This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 826304

THEME [SC1-DTH-03-2018]

Adaptive smart working and living environments supporting active and healthy ageing

BIONIC

body information on an intelligent chip

„Personalized Body Sensor Networks with Built-In Intelligence for Real-Time Risk Assessment and Coaching of Ageing workers, in all types of working and living environments"

Project Reference No	826304
Deliverable	D 9.5
Workpackage	WP9: Ethics strategy
Nature	D (Deliverable)
Dissemination Level	P (Public)
Date	03/10/2019
Status	Final
Editor(s)	Didier Stricker (DFKI) Eric Thil (DFKI)
Document Description	This document presents the ethical principes and strategies of the BIONIC project

CONTENTS

LIST OF TABLES

LIST OF FIGURES

No table of figures entries found.

Table 1: List of Abbreviations

Term / Abbreviation	Definition
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz
TUK	Technische Universität Kaiserslautern
IBV	Instituto de Biomecánica de Valencia
RRD	Roessing Research and Development
UPRC	University of Piraeus Research Center
IAW	Interactive Wear AG
HIKE	Hypecliq IKE
AC	Acciona Construcción
MTU	Rolls-Royce Power Systems - MTU
BAUA	Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
FLC	Fundación Laboral de la Construcción
DPO	Data Protection Officer
AI	Artificial Intelligence
CA	Consortium Agreement
WP	Work Package
GA	General Assembly

1 EXECUTIVE SUMMARY

The deliverable D 9.5 with the title "Ethics strategy" presents the project strategy related to the ethics issues and a general guideline on this topic.

The *Ethic strategy* document constitutes a fundamental task for European Research Projects since it allows to better design the planned system so that it guaranties privacy issues, has a positive impact on human wellbeing, on the society and on the environment.

To this end, the following activities are planned or have already been carried out right from the start of the BIONIC project:

1. General Assembly: A general meeting at which progress, problems and other administrative matters shall be discussed will follow each plenary meeting of the consortium. This meeting is called GA. If voting is necessary for a decision, each partner, regardless of the number of his members, has only one vote. The adopted voting procedure (paper, show of hands, etc.) will be discussed at the beginning of the meeting. In the eventuality of a tie, it is the host institute (as far as we are concerned, the DFKI) that will be responsible for deciding.
2. Data Protection Officer, DPO: The consortium has elected a DPO (Data Protection Officer) according to the proposal of the coordinating institution (DFKI). He will be in charge of personal data protection and GDPR compliance. All consortium members will get his contact details. For further information, see the document corresponding to the WP10, Deliverable 10.4 "POPD-Requirement No. 6".
3. Ethic committee: The GA will have to vote on an ethics committee comprising of three experts who will deal with all ethic issues of the project. If needed or asked, this committee will report potential problems and suggested solutions to the consortium during the GA. It would be preferable for the committee to be composed of people from different backgrounds in order to have an overview of the ethical issues related to the project on one hand and to avoid conflicts of interest on the other.

The Ethic Strategy of the consortium follows the guidelines proposed in the document of the European Commission "*Ethics guidelines for trustworthy AI"*.

2 ETHICAL PRINCIPLES AND POLICY

2.1 FRAMEWORK OVERVIEW

European sources present a document that can be consulted at the following address (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf), setting out the strategic framework for trustworthy artificial intelligence:

