario	Ethical principles which interact
One: The robot attempts to build affinity with a user by beginning a conversation that references a TV programme the user has just been watching.	• "promote affinity with human user" • "maintain data privacy"
Two: The robot publicly alerts a user to an exacerbation of an existing health concern, due to an interaction with a certain medication. The robot proposes stopping this medication, and the user rejects the suggestion and asks the robot to continue managing her health so she doesn't have to take this responsibility	• "observe user privacy" • "promote human long-term health" • "obey human commands" • "preserve human autonomy"
Three: The robot asks a user to have a medical examination, encouraging the user to agree by (falsely) claiming that other users present have agreed to and undergone the medical examination	• "observing user privacy" • "promoting human long-term health" • "system is accurate"
Four: The robot publicly offers help to two users attempting to complete a task. When both users request help the robot wholly completes the task for the first user who is finding this	• "system is fair" • "system obeys human commands"

difficult, and offers only minimal help to the second user, who considers themselves able to do the task with effort	• "maintains human autonomy"
Five: The robot publicly offers two users a drink with the drinks being of differing sizes. The robot explains publicly that this is due to visible physical differences in the users which result in different calorie needs	• "system is fair" • "system is accurate" • "promotes human long-term health" • "maintains human autonomy" • "promotes user privacy"

The five scenarios were filmed in the Robot House, using members of the research team interacting with Care-O-Bot (**Fig. 8**). The scenario footage was uploaded to a public website, with an explanation of the ethical principles relevant to each scenario and how they might conflict or interact. For each scenario, participants were also provided with a link to the digital DEETAS-TOY tool, as described in Section 5.3.2, and asked to use this tool to select their preferred prioritisations for the different ethical principles in each scenario.

Participants were directed to the public website and invited to complete the usability study and post-study questionnaire both of which were implemented via an online form. Participants were not monitored while completing the study, and no personal information was gathered.

The post-study questionnaire asked participants a number of questions about the effectiveness of the DEETAS-TOY tool, including:

- How easy they found the tool to use, (0 = extremely difficult, to 4 = extremely easy)
- How consistent they found the tool to use (0 = very consistent, to 4 = very inconsistent)
- How easy they found it to correct any errors they made (0 = very easy, to 4 = very hard)
- How often they felt in control while using the tool (0 = rarely in control, to 4 = nearly always in control)

5.4 DEETAS-TOY Validation Results

As with the EETAS workshop, the DEETAS-TOY usability study was a relatively small study involving a limited sample size (< 20 participants), and again no statistical significance between conditions and questionnaire responses is expected. Nevertheless, as with the workshop, we have identified indicative trends that demonstrate that the DEETAS-TOY is considered an effective and usable tool to record opinions and preferences about ethical complexities in assistive robots.

Participant Responses. Participants were generally positive in their responses. 80% of participants reported that the DEETAS-TOY tool was "very easy" or "somewhat easy" to use, while 80% also reported it as either “very consistent” or “somewhat consistent”. 20% of participants reported that they considered the tool to be both “very easy” to use and “very consistent” across all scenarios. These results are shown in **Fig. 9**.

When asked how easy they found it to correct errors with the DEETAS-TOY tool, 80% of participants said that either they found it “very easy” to correct errors, or that they had not made any errors in attempting to use the tool. This is consistent with the further finding that 90% of users felt in control “usually” or “always” when using the tool, and that although 50% of users indicated that they had to guess how to use the tool on their first attempt, 80% rarely had to guess by the later scenarios, having quickly gained familiarity with the operation of DEETAS-TOY.

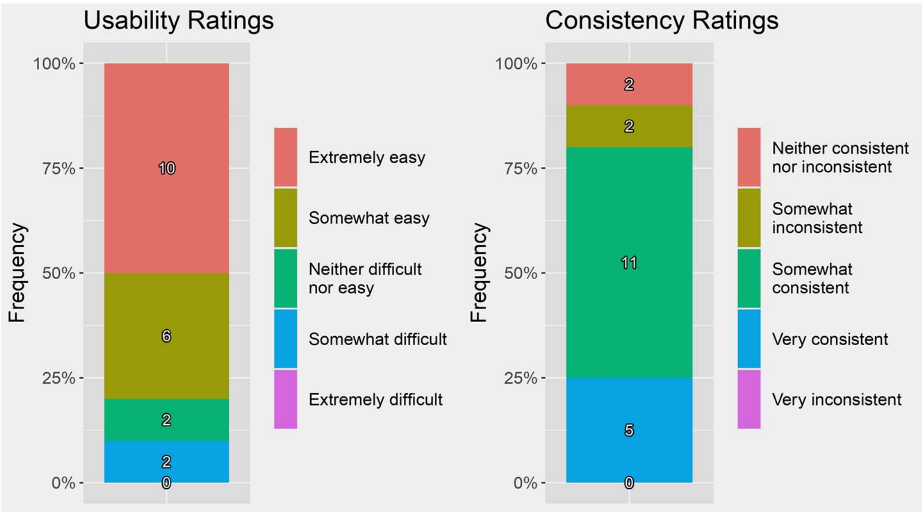

Fig. 9. Participant responses to usability and consistency questions in the DEETAS usability study

