Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 7335–7341

Transforming Experimental Cobot Cell to Industrial Realization – an Ethical AI Approach

Istvan Mezgár*, József Váncza*,, Imre Paniti*,****
József Tóth***

** EPIC Center of Excellence in Production Informatics and Control, Institute for Computer Science and Control, Budapest, Hungary (e-mail: mezgar.istvan@sztaki.hu).*
*** Department of Manufacturing Science and Technology, Budapest University of Technology and Economics, Budapest, Hungary (e-mail: vancza.jozsef@sztaki.hu, imre.paniti@sztaki.hu)*
**** Hepenix Ltd, Diósd, Hungary (e-mail: jozsef.toth@hepenix.hu)*

Abstract: Collaborative Robot (cobot) cells are getting more and more integrative building blocks of Cyber Physical Enterprises. These cells integrate the advantages of human workers with the special capabilities of robots in a safe manner. Cobots need advanced, in many cases artificial intelligence (AI) based control systems to harmonize the collaborative activities. When transferring/transforming experimental setups into industrial application, not only technological and business related, but also ethical aspects have to be taken into consideration. The paper introduces a novel workflow supporting this transformation and presents its application in a case-study of a cobot cell which uses advanced sensing, symbolic AI planning and mixed reality techniques for planning and explaining visually the operation of the cell. The work which takes the responsible artificial intelligence (RAI) approach combines the actual relevant AI standards with the explicit requirements of industrial practitioners.

Keywords: Collaborative robotics, Artificial intelligence, Ethics, Design methodology

---

1. INTRODUCTION

The development of different *artificial intelligence* (AI) technologies has speeded up in the last decade and in parallel increased their application in all sectors of the economy. AI systems have become a strategic factor in every domain, also in manufacturing. However, a limitation of AI system applications is the *trust* of users in the technology. According to different statistics the users need to have deeper knowledge on the operation of these systems, details on how the results have been generated. The demands of users are not focusing only on technical issues but have legal, social and economic implications as well. E.g., in Gillespie's survey (Gillespie, 2021) users' opinions in the USA, Canada, Germany, UK and Australia state that 28% of citizens are willing to trust AI systems in general. Two out of five citizens are unwilling to share their information or data with an AI system, and a third are unwilling to trust the output of AI systems.

It is obvious that trust is a key feature in using AI-based systems. The complex set of user demands on AI covers also the broad notion of *ethics*. By now it is commonly acknowledged that the ethical aspect in system design will be a strong competing factor already in the not too far future. The great problem of the AI community is how ethics can be decomposed to elements and transmitters through which ethical behavior can be realized in practical applications (Müller, 2021). While there are numerous attempts in creating sets of ethical principles, frameworks and guidelines that offer partial solutions for this problem (Corrêa, 2022), so far we know only a few successful applications of them (Leitão & Karnouskos, 2022); (Mezgár & Váncza, 2022).

In the *manufacturing* sector *robots* have an increasing role in many fields in automatizing production. In case of robots with AI software components, trust has even more importance as in other configurations (Kok, 2020). In some specialized fields (e.g., assembly) the collaborative robots (*cobots*) play significant role. Collaborative robots allow physical interaction with humans in a shared workspace to execute manufacturing or assembly tasks providing safe co-working based on sensors and special software modules (Sorell, 2022), (Cominelli, 2020), (Wang et al., 2019). These robots are designed to be reprogrammed easily, even by personnel without any programming background. The quick change-over time makes collaborative robots particularly interesting for small- and medium-sized enterprises that produce many different kinds of products in low volumes (Sherwani, 2020). Disadvantages of cobots are their limited speed and payload, and that they are not entirely autonomous.

Responsible AI (RAI) method has been developed for designing AI-based systems and embedded equipment taking into consideration ethical aspects along the whole life cycle, placing the human being (the user) in the center. Ethical guidelines try to provide help in system development, but there exists no general key, there is no standard route, there are rather individual solutions beyond the guidelines (Hagendorff, 2020).