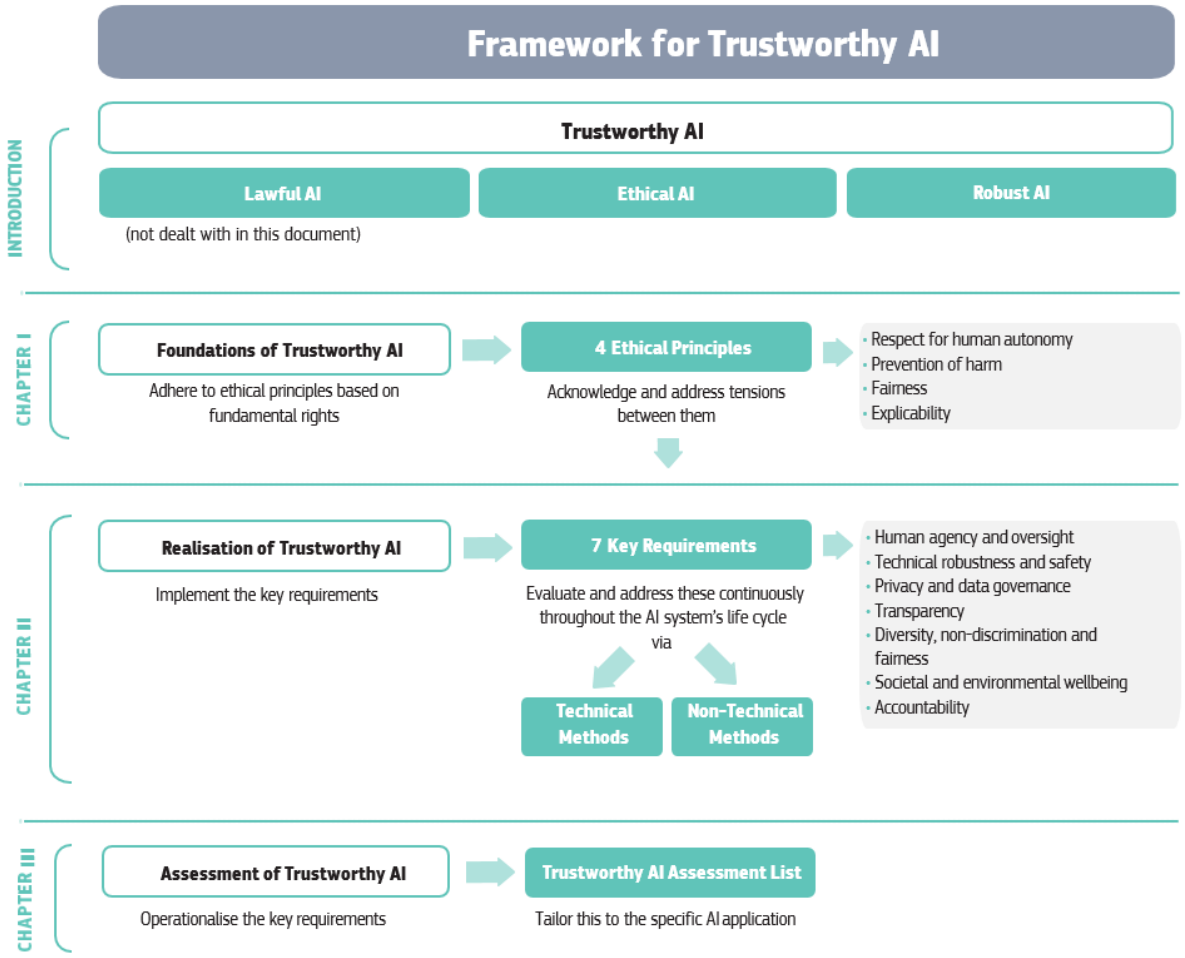

Figure 1: The Guidelines as a framework for Trustworthy AI

2.2 THE 4 ETHICAL PRINCIPLES

In view of the European Union's policy on research and the use of Artificial Intelligence, the BIONIC project must establish an ethical strategy that respects the rights of the data subjects and prevents the potential risks generated by the project.

In accordance with the document *Ethics guidelines for trustworthy AI*, and in order to create a trustworthy AI, the project must take into account the laws in force in each of the consortium partner countries to identify possible paradoxes and propose solutions. This duty is included in WP10 (10.3) and is a separate document.

Once completed, the project must ensure respect to ethical principles based on fundamental rights based on four rules:

1. Respect for human autonomy: any project must guarantee respect for the freedom and autonomy of human beings. It is crucial to provide help in the daily life or in the world of work.
2. Prevention of harm: AI systems should not harm human beings in any way.
3. Fairness: The deployment of such systems must be fair and help to combat existing injustices, as well as respect the principle of proportionality between means and ends, and consider carefully how to balance competing interests and objectives.
4. Explicability: finally, the principle of transparency must be respected to ensure a good understanding of the systems, their impact and the possible consequences of their implementation. This guarantees the freedom of the human being, thus closing the loop.

2.3 FRAMEWORK OVERVIEW

In order to concretize the ethical principles, the project must meet seven essential requirements decided by EU:

Figure 2: Interrelationship of the seven requirements: all are of equal importance, support each other, and should be implemented and evaluated throughout the AI system's lifecycle

2.4 THE 7 ETHICAL REQUIREMENTS

1. Human agency and oversight: including fundamental rights, human agency and human oversight.
2. Technical robustness and safety: including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility.
3. Privacy and data governance: including respect for privacy, quality and integrity of data, and access to data.
4. Transparency: including traceability, explainability and communication.