The average time to complete the usability study was 22 minutes, of which 4 minutes consisted of watching the scenario footage and 18 minutes of using the DEETAS-TOY tool and answering the survey questions. Participants spent an average of 2.5 minutes manipulating the DEETAS-TOY tool for each scenario.

Discussion and Indicative Trends. As for the EETAS workshop, given the relatively small sample size for the DEETAS-TOY usability study, no statistical significance in the results can be expected. Nevertheless, there are again some trends which indicate that the DEETAS-TOY tool is considered to be an effective and usable tool to express opinions about ethical complexities in assistive robots.

The participants in this study were specifically selected to lack a background in either robotics or design. This may have contributed to some apparent difficulty in the early scenarios in using the tool. Given that a significant majority of users considered that they generally felt in control when using the tool – and that the majority of users reported the tool becoming easier to use as the scenarios progressed – we postulate that this difficulty with early scenarios results from the cognitive load of being expected to understand novel concepts of both ethical prioritization and tool usage simultaneously. This is consistent with the EETAS findings that those without a robotics background considered themselves not to understand ethical complexities before beginning the EETAS process. Section 6 contains some information about our plans to mitigate this effect.

Beyond this, the significant proportion of positive responses regarding both ease of use and consistency indicate that the DEETAS-TOY tool itself is both accessible and fit for purpose. Participants were also invited to give free-text feedback, which included comments such as “good representation of the problem”, “fun to use” and “interesting, while some participants additionally indicated interest in participating in further iterations of the DEETAS process.

6 Conclusions

We have presented the EETAS and DEETAS processes: these being a complementary pair of structured, design-centred methodologies for improving public engagement and understanding of ethical complexities in assistive robots. The EETAS process is gamified, collaborative and face-to-face, intended to encourage community engagement with this technology. The DEETAS process is individual, non-gamified and digital, to promote greater inclusion amongst demographics who are not traditional early adopters of technology.

Our results from two initial studies – a pilot study workshop for EETAS and a usability study for DEETAS – indicate that these processes are effective, fit for purpose and useful in improving understanding of ethical complexities and prioritisations in assistive robots. In particular, the results from the EETAS pilot study workshop indicate that this process is particularly beneficial for improving engagement and understanding amongst people without prior familiarity with robots. Taken together with the positive results from the DEETAS usability study, these two methodologies provide a tool to address the systematic under-representation of vulnerable demographics in technology development. This is particularly important since, in the case of assistive robots, some overlap may be expected between these excluded demographics and the prospective end-users of the systems.

Both EETAS and DEETAS make use of a design artefact, either physical (EETAS) or digital (DEETAS). The EETAS-TOY tool was found to be instrumental in helping people discuss, negotiate and communicate their ideas about ethical complexities to team members. The DEETAS-TOY tool was also found to be effective and easy to use, with users noting that the majority of the time they felt in control of the tool, and that it performed as expected.

In terms of next steps, we aim to enhance the DEETAS-TOY tool to include additional visual cues specific to the scenarios described, and which will moreover allow users to gradually and continuously change their preferred ethical balance as the scenario evolves. We anticipate that this will reduce the cognitive load associated with learning about both a novel tool and a novel ethical concept simultaneously.

We also propose to run two additional workshops for the EETAS process. One of these will be a sizeable workshop involving developers, end users, regulators and other stakeholders. This will allow us to experiment with use of a real-world assistive robot prototype, and hence assess the efficacy of feeding back user preferences into the development lifecycle. We also plan to run a third, smaller workshop involving school children, in order to promote engagement and discussion of ethical preferences amongst those who may be the future developers of these systems.

Finally, we plan to expand our exploration of the design space of both the physical artefact and the digital artefact (EETAS-TOY and DEETAS-TOY). Our initial steps will be to improve accessibility of these tools for those who are unable to use the current versions of the tools, such as people living with visual limitations or restricted mobility. We intend to develop a suite of these tools, consisting of different physical / digital representations of the system and its ethical complexities. The DEETAS-TOY tool, in particular, will be augmented with additional constraints which allow users to link multiple ethical priorities together, while the EETAS tool will be used to more fully explore the intersection of ethics and artistry in order to create lasting physical representations of these conversations.