Our paper presents an attempt to solve the ethical design problem in case of an industrial assembly cobot cell using the "ethics by design" method. Ethical design can be focused in case of cobots by placing the human in the center and raising trust as the main aspect by providing transparency of cobot actions, going beyond evident aspects of safety and ergonomics.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2023.10.347

The main goal of our work is to establish and increase trust of the worker in the cobot cell and in the assembly process, by explaining the decisions of a symbolic AI control system and providing visual prediction (and explanation) of the cobot arm path for the worker while taking into consideration safety and ergonomics aspects as well. In this way, a more comfortable and safe working environment can be developed that raises the well-being of the worker and as a consequence results in a more effective overall operation and a higher quality output.

The structure of the paper is as follows: Section 2 discusses the relations between ethics and AI, and the way how ethical aspects can be transformed to practical industrial applications using guidelines and the responsible AI method. Section 3 introduces the architecture, components and operation of the collaborative work cell. In Section 4 a case study is presented on how a prototype assembly cobot workcell with AI-based control can be redesigned for an industrial environment taking into consideration ethical aspects. Finally, Section 5 discusses the lessons learned during the project and some dilemmas, giving recommendations on the ethically conscious design and development of cobot cells.

2. ETHICAL AI

2.1 The relations between AI and ethics

AI-based systems can be found in all segments of the society, and the interaction of such systems with humans generates questions that involve ethical aspects as well, so ethics has a number of different approaches and definitions. A short definition of ethics given by Kuipers is “Ethics is a set of beliefs that a society conveys to its individual members, to encourage them to engage in positive-sum interactions and to avoid negative-sum interactions” (Kuipers, 2020). Applied ethics assigns what a person can do in a particular situation on a selected field of action in real-life situations: there is e.g., machine ethics, ethics of technology (techno-ethics), cyber-, and digital ethics. In case of artificial intelligence, the ethics of AI defines the ethical and moral obligations and duties both of an AI system and its developers.

In the design of AI-based systems three levels can be distinguished (Dignum, 2018):

- *Ethics for design* – series of standards and certifications providing integrity of stakeholders during the life cycle of AI systems.
- *Ethics in design* – engineering techniques for evaluating the ethical capabilities of AI systems, and
- *Ethics by design* – the technical integration of ethical reasoning capabilities.

In the field of manufacturing, and especially in robotics, human-robot (HR) collaboration (HRC) has a distinguished role. This relationship is strongly defined by human trust to the machine and to the process. Trust can be defined as a “psychological condition comprising the trustor’s intention to accept vulnerability based upon positive expectations of the trustee’s intentions or behavior” (Rousseau et al., 1998). Hence, ethics and trust are in close reciprocal connection. High trust environment encourages to build better ethics, while ethical behavior supports trust building. Specifically, trust in the technology is “the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability” (Lee and See, 2004). In case of socio-technical systems trust appears in different forms. According to (Luhman, 2018) four types of trust can be defined: (1) intrapersonal trust, a kind of self-confidence basic trust, (2) interpersonal trust, an expectation based on cognitive and affective evaluation of the partners, (3) system trust in depersonalized systems, let it be legal, or technical, and (4) object trust in non-social objects. Our solution supports the emergence of the first three types of trust.

2.2 Principles and guidelines for ethics in AI

Different organizations, universities and research institutions have suggested various proposals and guidelines for the development of trustworthy AI systems and applications. Surveys have analyzed the different AI ethical frameworks and approaches, along with making conclusions on the most important themes and principles (Fjeld et.al, 2020), (Jobin et.al., 2019), (Müller, 2020), (Hagendorff, 2020), (Corrêa, 2022). Accordingly, a convergence around five main ethical principles—transparency, justice and fairness, non-maleficence, responsibility and privacy—can be observed. In the technology field eight main principles have been identified, such as privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility and promotion of human values. These ethical principles should be mapped through ethical frameworks, toolkits and guidelines into working AI applications. The final conclusions suggest principled, guideline-driven development efforts with substantive ethical analysis and adequate implementation strategies.

However, the application of the above guidelines is neither flawless nor general yet. As applied in different cultural, technical environments they can provide only the first steps in the direction of developing ethical AI systems (Corrêa, 2022). According to (Hagendorff, 2020), “in practice, AI ethics is often considered as extraneous, as surplus or some kind of “add-on” to technical concerns, as unbinding framework that is imposed from institutions “outside” of the technical community”. There seems to be a gap between established principles and their actual use. The related documents only prescribe normative claims without the means to achieve them, while the effectiveness of pragmatic methodologies, in the majority of cases, remain extra empirical (Fjeld et al., 2020). Meanwhile, so-called “ethics washing” has become a common practice of tech companies (Bietti, 2021). This means the reductive instrumentalization of the rhetoric of ethics and morality using ethics only as a buzzword for marketing purposes.