5. **Diversity, non-discrimination and fairness**: including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation.
6. **Societal and environmental wellbeing**: including sustainability and environmental friendliness, social impact, society and democracy.
7. **Accountability**: including auditability, minimisation and reporting of negative impact, trade-offs and redress.

Once these seven requirements have been met, the product produced by the BIONIC project can be considered ethical.

3 ETHICAL SELF-ASSESSMENT

3.1 SOURCES

The procedure employed to perform self-assessment comes from a document provided by the EU, which can be consulted at the following address:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf

3.2 RESULTS OF THE SELF-ASSESSMENT

The results of the self-assessment are listed in the table hereunder. It came out that the first issue is related to the protection of personal data as BIONIC monitors individuals over a long period and with a very accurate way.

Three colours have been used to make the document easier to read. It is quite obvious that the BIONIC project is in line with the European requirements for a trustworthy AI.

	Yes	No	No relevance	
Questions				
111	D10.3			
112	x			
112 a		x		
112b	x			
112c			x	
121			x	
121a			x	
121b			x	
131			x	
131a			x	
131b			x	
131c			x	
131d			x	
132			x	
132a			x	

132b			x	
211			x	
211a			x	
212			x	
213		x		
214			x	
221			x	
222			x	
222a			x	
222b			x	
222c			x	
222d			x	
223			x	
223a			x	
223b			x	
223c			x	
224	x		x	
224a	x			
224b	x			
231	x			
231a	x			
231b	x			
231c	x			
232			x	
233		x		
234	x			
241	x			
241a	x			
241b	x			
241c	x			
241d	x			
241e	x			
311	x			
312	x			
314	x			
315	x			
316	D10.4			
321	x			
322	x			
323	x			
324	x			

331	D10.1			
331a	x			
331b	x			
331c	x			
411	x			
411a	x			
411b	x			
411c	x			
421	x			
421a	x			
421b	x			
421c	x			
421d	x			
422	x			
423	x			
423a	x			
423b	x			
423c	x			
431	x			
432	x			
432a	x			
432b	x			
432c	x			
432d	x			
433	x			
433a	x			
433b	x			
434	x			
434a	x			
434b	x			
511	x			
511a	x			
511b	x			
511c	x			
511d	x			
512	x			
512a	x			
512b	x			
513	x			
513a	x			
513b	x			
514	x			

514a	x			
514b	x			
514c	x			
521	x			
521a	x			
521b	x			
521c	x			
522	x			
522a	x			
522b	x			
522c	x			
531	x			
532	x			
611			x	
612			x	
621			x	
621a			x	
621b			x	
622	x			
631	x			
711	x			
712	x			
721	x			
722	x			
722a	x			
722b	x			
722c	x			
723	x			
724	ethic board			
731			x	
732			x	
741		x		
742		x		

One of the main shortcomings of the BIONIC project, which will be studied in the coming months, remains the problem of redressing the artificial system in case of major problems.

4 BIONIC ETHICS STRATEGY

4.1 OVERALL STRATEGY

This strategy is used to make trustworthy AI systems which is explained in detail in Deliverable 1.5 "High-level System Architecture and Technical Specifications" with iterative prototype definition. This fits in perfectly with the European perspective provided by the above-mentioned document and visible in the following figure:

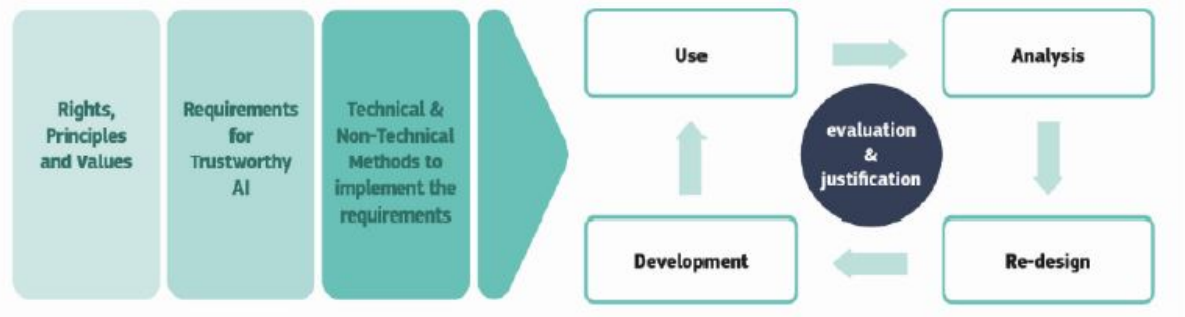

Figure 3: Realising Trustworthy AI throughout the system's entire life cycle

4.2 IMPLEMENTED POLICY

4.2.1 Data Protection

A DPO was elected by the absolute majority of the GA of 04/06/2019 and will be responsible for the BIONIC project. For more information, see Deliverable 10.4 "POPD – Requirement No.6"