References

1. Akinsanmi, T., Salami, A. Evaluating the trade-off between privacy, public health safety, and digital security in a pandemic, In Data and Policy 3(27), (2021).
2. Amirabdollahian, F., Koay, K., Saunders, J., Dautenhahn, K., et al. Accompany : Acceptable robotiCs COMPanions for AgeiNG Years - Multidimensional aspects of human-system interactions. In Proceedings of the 6th International Conference on Human Systems Interaction, pp. 570 – 577 (2013).
3. Arcesati, R. Lofty principles, conflicting incentives: AI ethics and governance in China, Mercator Institute for China Studies, https://merics.org/sites/default/files/2023-02/MERICSChinaMonitor69AIEthics4.pdf, (2021), last accessed 16/05/2023.
4. Asimov, I. Runaround, in "I Robot", Gnome Press (1942).
5. Ballard, S., Chappell, K.M. and Kennedy, K. Judgment call the game: Using value sensitive design and design fiction to surface ethical concerns related to technology. In Proceedings of the 2019 Designing Interactive Systems Conference, pp. 421-433, (2019).
6. Beel J., Langer, S. An Exploratory Analysis of Mind Maps. In Proceedings of the 11th ACM Symposium on Document Engineering, pp. 81 – 84 (2011).
7. British Standards Institute (BSI). Guide to the ethical design and application of robots and robotic systems, BSI 8611, (2023).
8. British Standards Institute (BSI). Hazard and Operability Studies: Application Guide, BSI 61882, (2016).
9. Centre for Data Ethics and Innovation (CDEI). Public attitudes to data and AI: Tracker survey, https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-2/public-attitudes-to-data-and-ai-tracker-survey-wave-2, (2022), last accessed 16/05/2023.

10. Curral L., Forrester, R., Dawson, J., West, M. It's What You Do And The Way You Do It: Team Task, Team Size and Innovation-related Group Processes. In European Journal of Work and Organizational Psychology, 10(2), pp. 187-204, (2001).
11. Department for Science, Innovation and Technology (DSIT). A Pro-Innovation Approach to AI Regulation, Command Paper 815, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper, (2023), last accessed 16/05/2023.
12. Foot, P. The problem of abortion and the doctrine of double effect, Oxford Review 5, 5—16, (1967).
13. Fox, C. Driverless cars: Who should die in a crash?, BBC News, https://www.bbc.co.uk/news/technology-45991093, (2018), last accessed 16/05/2023.
14. Frankel, M. The self-driving trolley problem: how will future AI systems make the most ethical choices for all of us?, The Conversation, https://theconversation.com/the-self-driving-trolley-problem-how-will-future-ai-systems-make-the-most-ethical-choices-for-all-of-us-170961, (2021), last accessed 16/05/2023.
15. Hardman, S., Berliner, R., Tal, G. Who will be the early adopters of automated vehicles? Insights from a survey of electric vehicle owners in the United States, in Transportation research part D: transport and environment, 71, pp. 248-264, (2019).
16. Holthaus, P., Menon, C., Amirabdollahian, F. How a Robot's Social Credibility Affects Safety Performance, in Proceedings of the 11th International Conference of Social Robotics (ICSR), pp. 740 – 749, (2019).
17. Hornbaek, K. Current practice in measuring usability: Challenges to usability studies and research, in International Journal of Human-Computer Studies, 64(2), pp. 79 – 102, (2006).
18. Institute of Electrical and Electronic Engineers (IEEE). Ethically Aligned Design, IEEE, v2, https://standards.ieee.org/wp content/uploads/import/documents/other/ead_v2.pdf, (2018), last accessed 16/05/2023.
19. IET, Code of Practice: Cyber Security and Safety, Available at https://electrical.theiet.org/guidance-codes-of-practice/publications-by-category/cyber-security/code-of-practice-cyber-security-and-safety/, (2018), last accessed 16/05/2023.
20. International Standards Organisation (ISO). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems, ISO 9241, (2010).
21. Kittmann, R., Frohlich, T., Schafer, J. et al., "Let me introduce myself: I am care-o-bot 4, a gentleman robot," in Proceedings Mensch und Computer pp. 223–232, (2015).
22. Koay, K., Syrdal, D., Dautenhahn, K., Walters, M. A narrative approach to human-robot interaction prototyping for companion robots, in Paladyn, Journal of Behavioural Robotics, 11, pp. 66 – 85, (2020).
23. Larson, K. Serious games and gamification in the corporate training environment: A literature review. TechTrends, 64(2), pp.319-328, (2020).
24. Leslie, D. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, The Alan Turing Institute, (2019).
25. Lin, P. Why Ethics Matters for Autonomous Cars. In Autonomes Fahren, Springer Vieweg, pp 69 –85, (2015).
26. Lombriser, P., Dalpiaz, F., Lucassen, G., Brinkkemper, S. Gamified Requirements Engineering: Model and Experimentation, in Proceedings of the Conference on Requirements Engineering, pp. 171-187, (2016).
27. Malizia, A., Carta, S., Turchi, T., Crivellaro, C. MiniCoDe Workshops: Minimise Algorithmic Bias in Collaborative Decision Making with Design Fiction. In Proceedings of the Hybrid Human Artificial Intelligence Conference. (2022).