2.3 Responsible AI

Responsibility is a main factor in handling highly automated and autonomous systems. In the relation of AI and ethics, *responsible AI* (RAI) represents a methodology that focuses on human responsibility along all phases of the system development process. RAI development warrants that AI systems have an acceptably low risk of harming their users while being beneficial (Askell, 2019). The responsible AI development pursues the phases and the theoretical considerations included in ethical frameworks (Peters, 2020). RAI is a guidance framework focusing on designing and implementing ethical, transparent and accountable AI solutions that help maintain individual trust and minimize privacy attack. RAI places humans in the center and implementing RAI means to meet relevant laws, regulations, and standards. During development the relevant standards have to be taken into consideration as well, such as (BS, 2016) or (IEEE, 2022).

3. DESIGN OF ASSEMBLY COBOT WORKCELL

3.1 The operation and architecture of the workcell

The technical goal of the cobot cell under investigation is to glue components of a workpiece and to fix cables on its surface. Operations of the robot and the human worker are executed in the same workspace, as far as possible, simultaneously. The economical goal is to increase the efficiency of the assembly process by reducing cycle time and improving gluing quality. The overall objective is to create and sustain trust of the human worker in the robotized workcell through the design of an ethical environment.

The basic tasks to be accomplished on the workpiece are (1) *gluing* two components with line gluing, (2) *locating* and pressing glued components, and (3) putting and fixing a set of cables on the workpiece, i.e., *cabling*. The gluing is done by a cobot while locating and cabling are realized by a human worker. Because of technological reasons and cycle time reduction these tasks should be executed in parallel, as far as possible. As in any cobot setting, it is essential to avoid collisions which in a tight workspace may easily happen.

The cobot system consists of two main subsystems:

- *Mixed Reality / Virtual Reality (MR/VR) cobot with AI-based planning and control system*. This takes as input the geometric model of the components and the completed workpiece in STEP files, the specification of sealings (gluing lines) and produces as output a sequence of points for controlling the cobot tool center point traversing the gluing line. The cobot control has a built-in module for sensing the possible collision with the human, and in case the cobot arm senses a collision it moves slower to avoid hard collision or harm of the worker or it stops.
- *MR/VR system for visualizing cobot arm motions*. This module visualizes the cobot arm path and the movements of the worker in real time. It takes as input sequences of the points of the cobot arm path and representation of the arm or any other body part of the worker. As output, it visualizes the motion of the worker and the cobot arm in 3D in a *mixed reality* (MR) glass in real time. The visualized cobot path is appropriate to explain the result of the AI-based planning.

The *mixed reality* (MR) system visualizes the movements of the human worker's arm and the cobot's arm (planned predicted gluing path). The possible collisions can be projected and detected in this way before they happen. The worker (who has MR glasses on) can modify his/her motion in time. If this action is too late, the collision avoidance system slows down the cobot arm not to cause any harm for the human. This impact is represented on the display as well. Hence, the MR/VR system adds extra safety as the calculated cobot path is handled as a predicted path that can be modified in real time according to real physical circumstances. The MR/VR display/representation can be used also as a qualitative *explanation* of the AI generated path as well. This explainable operation of the robot is the key to multi-agent teamwork in HRC (Kemény et al., 2021).

In order to develop an ethical working environment that generates and maintains trust in the worker the RAI development method has been applied during the design, implementation and operation of the workcell. Trust generation happens on 3 levels:

- The sensor-based collision avoidance system stops the robot when it gets into contact with the worker.
- The control algorithm alerts and executes an automatic stop when the end effector of the cobot moves outside of the given tolerance lane of gluing points/line.
- The explanation of the operation of the system is provided by (1) visual monitoring with the MR/VR of the robot arm moving with possibilities of emergency stop and replanning path, and (2) the declarative AI algorithm which can explain the path generation with guarantees.