Data processing in the BIONIC system will be in accordance with the requirements set by GDPR. This can be verified through the following three deliverables.

- Deliverable 1.4 "Privacy and Data Protection Framework"
- Deliverable 6.4 "Data Protection (security and privacy) Measures"
- Deliverable 6.5 "Data Protection Policies"

4.3 LEGISLATION

The deliverable 10.3 "POPD – Requirement No.5" ensures that the consortium leader of the BIONIC project has checked the legality of the partner's countries to avoid any possibility of legal problem. DFKI has therefore asked each of the partners to find out about the legislation in force in their country and its concordance with European directives.

4.4 ETHIC ADVISORY BOARD

An ethic advisory board was established under the guidance of Prof. Dr. Karen Joist from Technische Universität Kaiserslautern. It also includes Dr. Lada Tmotijevic from University of Surrey (UK).

4.5 TRUSTWORTHY AI ASSESSMENT LIST

The AI Ethics Guidelines list from EU has been checked and attached in the appendix of this document, see section 10.1 in Annex.

The results presented in section 3.2 of this document have been divided in three categories (yes, no and no relevance) which can be found in different colors in the relative table of section 3.2.

4.6 CONCLUSION

4.6.1 Consequences for the BIONIC Project

The BIONIC project consists of micro body-sensors, embedded in the clothing of workers, which aim to measure the physical strain and other vital signs of the workers.

As these sensors only serve as monitoring devices without any ability of processing or any other form of intelligence, they are in no way a brake on human autonomy. Nevertheless, the solutions implemented by the project (and developed by the RRD partner) challenge the notion of autonomy (this point will be discussed later).

Embedded into clothing, the sensors do not directly threat the privacy / autonomy of the people involved in the project. Nevertheless, it is important to note the long-term effects of the radiation due to the the wifi or/and bluetooth communication channels.

The aim of the project is to provide support for the working environment, first within selected partners (AC and MTU), and then to extend it to the labour market so that all can benefit from it. In this way, BIONIC is against a potential injustice and for fairness in the professional world.

Participants (AC and MTU employees) will agree to wear suits with micro bodysensors to measure their strain and several other vital signs. Each employee (data subject) will have the option to share or not share her/his personal data (raw data from the body sensors or processed data) with the doctors and other medical authorities involved in the project (BAUA and MTU). Questionnaires and explanatory notes, as well as personalised interviews, will make it possible to access and understand the ins and outs of this project (see D10.1). The presence of a DPO will ensure the smooth progress of the project and will allow all partners / participants to resolve any questions they deem necessary or useful.

Finally, the collection of data will be treated separately (WP6, D6.1, D6.2, D6.3, D6.4, D6.5), given the increasing importance of data collection in European policy in recent years.

5 ETHICS POLICY

5.1 THE BIONIC PROPOSITIONS

The BIONIC project guarantees the maintenance of the fundamental rights of each participant, which is in line with the first European requirement. In addition, the project involves three industrial partners (AC, MTU, and FLC) who will test the BIONIC prototypes on their teams (the latter will have freely identified themselves and will have been properly informed according to the principle of transparency below). Since the prototype incorporates only sensors embedded in clothing, they are subject to human control and thus also comply with the first requirement. Also, the sensors used for the BIONIC project do not produce meaningful data by themselves. They simply capture movements and some other vital signs in the form of raw data that needs further processing for producing some meaningful result. Human "oversight" will be therefore necessary something that again satisfies the first requirement.

The noun “robustness” must be understood here in the sense of preventing any damage. Systems developed for data collection will have to be tested to ensure their security and resistance to external attacks to prevent damage of any kind. Again, since the BIONIC prototype only consists of body

sensors and the related software, it does not put in danger the safety of participants and does not risk being misused, which is the second requirement. Similarly, since the sensors only produce data that can be only interpreted by the medical authorities, in the case that the participants allow access to the data, the system itself cannot judge the good or poor quality of the data. At least, In the event of application/system errors, the designers can double-check the software and/or the employed infrastructure in order to make it reproducible and reliable, as imposed by the second requirement.

