X-HAZOP: A Family of Techniques For Ethical Hazard Analysis of Assistive Robots

Catherine Menon, Austen Rainer, Patrick Holthaus, Sílvia Moros, Gabriella Lakatos

***Abstract*—Determining the ethical acceptability of assistive robots is a complex task, not least because the ethical hazards of such systems manifest differently across different demographics. Users of these robots are potentially likely to be already marginalised by technology and thus vulnerable to a wider range of ethical hazards than other demographics. Adequate assessment and mitigation of the robot's ethical hazards therefore requires creativity, collaboration and a range of diverse perspectives. This paper presents X-HAZOP, a family or toolbox of techniques for conducting ethical hazard analysis of an assistive robot by utilising structured, facilitated workshops. We present the findings from multiple workshops, demonstrating that the use of X-HAZOP techniques with a suitably diverse group of participants improves creation of accessible descriptions of the robot, aids understanding, and leads to an effective identification of a range of ethical hazards.**

***Index Terms*—Ethics, robotics, ethical AI, creativity**

I. Introduction

THE ethical hazards of a system manifest differently across different demographics, and cause differing degrees of ethical harm. Demographics which have been traditionally marginalised by technology - for example those people affected by economic disadvantage, bias or institutionalised discrimination - are more likely to be vulnerable to ethical hazards, not least because they often lack the opportunity to become involved in the design and development of technological systems and the consequent identification and mitigation of their ethical harms.

This poses a particular problem for assistive robots, whose users might reasonably be expected to fall into these marginalised demographics for reasons of age, frailty or cognitive and physical vulnerability. This concern is compounded by the fact that, as there is no existing standardised process for ethical hazard analysis, development teams will often resort to ad hoc or "best guess" methods. These ad hoc methods require a deep understanding of the specification and behaviour of the assistive robot, which militates against the inclusion of any participants from marginalised demographics who are unlikely to have the required technological background. This in itself can lead to ethical harm by further marginalising those from already-marginalised demographics.

In this paper we present a proof-of-concept of X-HAZOP: a family or toolbox of participatory design process methodologies for ethical hazard analysis of an assistive robot. All X-HAZOP techniques are based on Hazard and Operability Analysis (HAZOP) [1], an existing standardised process for safety assessment which makes use of pre-defined guide words to identify safety hazards.

Catherine Menon (c.menon@herts.ac.uk), Patrick Holthaus (p.holthaus@herts.ac.uk), Sílvia Moros (s.moros2@herts.ac.uk), Gabriella Lakatos (g.lakatos@herts.ac.uk) are with the Department of Computer Science, University of Hertfordshire

Austen Rainer (a.rainer@qub.ac.uk) is with Queen's University, Belfast

The two X-HAZOP techniques used for ethical hazard analysis are CHAZOP (creative HAZOP) and EHAZOP (ethical HAZOP). These methodologies are used to firstly create understandable and accessible narrative text descriptions of an assistive robot, and to secondly utilise these narrative texts for guided identification of the ethical hazards associated with an assistive robot.

As can be seen, the name "X"-HAZOP refers to the overall family of methodologies, with each individual methodology in the family named to represent its individual purpose, as shown in Table I.

TABLE I: X-HAZOP methodologies

X-HAZOP	A family of HAZOP-based techniques:
CHAZOP	Creative HAZOP: Create narrative texts describing the robot
EHAZOP	Ethical HAZOP: Utilise the CHAZOP texts for ethical hazard analysis

Section II of this paper describes the background and related literature. Section III introduces CHAZOP and EHAZOP, while Section IV summarises the workshops held to validate these methodologies. Section V presents discussion and analysis, and Section VI concludes.

II. Literature Review

Existing work on the ethics of robots and AI systems [2]–[4] describe the ethical properties which are desirable for these systems, including fairness, transparency, freedom from bias and preservation of privacy. However, the majority of these works do not specifically discuss the identification and mitigation of ethical hazards, nor do they present a reproducible or generalised process for this.