3.2 AI in real-time robot task sequencing and path planning

One basic task of the workcell, the *dispensing of the glue material* on the surface of the workpiece, can be fully automated and robotized. For the solution of this problem, we suggest a generic and declarative approach which provides (1) a powerful, rich representation language, and (2) advanced search techniques for modelling and solving *process planning and sequencing* problems in industrial robotics. In particular, the ProSeqqo open-source system developed in our laboratory combines operational research (OR) and AI techniques for diverse industrial robotic application domains where the sequencing of various tasks should be accomplished together with making choices on how the tasks are performed (Zahorán and Kovács 2022).

Following the best traditions of theoretical research in AI planning, the problem can be defined using an intuitive, easy to comprehend and edit *problem definition language*. This representation is hierarchical: there are (1) processes on the top level, (2) alternatives for the possible ways of executing a process, (3) series of elementary tasks for executing an alternative, (4) multiple candidate motions for performing each task, and (5) for each motion, the sequence of configurations the robot must visit (see Fig. 1). Precedence constraints can be defined between two processes or two motions. Along with the problem, a rich set of optimization criterion can be defined in terms of cost factors of using and changing resources, of making moves, or of penalties for violating some requirements. All in all, while keeping overall costs minimal, three types of decisions must be made to solve a planning problem: (1) processes must be sequenced, (2) an execution alternative should be selected for each process, and (3) a motion has to be chosen for each task. The solver transforms this declarative problem definition into a generalized travelling salesman problem (GTSP) formalism and applies a combination of mixed-integer programming and local search methods to solve it. The efficiency of ProSeqqo makes it amenable for on-line application, too.

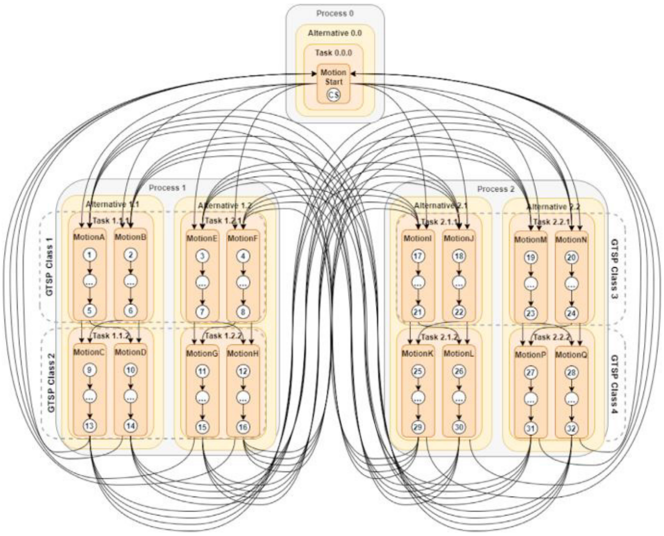

Fig. 1. Model hierarchy and GTSP representation of a robotic sequencing problem (Zahorán & Kovács, 2022).

The planning model which was motivated by robotic remote laser welding was successfully applied in diverse domains like robotic pick-and-place in semi-defined work environment, robotic laser engraving or 6D robotic grinding and polishing of free-form surfaces. In the gluing domain, we assume a glue dispenser is attached to the robot. The problem consists in grasping the appropriate dispenser, sequencing the sealings, choosing the right directions and the fine approach/detach motions. Processes define various sealings, with optional precedences between them. Various dispensers define execution alternatives for making the sealings with different tools and from different directions. Within each alternative, elementary tasks include approach, dispensing and detaching. For realizing the tasks, series of robot configurations can be pre-determined so as to avoid collision with any elements of the workcell, most specifically with the workpiece. The cost function is the total travel time assuming limited robot joint velocities and accelerations. This approach to the planning problem needs tedious preparation, but it is sufficient to focus on the hierarchical representation of the problem domain. The representation can be checked against the declared principles or guidelines of a design framework. At execution time, the solver warrants compliance with all declared constraints.

3.3. Mixed reality for robot path prediction

In parallel with the robot's operation, the human worker also executes some tasks in a shared workspace with the robot. In particular, the worker (1) locates the part to be fixed by the glued sealings and applies pressure to it, and (2) mounts some flexible cables on the workpiece. It is an elementary safety requirement to avoid collision between the two partners. In this partially defined and dynamically changing work environment we apply a double effort in which both parties mutually "see" each other and make predictions of the other's intentions thereby anticipating future conflicts. With appropriate look-ahead, actual collisions can be avoided.