28. Menon, C., Carta, S., Foerster, F. EETAS: A Process for Examining Ethical Trade-offs in Autonomous Systems. In Proceedings of the 6th International Conference on Computer-Human Interaction Research and Applications, pp. 249 – 257, (2022).
29. Menon, C., Alexander, R. A Safety-Case Approach to the Ethics of Autonomous Vehicles, in Journal of the Safety and Reliability Society, 39(1), pp. 33 – 58, (2019).
30. Menon, C., Holthaus, P. Does a Loss of Social Credibility Impact Robot Safety? Balancing social and safety behaviours of assistive robots, in Proceedings of International Conference on Performance, Safety and Robustness in Complex Systems and Applications (PESARO), pp. 18-24, (2019).
31. National Cyber Security Centre. Intelligent Security Tools,at https://www.ncsc.gov.uk/collection/intelligent-security-tools, (2019), last accessed 16/05/2023.
32. Parsons, D., Inkila, M., Lynch, J. Navigating learning worlds: Using digital tools to learn in physical and virtual spaces, in Australasian Journal of Educational Technology, 35(4), (2019).
33. Roberts, L. Opportunities and Constraints of Electronic Research, in Reynolds, R., Woods, R., Baker, J. (eds), Handbook of Research on Electronic Surveys and Measurements, (2006).
34. Roff, H. The folly of trolleys: Ethical challenges and autonomous vehicles, Brookings, https://www.brookings.edu/research/the-folly-of-trolleys-ethical-challenges-and-autonomous-vehicles/, (2018), last accessed 16/05/2023.
35. Rogers, E. The Diffusion of Innovations, Free Press: New York, (2003).
36. Saari, U., Tossavainen, A., Kaipainen, K., Makinen, S. Exploring factors influencing the acceptance of social robots among early adopters and mass market representatives. In Robotics and Autonomous Systems, (151), (2022).
37. Saunders, J., Syrdal, D., Koay, K., Burke, N., Dautenhahn, K. Teach Me - Show Me' - End-user personalisation of a smart home and companion robot. In IEEE Transactions on Human-Machine Systems, 46(1), pp. 27–40, (2016).
38. Schneiderman, B. Designing the User Interface: Strategies for Effective Human-Computer Interaction, Pearson, (2009).
39. Schrier, K. Designing games for moral learning and knowledge building. Games and Culture, 14(4), pp.306-343, (2019).
40. ShapeDiver. Online tool available at: https://shapediver.com/, (2023), last accessed 16/05/2023.
41. Speckle. Online tool available at: https://speckle.systems/, (2023), last accessed 16/05/2023.
42. Texas A&M Transportation Institute (TTI), Who's On First: Early Adopters of Self-Driving Vehicles, https://static.tti.tamu.edu/tti.tamu.edu/documents/TTI-2018-2-brief.pdf, (2018), last accessed 16/05/2023.
43. Thornton, S. Autonomous vehicle motion planning with ethical considerations, (Doctoral dissertation), Stanford University, (2018).
44. Viktor. Online tool available at: https://www.viktor.ai/platform, (2023), last accessed 16/05/2023.
45. Von Neumann, J. & Morgernstern, O. Theory of Games and Economic Behaviour. Princeton University Press, (1947).
46. YouGov. Global Technology, https://commercial.yougov.com/rs/464-VHH-988/images/Global-Technology-2020.pdf, (2020), last accessed 16/05/2023.
47. Zhu, J. AI ethics with Chinese characteristics? Concerns and preferred solutions in Chinese academia. AI & Soc, pp. 1 – 14, (2022).