One exception is [5] which considers how ethical hazards might be identified within a hypothetical smart toy as part of a prototype case study. Similarly, [6] proposes a reproducible methodology for ethical hazard analysis, but presents this as a proof of concept only. Two existing standards [7], [8] present an enumeration of ethical hazards associated with, respectively, robotic systems and empathic systems. While both these standards note the importance of identifying ethical hazards, they also do not include a description of the process by which this could be achieved. Existing hazard assessment methodologies, including both HAZOP itself [1] and Failure Modes and Effects Analysis [9], present structured and justifiably effective techniques for hazard identification, but do not offer any suggestion as to how these might be extended to consider ethical issues.

In any form of ethical assessment, it is essential to consider diversity of stakeholder perspectives. Existing work such as [10] identifies a lack of voices from the Global South, while [11], [12] note the importance of diverse participation in AI and technology development, and highlight that it is those who lack economic or social power who have traditionally been excluded from such participation. Works such as [13], [14] acknowledge the importance of narratives and narrative texts in constructing public understanding of software and AI; this is particularly relevant to ethics, as the *perception* of bias can cause ethical harm regardless of its objective presence. Existing work [15] also considers how different forms of prose, such as poetry, can aid our underlying conceptual understanding. However, these works largely do not extend to recommendations of how narratives and narrative texts can be constructed or utilised to increase awareness of ethical hazards and aid inclusive design.

In this paper we address these gaps via the identification of a proof-of-concept participatory design process methodology, consisting of members of the X-HAZOP family. The X-HAZOP methodologies build on [6], using sets of pre-defined guide words which are in turn applied to elements of the narrative texts or elements of the assistive robot. The application of these guide words creates open-ended "what if?" questions which are used to springboard deeper understanding of how the robot functions and the ethical hazards it can create.

III. EHAZOP AND CHAZOP

The two X-HAZOP methodologies we discuss in this paper are EHAZOP (Ethical Hazard Analysis) and CHAZOP (Creative Hazard Analysis). CHAZOP facilitates the creation of *narrative texts* describing the system in an accessible, understandable way, in order to increase the pool of potential participants in the ethical analysis. Narrative texts will typically include vignettes, user stories, personas and other textual descriptions of the system and its functions. They may also include accessible fictional texts which explore the societal integration and positioning of the system, and may take many forms including stories, poems or flash fiction.

The narrative texts generated with CHAZOP then serve as input to the EHAZOP methodology. Using these narrative texts, EHAZOP facilitates the identification of ethical harms which might eventuate from use of the system or its functions.

The full process, from generation of narratives to identification of ethical hazards, is shown in Figure 1

A. *CHAZOP: Creative HAZOP*

The CHAZOP methodology makes use of a set of pre-defined guide words, and produces accessible and understandable narrative texts which describe aspects of the system. The CHAZOP guide words are based on the existing HAZOP guide words [1], which reflect the ways in which information relevant to the system might be compromised in the act of transmission to the user (in this case, reader). The CHAZOP guide words are shown in Table II.

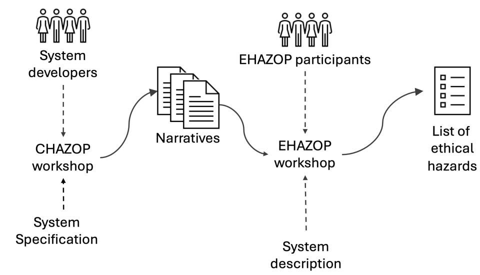

Fig. 1: Ethical hazard analysis using CHAZOP and EHAZOP

TABLE II: CHAZOP guide words

TOO MUCH	Narrative text over-emphasises this aspect, at a cost to understandability or reader attention
EARLY	Narrative text conveys information to reader earlier than apparently intended
LATE	Narrative text conveys information to reader later than apparently intended
DIFFERENT	Information is different to or inconsistent with other information within the narrative text
IN ADDITION	Narrative text conveys this information to reader in addition to other explicitly-stated informationt
NEVER	Narrative text never conveys this information to reader with sufficient clarity

The CHAZOP guide words are applied to different components of an initial narrative text (termed *narrative components*) as shown below:

- Plot - the structure and relationship of events within the narrative text
- Setting - the environment (physical, thematic etc.) wherein the narrative text takes place
- Character - the people or other entities described within the narrative text
- Theme - the development of underlying ideas, perspectives or resonances not necessarily directly stated in the narrative text
- Structure - the composition and organisation of the narrative text
- Point of view - the perspective from which the narrative text is told

The application of a guide word to any given narrative component creates a "what if?" question intended to help improve accessibility and understandability of the text. Examples of this might include:

- What if this ⟨ CHARACTER ⟩ were encountered ⟨ EARLIER ⟩ in the narrative text?
- What if the ⟨ SETTING ⟩ were ⟨ DIFFERENT ⟩ at different points in the narrative text?
- What if the ⟨ THEME ⟩ is being presented with ⟨ TOO MUCH ⟩ emphasis in the narrative text?

CHAZOP participants collaborate to explore these questions and amend the narrative texts based on the ensuing discussions. Potential CHAZOP output narrative texts might include: an accessible factual description of the robot's behaviour, a simple use case scenario describing how an end-user might interact with the robot, or a fictional story about the potential impact of assistive robots on society.

The purpose of the narrative texts is to describe and site the system (in this case, an assistive robot) in diverse textual forms which are understandable by a range of stakeholders. Once complete, these narrative texts are used to aid the EHAZOP process as described below.

B. EHAZOP

The EHAZOP methodology also makes use of a set of pre-defined guide words. As with CHAZOP, the EHAZOP guide words are based on existing HAZOP guide words, reflecing how information might be compromised in the act of transmission to the user (in this case, end-user of the assistive robot).

Before the EHAZOP process begins, the participants are provided with information about the robot, in the form of the narrative texts output from CHAZOP. Using these texts, EHAZOP participants work together to apply the pre-defined guide words to the different characteristics of the robot. The EHAZOP guide words are presented in Table III.

TABLE III: EHAZOP guide words

MORE	This characteristic or function of the robot is greater or increased from that expected by the user
LESS	This characteristic or function of the robot is less or diminished from that expected by the user
EARLY	This characteristic or function of the robot occurs or is encountered earlier than the user expects
LATE	This characteristic or function of the robot occurs or is encountered later than the user expects
OPPOSITE	This characteristic or function of the robot is the opposite of that expected by the user
DIFFERENT	This characteristic or function of the robot is in some way different to that expected by the user
IN ADDITION	This characteristic or function of the robot is performed or encountered in addition to a different one expected by the user
NEVER	This characteristic or function of the robot is not performed or encountered despite being expected by the user

EHAZOP guide words are applied in turn to the following characteristics of the assistive robot:

- Robot functions and behaviour
- Robot physical design
- Extent of robot autonomy implemented for any given function or behaviour

EHAZOP guide words may also be applied to a combination of robot characteristics simultaneously. Applying a guide word to any given function and/or property creates a "what if?" question relating to the user's expectations, which can be further explored to identify the ethical harms. Examples of this might include:

- What if this function were provided ⟨ EARLIER ⟩ than the user expects?
- What if this function had the ⟨ OPPOSITE ⟩ effect to the user's expectations?
- What if this function were provided with ⟨ LESS ⟩ ⟨ AUTONOMY ⟩ than the user expects?

EHAZOP participants collaborate to explore whether any ethical harm could result from these "what-if" situations. EHAZOP output might include a list of diverse, creative ethical hazards which take into account all participant perspectives.

A characteristic of all HAZOP-based techniques is that not all applications of every guide word will be relevant. Moreover, there are multiple interpretations of how the guide words might be applied for any given situation. The "what if?" questions produced by CHAZOP and EHAZOP should therefore be used as starting points for further discussion.

IV. Validation workshops

Two validation workshops were held for the CHAZOP process, and two for the EHAZOP process. The workshops were approved by Queen's University Belfast Faculty of Engineering and Physical Sciences' Ethics Committee under protocol number EPS 22297 and the University of Hertfordshire Ethics Committee under protocol number SPECS-SF-UH 05738.