- The robot continuously monitors the environment with a depth camera, compresses the point cloud into a voxelized digital twin of the worker and makes preventive actions (slowing, avoiding, stopping) if its projected path gets too close to the image (see Fig. 2-A).
- In a mixed reality the robot's future path is projected for the human worker (see Fig. 2-B). Hence, the worker is informed about the "intentions" of the robot and can plan and execute its own actions accordingly.
- In case we follow the movement of the worker with a depth camera and build a voxelized digital model of the human in the virtual space we have created, then we can examine collisions with the cobot's model moving forward in time (see Paniti et al., 2021). The module that draws the digital twin pair model of the cobot (see Erdős et al., 2020) must be informed of the starting section of the planned path, divided into joint angles, before issuing the signal controlling the robot's movement.

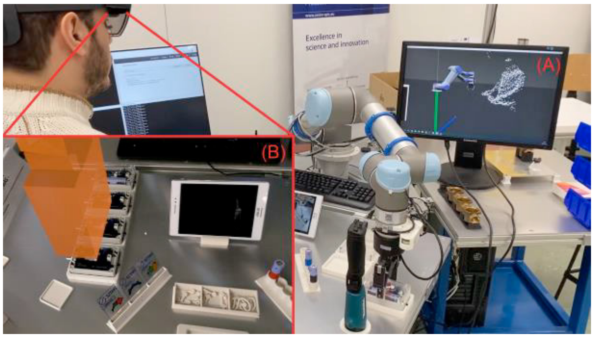

Fig. 2. (A): As the robot "sees" human movement in terms of voxelized point cloud in a single gripper set-up. (B): As the human sees the projected future cobot path.

In order to prevent the movement of the cobot and its digital twin from drifting apart in time during non-deterministic movements (e.g., when grasping tools of different sizes), synchronization points must be placed at certain points in the robot program. The projected movement of the cobot is followed with mixed reality glasses, and in addition to the visual warning, it can also emit a sound or haptic signal about the prospective virtual collisions (e.g., using a smart watch) (Paniti et al., 2021).

The system can adjust how far the cobot's digital twin should move forward, allowing enough time to correct the human's body position to avoid a real collision. If this does not happen within a given time, the robot's movement is slowed down, and then accelerated in the remaining time to maintain the cycle time. The visually predicted cobot path which explains continuously for the worker the operation of the robot fulfils the highest transparency level 5. according to the (IEEE, 2022) "Transparency requirements for users" standard.

4. ETHICS BY DESIGN – A CASE STUDY

When deploying AI-based systems to real industrial practice, a thorough, rigorous system design is of essential importance. This is especially so for systems where humans and robots meet or even collaborate at times. Beyond complying with the engineering requirements and the actual standards (see Safety Toolkit by COVR, 2023), we suggest here to take an "ethics by design" approach. First, such an attitude in system design can build confidence in the human users or collaborators of the system. Secondly, it is better if ethics related critical issues emerge as early and as explicitly as possible, so as to avoid the need of subsequent substantial modifications in the operating system.

In what follows we present a novel workflow for supporting the transformation of the above prototype gluing cobot cell using advanced sensing, planning and mixed reality techniques into an industrial system. The case-study takes a RAI approach involving also ethical aspects. In this process we rely on the integration of the Responsible AI Guidelines which operationalizes the US Department of Defense (DoD) general ethical principles of developing and deploying AI powered systems (Dunnmon et al., 2022), and the EC Guidance Ethics by Design and Ethics of Use Approaches for Artificial Intelligence (EC, 2021). Standards such as (BS, 2016), (IEEE, 2022) have been taken into consideration, too.

4.1 Phase 1: Planning

Identification of end users, stakeholders, and responsible mission owner is the first step of the planning phase. Assigning a mission owner is crucial as no AI system can be held accountable for its outcomes. In our case stakeholders include the System Developer, System Integrator, 3rd party Auditor, and End User. Mission owner tasks like assembly task design and robot programming are assigned to the System Developer.

Definition of tasks, quantitative performance metrics, and a baseline: AI algorithms aid in generating optimal paths for robots performing tasks such as gluing assembly components. The robot control task involves defining an optimized path for the end-effector, while avoiding collisions with the human worker in the permitted working space. Key success factor is the efficiency of the planner which makes it applicable in online manner. Accuracy of finding the gluing point and process time are also important metrics. As a feedback, gluing line accuracy can be measured with a camera on the robot arm. The insertion of cables is supervised by the worker and the key demand is the stable fixturing of the cables at the prescribed points. The primary criteria are cycle-time, end-effector accuracy, gluing line accuracy and cable stability.

Ownership of, access to, provenance of, and relevance of candidate data and models: During development, the developing company owns the models and datasets, but the industrial partner owns them after implementation. The stakeholders include assembly task designers and robot programmers. STEP format datasets are used to define the gluing path and stored securely, without collecting personally identifiable information. Worker identification data can be added for traceability. Quality of the complete workpiece, including gluing and cabling can also be connected to a worker. Login and data identification are stored for problem-checking, but no direct documentation tool will be used.

Risk modelling to assess likelihood and magnitude: A collision between the robot arm and a worker can cause physical injury under certain conditions, but ISO/TS 15066 mandates that collisions must remain under a threshold value. Malfunctions in the control system may restrict access to data sets and models, but only within the cell. Workers may experience mild psychological stress while adjusting to collaborative work. Worst-case scenarios include control and MR subsystem failures that interrupt automated production. Manual guiding of the robot or focused handling by the worker can solve these problems. Operational risks include decreased productivity, increased stock levels, and workers feeling unsafe or halting work. Some examples of ethical hazards and harms can be found in Table 1.

System rollback, error identification and correction: For automation failures, manual processes must be applied while reporting to the System Integrator. Serious problems may involve the System Developer, too. The process for error identification and correction should be defined, along with a list of potential errors and remedies. Project leaders can make decisions on necessary design changes.

Table 1. Ethical hazard identification of the cobot cell (excerpts).

Tasks/Event	*Ethical Issue*	*E. Hazard*	*E. Risk*	*E. Harm*	*Mitigation*	*Case in phase*
High job demands	Psychological	Tension in worker	High	Stress, addiction	Reschedule tasks in assembly cell	Operation phase
Handling video of activities	Privacy	Unauthorized access to data	Low	Stress, distrust	Implement security tools	Design phase
HR collision avoidance	Psychological	Fear of collision with robot	High	Fear, stress, uncertainty	Visual check on predicted robot path	Operation phase

4.2 Phase 2: Development

Preventing manipulation of data or model outputs: Theoretically false data elements can be insert into the database, but the possible intentional disturbing motions of the worker can force continuous re-planning of the gluing path. So, both direct cyber- and physical attacks can happen but with very low probability. In worse case scenarios, infected data can cause instability and a permanent stop of the robot arm. In this case the data sets must be checked and reloaded with new cleaned data. However, all this incurs financial losses.

Performance monitoring: This process includes the definition of procedures and reporting processes for system performance and post deployment monitoring and identification of responsible persons for implementation of these procedures.

Planned systems verification: Objective evidence should prove that design requirements are met. The results of the system will be verified continuously by the worker through the 3D visual representation and automatically by the built-in tolerance limit monitoring. If the predicted value of the gluing line is out of tolerance the automatic monitoring system invokes real time re-planning. The worker can also initiate re-planning if the gluing line looks incorrect or there is a danger of collision in the visualization. Decisions will be made automatically or by the worker, on the fly.

Third-party system audits will be carried out on requests. The audits will follow the predefined structure of audit reports in each professional field. The reports will be collected and stored in the project documentation repository.

4.3 Phase 3: Deployment

Here, functional testing is performed at each system components to ensure correct operation with reliable data sources. Deviations in performance are identified and fixed. Rollback processes remain the same. Continuous dialogue with responsible mission users is necessary for additional testing and fine-tuning proposals. Post-deployment monitoring provides useful information on system operation and gives directions for further development.

Task definition check: Tracing of the AI components until test deployment is essential. The System Integrator may become the system's vendor, requiring technical, legal, documentation and organizational changes. Operational requirement changes are tracked in log files. Quality check programs and internal auditors evaluate changes to data inputs and outputs for correct and optimal results.