A. CHAZOP Validation

The CHAZOP validation workshops were carried out at the Crescent Arts Centre, Belfast. This venue was chosen as it is a collaborative teaching and writing hub for the professional writing community, and therefore enabled us to source the necessary professional writing participants.

Five participants for each CHAZOP workshop were recruited using social media advertising and personal outreach by Crescent Arts. The participants were either professional or amateur writers of fiction with an expressed interest in improving creativity. All participants were asked to submit a written narrative text prior to the workshop.

For both workshops, after obtaining consent participants were provided with the CHAZOP guidewords and given a short introduction in how to apply these. Participants were also asked to confirm that they had read the submitted narrative texts. Participants were then divided into groups of 5 - 7, and asked to conduct an CHAZOP process on each of the written narrative texts in turn, facilitated by staff from Queen's University Belfast and the University of Hertfordshire.

1) Post-study questionnaires: Following both of the workshop, participants were asked to complete an anonymous post-study questionnaire, requesting both qualitative and quantitative feed-back. The questionnaire included the following questions, in order to determine the effect that CHAZOP had on participants' perceived effectiveness of their narrative texts.

- Participants were asked for their level of writing experience (experienced, partially experienced, or none)
- Participants were asked how helpful they considered the CHAZOP process across different forms of text, being: story, narrative, flash fiction, poetry (0 = very unhelpful, 5 = very helpful)
- Participants were asked how helpful they found the CHAZOP process as a writer vs as a reader of the narrative text (0 = very unhelpful, 5 = very helpful)
- Participants were asked how helpful they considered the CHAZOP process to both novice and professional writers (0 = very unhelpful, 5 = very helpful)

B. EHAZOP Validation

The EHAZOP validation workshops consisted of a proof-of-concept workshop carried out in the University of Hertfordshire's Robot House[1], and a half-day workshop carried out at Bishops Girls' School, Hatfield. The Robot House was selected as a venue as this enabled us to use the social robot ARI, which is installed within the Robot House; ARI itself was chosen as it is a social and collaborative humanoid robot equipped with a touchscreen, gaze direction and movement control [16].

The Bishop's Girls' School was selected as a venue as we wished to reach a specific demographic (schoolchildren), as a population more likely to be impacted by emerging technologies such as assistive robots. The school offered as participants a class for whom a computer ethics outreach event had already been organised; we therefore had the opportunity to conduct the workshop as part of this event. The robot used in this workshop was Pepper, a smaller humanoid robot similarly equipped to ARI [17]. Pepper was chosen as the schoolchildren had already had some experience viewing videos of this, and due to its small size making it able to be transported and used within a classroom. Thirteen participants between 13 - 17 years old took part in this second workshop, all of whom were secondary students at the school. Consent was obtained from both the guardians and the students themselves.

The proof-of-concept workshop using ARI is shown in Figure 2. Five adult participants took part in the workshop, with backgrounds ranging from architecture to robotics. The participants were all university researchers and were chosen for their range of backgrounds and previously-expressed interest in taking part in proof-of-concept workshops.

Fig. 2: EHAZOP workshop with ARI in the University of Hertfordshire's Robot House

All participants in both EHAZOP workshops were provided with the following three functions to assess for ethical hazards for ARI or Pepper.

1) Cognitive function **(Cog1)**: "At a specified time the robot reminds the user to take their medication".
2) Social function **(Soc1)**: "From monitoring of user activity and expression, the robot detects that the user is lonely and offers to set up a video call with a relative".
3) Coach function **(Coa1)**: "After an interval has gone past without any user physical movement the robot suggests the user engage in a sequence of stretching exercises, during which it monitors and provides feedback".

[1]https://robothouse.herts.ac.uk

For both workshops, after obtaining consent participants were provided with the EHAZOP guidewords and given a short introduction in how to apply these. Participants were also introduced to the robot being used - ARI or Pepper - shown its functionality and given a chance to ask questions about this. Participants were then divided into groups of 5 - 7, and asked to conduct an EHAZOP process on the robot functionality as described.

1) Post-study questionnaires: Following the half-day workshop, participants were asked to complete an anonymous post-study questionnaire, requesting both qualitative and quantitative feed-back. The questionnaire included the following questions, to elicit user feedback about the effectiveness of the EHAZOP process:

- Participants were asked whether they thought any of their identified ethical hazards would not have been identified without EHAZOP (0 = not at all, to 5 = certainly)
- Participants were asked how helpful they found the different aspects of EHAZOP, such as team-working (0 = very unhelpful, 5 = very helpful)
- Participants were asked to rank the three functions **Cog1, Soc1, Coa1** in terms of how ethically hazardous they considered these, following completion of EHAZOP

The ethical hazards identified by participants using EHAZOP were also recorded, as were the associated discussions.

C. Results - CHAZOP

Due to the background of the participants, all submitted narrative texts were pieces of fiction (story or narrative), some of which were incomplete. All participants had at least some partial experience with writing when all four assessed forms of story, narrative, flash and poetry were considered, but no participant was fully experienced with all. The distribution of writing experience is shown in Figure 3.

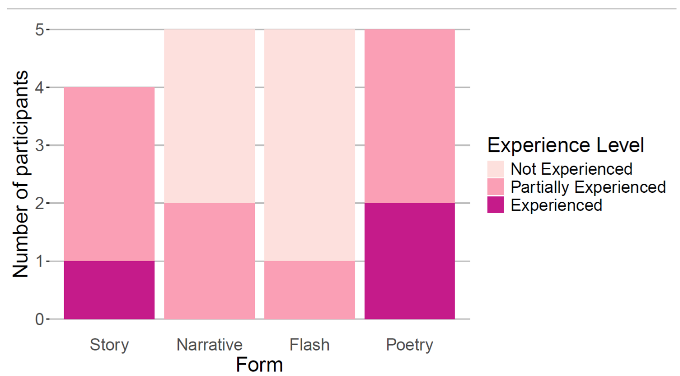

Fig. 3: Participants' writing experience across different forms

From the post-study questionnaire, participants considered CHAZOP to be of benefit across all writing forms, with the strongest agreement for the forms of story and poetry (40% - 100% across the forms of narrative, story and poetry). There was no correlation between participants' own experience in these forms, and their consideration of whether CHAZOP was useful for each. These results are both shown in Figure 4.

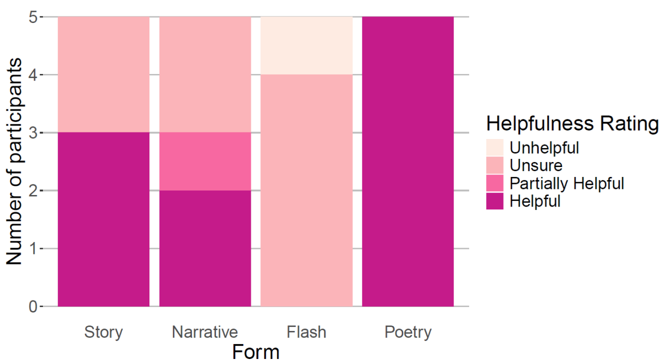

Fig. 4: Participants' rating of CHAZOP helpfulness across different forms

Participants considered CHAZOP to be helpful in assisting them to produce accessible and understandable narrative texts (87.5% average strength of agreement). Additionally they found CHAZOP helpful as a team member critiquing and improving the narrative texts (93.5% average strength of agreement). These results are as shown in Figure 5. (Note that this figure reflects four participants only: one of the participants had to leave mid-way through the workshop due to illness, which did not affect any of their prior answers).