Protection of input data: New data can be managed as old data with defined data interpretation capabilities. Adjustments in data preparation are logged and linked to new workpiece and task data in the database. Managing this process is responsibility of System Developer and System Integrator, according to the access roles defined in Phase 1.

Capability check: The System Developer post-examines models for consistent path generation in automatic gluing. Model performance changes are tracked as new items. Periodic capability reviews are managed by the System Integrator through functional testing.

4.4 Phase 4: Operation

To prevent unexpected actions in an industrial AI system, regular tests are necessary at user and developer sites. The essential procedures are as follows.

Performance deviations: Procedures for monitoring and reporting system performance are defined. Deviations and values are monitored, and gluing line coordinates are stored in a database. Deviations trigger output analysis and setup modification. Serious cases prompt control model checks and functional tests.

Rollback mechanism: The cobot system comprises hardware and software components, and the rollback plan considers monitoring of physical parts and control commands. Malfunctions in one component can affect the other. Appointed staff manage the rollback system both at the End User and the System Integrator.

Risk assessment: System risks must be identified, countered, and documented with details (Liu et al. 2020). Model errors can specifically lower user trust. The System Developer creates a risk model, which is adjusted based on the End User's environment.

Post-development monitoring and auditing: Continuous monitoring and auditing of the AI system's performance is crucial, just like recording data of the operation of the cell. Regular tests, with industrial and test data sets, are done to recover and prevent potential errors and malfunctions. The system's capability is evaluated in its post-deployment phase too, on a regular basis,

5. DISCUSSION AND CONCLUSIONS

The AI community is in common that ethical aspects have to be taken into consideration during the life cycle of AI systems as this is the guarantee that users will trust and use routinely these systems. The operation and the results of the systems should be explainable and transparent, warranting at the same time privacy and security of all their recorded data. Guidelines that are based on ethical principles have been created, but numerous analysis and reviews state that it is not possible to develop general guidelines that can fit all kinds of application. Hence, these guidelines have to be adapted to individual cases, just like ethical norms to individual situations. An additional problem is that today technical aspects often overwrite the ethical ones.

The paper introduced a workflow for ethics oriented design of a cobot assembly cell integrating different guidelines for ethical development. The architecture and components of the workcell has been constructed and selected to increase the worker's trust in the overall system. This was facilitated by giving declarative explanation of the cobot's motion plan and by showing with 3D streaming the planned arm path in real time. Additional factors for trust building are the predictive collision avoidance system and the automatic warning in case of leaving the glue-line tolerance limits by the gluing head.

The next step is the refinement of ethical design workflow in order to develop a routine process for ethical design of cobot cells and extending the explanation possibility for machine learning controlled robots as well (Liu et al. 2022). The ethics by design approach has to go beyond the existing guidelines as different fields, environments, and user demands need different solutions from the developers. However, applied in a generic ethics conscious workflow, current technologies available in cyber-physical production systems can well support such developments.

ACKNOWLEDGEMENT

This research has been supported by the ED_18-2-2018-0006 NRDIO grant on "Research on prime exploitation of the potential provided by the industrial digitalization" and the European Commission by the H2020 EPIC grant No. 739592.

REFERENCES

Askell, A., Brundage, M., & Hadfield, G. (2019). The role of cooperation in responsible AI development. *arXiv preprint arXiv:1907.04534*.

Bietti, E. (2021). From Ethics Washing to Ethics Bashing: A moral philosophy view on tech ethics. *Journal of Social Computing*, *2*(3), 266-283.

BS, (2016). *Robots and Robotic Devices: Guide to the Ethical Design and Application of Robots and Robotic Systems*. British Standard BS8611, 2016.

Cominelli, L., Feri, F., Garofalo, R., Giannetti, C., et al. (2021). Promises and trust in human–robot interaction. *Scientific Reports*, 11(1), 1-14.

Corrêa, N. K., Galvão, C., Santos, J. W., et al. (2022). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. arXiv preprint arXiv:2206.11922.

Dignum, V. (2018). Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Inform. Technology*, 20(1), 1-3.