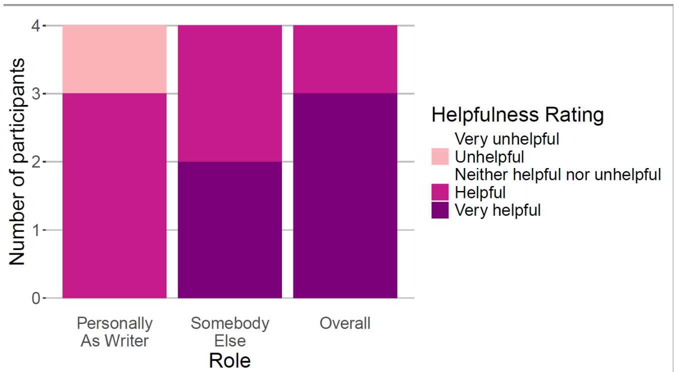

Fig. 5: Helpfulness of CHAZOP for different workshop roles

Participants considered CHAZOP to be of most benefit to inexperienced writers than to professional writers (80% strength of agreement), and to be equally helpful when considering both complete and incomplete narrative texts (80% strength of agreement across both).

Free-text responses are not fully analysed here but include the following:

- "This is an excellent roadmap for giving beginner writers the tools to evaluate their own and others' writing"
- "Very helpful to enhance creativity - a sense of structure can inspire new ways of thinking"
- "Excellent workshopping process which helped me deal with structural elements of my work in progress"

TABLE IV: Ethical hazards identified by EHAZOP

Function	Hazard	Notes
All	Lack of privacy	The user's privacy is compromised by the robot's monitoring
All	Lack of informed consent	The user did not consent to monitoring by the robot, or has forgotten this
All	Dehumanisation	The user begins to see the robot as an authority figure
All	Dehumanisation	The user is physically or mentally intimidated by the robot
All	Deception	The user believes the robot is monitoring them when it is not
All	Loss of trust	The user no longer trusts the robot to perform its functions
Cog1	Inappropriate trust (deception)	The user begins to trust the robot to facilitate wider medical activities
Coa1	Loss of human autonomy	The user loses ability to recognise body cues for exercise, or to perform these without coaching
Coa1	Loss of human control	The user temporarily loses ability to concentrate or focus due to repeated interruptions
Coa1	Robot addiction	The user begins to prefer interacting with the robot to other people, as a result of these interruptions
Coa1	Lack of respect for cultural diversity & pluralism	The user's culture does not align with the wellness values the robot facilitates
Soc1	Lack of respect for cultural diversity & pluralism	The user's culture does not align with the social expectations the robot facilitates
Soc1	Dehumanisation	The user begins to consider their own facial expressions problematic
Soc1	Erosion of confidence*	The user begins to question their own desires and feelings based on the robot's prompts
Soc1	Lack of associative control*	The user's mental associations with socialising alter as a result of their interactions with the robot
Coa1	Cultural flattening*	The user begins to lose appreciation of, or understanding of, their own cultural characteristics as a result of question the robot's prompts

D. Results - EHAZOP

Time constraints meant participants in both EHAZOP workshops considered only a subset of the EHAZOP guide words and the robot functions, and therefore the results should not be taken to be exhaustive. Nevertheless, a comprehensive selection of ethical hazards was identified in both workshops, and shown in Table IV.

Ethical hazards in Table IV align where possible with the hazards listed in BS8611 [7]. Where a hazard presents in a number of different unique ways, these have been included as distinct entries. Novel ethical hazards not identified in BS8611 are marked with a * and discussed further in the following sections.

From the half-day workshop, of those who completed the relevant question in the questionnaire, 75% considered that they would have failed to identify at least one of the ethical hazards in Table IV without the use of EHAZOP: these hazards being **dehumanisation, lack of privacy** and **cultural flattening**. Participants also identified that the primary strengths of EHAZOP were team-working (97%) and the structured nature of the what-if questions (85%).

V. Discussion and Analysis

Although the relatively low participation numbers for both the CHAZOP and EHAZOP workshops mean that no results of statistical significations were obtained, there are nonetheless some promising correlations and trends.