Dunnmon, J., Goodman, B., Kirechi, P., Smith, C. & van Deusen, A. (2022) *Responsible AI guidelines in practice*. Defense Innovation Unit, https://www.diu.mil/responsible-ai-guidelines (accessed at 27.03.2023)

EU Commission, (2021). *Ethics By Design and Ethics of Use Approaches for Artificial Intelligence*, European Commission DG Research & Innovation RTD.03.001- Research Ethics and Integrity Sector Version 1.0, 25 November 2021.

Erdős, G., Paniti, I., & Tipary, B. (2020). Transformation of robotic workcells to digital twins. *CIRP Annals*, 69(1), 149-152.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, (2020-1).

Gillespie, N., Lockey, S., Curtis, C. (2021). *Trust in Artificial Intelligence: A Five Country Study*. The University of Queensland and KPMG Australia. doi: 10.14264/e34bfa3

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, 30(1), 99-120.

IEEE, (2022). *IEEE Standard for Transparency of Autonomous Systems*. IEEE Std 7001-2021.

Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.

Kemény, Z., Váncza, J., Wang, L., & Wang, X. V. (2021). Human–robot collaboration in manufacturing: A multi-agent view. In *Advanced Human-Robot Collaboration in Manufacturing*, pp. 3-41, Springer, Cham.

Kok, B. C., & Soh, H. (2020). Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1(4), 297-309.

Kuipers, B. (2020). Perspectives on ethics of AI. In *The Oxford Handbook of Ethics of AI* (p. 421). Oxford University Press.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80.

Leitão, P., & Karnouskos, S. (2022). The emergence of ethics engineering in industrial cyber-physical systems. In 2022 *IEEE 5th International Conference on Industrial Cyber-Physical Systems* (ICPS), pp. 1-6.

Leslie, D. (2019). *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. https://doi.org/10.5281/zenodo.3240529.

Liu, Z., Wang, X., Cai, Y., Xu, W., Liu, Q., Zhou, Z., & Pham, D. T. (2020). Dynamic risk assessment and active response strategy for industrial human-robot collaboration. *Computers & Industrial Engineering*, 141, 106302.

Liu, Z., Liu, Q., Xu, W., Wang, L., & Zhou, Z. (2022). Robot learning towards smart robotic manufacturing: A review. *Robotics and Computer-Integrated Manufacturing*, 77, 102360.

Luhmann, N. (2018). *Trust and power*. John Wiley & Sons.

Maurice, P., Padois, V., Measson, Y., & Bidaud, P. (2017). Human-oriented design of collaborative robots. *International Journal of Industrial Ergonomics*, 57, 88-102.

Mezgár, I., & Váncza, J. (2022). From ethics to standards–A path via responsible AI to cyber-physical production systems. *Annual Reviews in Control*, 53, 391-404.

Moencks, M., Roth, E., Bohné, T., Romero, D., & Stahre, J. (2022). Augmented Workforce Canvas: A management tool for guiding human-centric, value-driven human-technology integration in industry. *Computers & Industrial Engineering*, 163, 107803.

Müller, V. C., & Zalta, E. (2021). Ethics of artificial intelligence and robotics. In *The Stanford Encyclopedia of Philosophy*. Stanford, CA, (Summer 2021 Edition), https://plato.stanford.edu/archives/win2020/entries/ethics-ai, (accessed at 27.03.2023).

Paniti, I., Nacsa, J., Kovács, P., & Szűr, D. (2021). Human-robot collision predictor for flexible assembly. *ACTA IMEKO*, 10(3), 72-80.

Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—Two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, *1*(1), 34-47.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393-404.

Safety Toolkit by COVR, https://www.safearoundrobots.com/toolkit/, (accessed at 27.03.2023)

Sorell, T. (2022). Cobots, "co-operation" and the replacement of human skill. *Ethics and Information Technology*, 24(4), 1-12.

Wang, L., Gao, R., Váncza, J., Krüger, J., Wang, X. V., Makris, S., & Chryssolouris, G. (2019). Symbiotic human-robot collaborative assembly. *CIRP Annals*, 68(2), 701-726.

Zahorán, L., & Kovács, A. (2022). ProSeqqo: A generic solver for process planning and sequencing in industrial robotics. *Robotics and Computer-Integrated Manufacturing*, 78, 102387.