A. EHAZOP discussion

EHAZOP participants considered EHAZOP to be of benefit in terms of identifying hazards that would otherwise have been missed. One ethical hazard identified in both EHAZOP workshops is **lack of respect for cultural diversity & pluralism**: in discussions, users identified that different cultures might place greater or lesser value on practices such as stillness, silence and meditation, which the robot's **Soc1** and **Coa1** functions do not accommodate. It is worth noting that the participants in the half-day workshop identified this ethical hazard more quickly than participants in the proof-of-concept workshop: we partially attribute this to the different demographics and suggest that this could usefully lead to further future investigation.

Three novel ethical hazards not included in [7] also emerged from the EHAZOP workshops: **erosion of confidence, cultural flattening** and **lack of associative control**. The first of these relates to the impact of a robot's continual prompts based on an assumption about the user's desires or requirements; participants noted that some users may be convinced against their own beliefs that they did in fact feel the prompted desire to socialise. The second, **cultural flattening**, is related to the ethical hazard of **lack of respect for cultural diversity and pluralism** [7], in that cultural flattening occurs when the user begins to believe that the robot's prompting - e.g. to engage in movement - reflects a culture in which they *ought* to participate, and as a consequence begins to lose familiarity with or appreciation of their own culture. The third, **lack of associative control**, arises from the fact that the robot presents socialising as a remedy for a negative situation (boredom), rather than an enhancement to a positive situation. Participants identified this risked affecting the user's mental associations with social activities.

B. CHAZOP discussion

The CHAZOP process was considered by a majority of the participants to be helpful for both complete and incomplete narrative texts. This indicates that CHAZOP may be used at any stage of constructing narrative texts, hence enabling ethical analysis to be interleaved with design.

The participants also considered CHAZOP to be of benefit to all writers, but most emphasised its value to novice writers. As those involved in designing or running ethical hazard analyses for assistive robots are unlikely to also be professional writers, this indicates that CHAZOP may be of significant real-world benefit when attempting to construct narrative texts which aid in diverse participant understanding.

For both CHAZOP and EHAZOP, participants all considered that the most helpful results were obtained in the discussion stemming from the guide words. The associative links between guide words, narrative texts and ethical hazards is extremely valuable, and thus we would suggest that the facilitator role should transition from facilitator to amanuensis in the latter part of the workshops.

We also found that not only was creativity enhanced during both CHAZOP and EHAZOP workshops, but collaboration was also fostered. Participants self-reported that they were able to consider suggestions for their narrative texts – or suggestions for amendments to their identified ethical hazards – and build on these as part of a team. In addition to this, the majority of the participants in both workshops identified that such suggestions would be viewed more positively due to the collaborative focus of both workshops.

C. Workshop limitations

Although promising, there are some limitations of these workshops which mean that these results should be considered as proof-of-concept only, and require further confirmation and validation as described in Section VI. The small sizes of the workshops and the timing restrictions have constrained the full application of both methodologies, and hence these initial results must be substantiated with additional more extensive and longer workshops which fully exercise the methodologies and focus on diversity of participants.

Furthermore, the workshops have utilised intentionally simple scenarios focusing only on assistive robots. Within industry, robots may be used for factory, cleaning, manipulation and human-interaction tasks that increase the complexity of performing both EHAZOP and CHAZOP. In particular, there may be little end-user understanding of the functionality or narratives around industrial robots used for specialised activities such as assembly and disassembly on factory lines, as well as for robots used within safety-critical systems. This represents a practical challenge with extrapolating these results to more complex real-world systems, and will be addressed in future work.

VI. Conclusion

We have proposed X-HAZOP, a family or toolbox of methodologies for conducting ethical hazard analysis, and considered the case study example of an assistive robot. We have described how the use of CHAZOP, one of the X-HAZOP methodologies, enables development teams to produce understandable, accessible narrative texts describing the robot. These texts facilitate inclusion of end-users or other people without a technological background in the ethical analysis process, thereby increasing diversity and representation. We have also described how the use of another X-HAZOP technique, EHAZOP, facilitates guided, collaborative identification of ethical hazards by drawing on the accessible narrative texts produced from CHAZOP. We have also presented the results of preliminary workshops, which demonstrate the ability of CHAZOP to contribute to understandable and clear narrative texts, and the ability of EHAZOP to aid in generation of novel ethical hazards.

We propose to build on this work by running a future combined EHAZOP and CHAZOP workshop with a range of participants across demographics. This workshop will allow us to assess the effect of an end-to-end complete ethical hazard identification process, from generation of narrative texts to identification of ethical hazards. Furthermore, we also intend to complement the EHAZOP and CHAZOP methodologies with further X-HAZOP methodologies which focus more specifically on challenges which impact multiple domains, such as the intersection of ethics, safety and security of public-facing systems. A more detailed application of X-HAZOP methodologies would also require further scrutiny of the guide words used, and we propose undertaking further work which compares the existing HAZOP-inspired guide words with words which specifically map to known ethical or textual vulnerabilities of a system or narrative.

REFERENCES

[1] British Standards Institute, "BS61882 - Hazard and Operability Studies Application Guide," standard, British Standards Institute, 2016.

[2] Institute of Electrical and Electronic Engineers (IEEE), "Ethically Aligned Design," standard, Institute of Electrical and Electronic Engineers (IEEE), 2018.

[3] D. Leslie, "Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of ai systems in the public sector," tech. rep., Turing Institute, 2019.

[4] Observatory of Public Sector Innovation, "Ethical OS Toolkit." https://oecd-opsi.org/toolkits/ethical-os-toolkit. Acccessed 2024-11-06.

[5] A. Winfield, A. van Maris, K. Winkle, M. Jirotka, P. Salvini, H. Webb, A. Scott, J. Freeman, L. Kunze, P. Slovak, and N. Theofanopoulou, "Ethical Risk Assessment for Social Robots: Case Studies in Smart Robot Toys," *Towards Trustworthy Intelligent Artificial Systems*, pp. 61–76, 2022.

[6] C. Menon, A. Rainer, P. Holthaus, G. Lakatos, and S. Carta, "EHAZOP: A Proof of Concept Ethical Hazard Analysis of an Assistive Robot," in *Proceedings of the 2024 IEEE International Conference on Robotics and Automation Workshop on Robot Ethics*, 2024.

[7] British Standards Institute, "BS8611: Robots and robotic devices – Ethical design and application of robots and robotic systems – Guide," standard, British Standards Institute, 2023.

[8] Institute of Electrical and Electronic Engineers (IEEE), "IEEE7014: IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems," standard, Institute of Electrical and Electronic Engineers (IEEE), 2024.

[9] International Electrotechnical Commission, "IEC61508: Functional safety of electrical - electronic - programmable electronic safety-related systems," standard, International Electrotechnical Commission, 2010.

[10] C. Roche, P. Wall, and D. Lewis, "Ethics and diversity in artificial intelligence policies," *AI Ethics*, pp. 1095–1115, 2022.

[11] A. Birhane, W. Isaac, V. Prabhakaran, M. Diaz, M. Eilish, and I. Gabriel, "Power to the People? Opportunities and Challenges for Participatory AI," in *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–6, ACM/IEEE, 2022.

[12] M. Cheong, K. Leins, and S. Coghlan, "Computer Science Communities: Who is Speaking, and Who is Listening to the Women? Using an Ethics of Care to Promote Diverse Voices," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 106–15, ACM/IEEE, 2021.

[13] A. Aldave, J. Vara, D. Granada, and E. Marcos, "Leveraging creativity in requirements elicitation within agile software development: A systematic literature review," *Journal of Systems and Software*, pp. 1–25, 2019.

[14] S. Cave, K. Dihal, and S. Dillon, *AI Narratives: A History of Imaginative Thinking About Intelligent Machines*. Oxford University Press, 2020.

[15] K. Simecek and K. Rumbold, "The Uses of Poetry," *Changing English: Studies in Culture and Education*, pp. 309–313, 2016.

[16] PAL Robotics, "ARI: The Social and Collaborative Robot." https://pal-robotics.com/robots/ari/. Acccessed 2024-11-13.

[17] Aldebaran United Robotics Group, "Pepper: The Humanoid and Programmable Robot." https://aldebaran.com/en/pepper/. Acccessed 2024-11-13.