Guaranteeing privacy protection and data sovereignty is one of the essential axes of the BIONIC project and also the third requirement. The data collected during the final phase of the project will be stored in a secure storage area and will be made accessible only to authorised users. A series of security measures and privacy enhancing technologies will be employed in order to protect all data during the storage or/and transmission through the various communication channels, thus ensuring their quality and integrity. It is the participant's choice to disclose them or not. Nevertheless, it will be necessary to ensure equity of opportunity for participants by giving them the same opportunity to access their data regardless of the social background from which they come. That would fulfil the third requirement.

In order to ensure greater transparency and traceability, project partners will produce documents with the details of the system architecture, the types of data collected, the way they are processed and the people (roles) that are authorized to access them. All these documents will be available through a common document repository. In addition, the description of all use cases that will be implemented during the project, along with their results, will be present, as annexes to the deliverables submitted. The project, its purpose and the data collected will be interpreted by specialists and properly explained to volunteers in the final testing phase with medical authorities from AC, MTU and BAUA. In addition, the DPO will be freely accessible to each of the users who will have their contact details in advance. Explainability as Communication will be respected which fulfils requirement number 4.

As the project aims to address all types of fields and populations (although with priority given to the workers' field, at least in the final phase), it will be necessary to ensure the recruitment of people from different contexts, cultures and disciplines to guarantee both diversity and non-discrimination. The use of software programs managing participants' data will be simple in order to create a universally equitable and accessible AI, especially for future marketing. All participants will be consulted and informed of the process. BIONIC is above all an exchange between developers and users, so the project complies with requirement 5.

In view of the growing importance that the European Union pays to the environment, it will be necessary to take into account the impact of the product developed in the project. Since the aim is to integrate electrical circuits into clothing, one of the major challenges is to orchestrate as little waste as possible. The energy required to operate the data collection system will work as follows: some sensor nodes, consisting inertial- and environmental sensors, will be integrated into the clothes and these sensor nodes will be connected to the application edge device. We call this Body Sensor Network (BSN) which will be powered through 5V-batteries. Data transmission to the application devices, such as mobile phones or PCs, will be performed through wireless communication such as Wifi or Bluetooth. Data will be and must only be stored in a secure Cloud storage that has to be decided by MTU and AC. The storage Cloud was developed by Hypercliq and will be use only during the running time of the BIONIC project. It should be taken into account that the tests involve a porous boundary between private and professional life. By using the game application developed in the BIONIC project, work affects the private life of participants. Therefore, to avoid such confrontation, strategies will have to be proposed. Created for public health purposes, the sensors of the BIONIC project should not have

a societal impact or cause problems with the freedom policy in place. Thus the requirements of point 6 are also respected.

The evaluation of the various processes and experiments listed above will have to be validated by an ethics committee and by the DPO. In order not to fall into arbitrary judgment, this committee will be composed of specialists who are not part of the consortium or who are closely linked to its members. In the event that problems arise during the project, they will be submitted to the ethics committee and the DPO. As it is difficult to predict the potential problems during the project's life, the implementation of solutions will be done on a case-by-case basis, but always listed and documented, so the last requirement will be also respected.

6 DATA POLICY

AI systems must guarantee privacy and data protection throughout the system's entire lifecycle. To allow individuals to trust the data gathering and processing process, it must be ensured that data collected about them will not be used for illegal or unfairly discriminate purposes.

Processes and data used must be tested and documented at each step such as planning, implementation, training, testing and deployment. Each partner of the consortium will be responsible to document its own activities. They will also ensure that they comply with the laws intrinsic to their respective countries. These files will be sent to the DPO, who will check them, classify them and report them to the ethics committee.

Data access policies should also be put in place. These policies will outline who can access data, in which way (read/write/erase) and under which circumstances. This policy should be approved by the DPO. It is obvious that everyone's access to their own data must be made possible by the project.

In cases where data must travel from one country to another, in order to be examined by the project partners, they will be made anonymous or pseudonymized in order to avoid any potential danger.

7 MEDICAL APPROACH

In agreement with the medical authorities involved in the project, a medical protocol must be created by the host institution. DFKI has therefore called upon the medical entities as well as the legal services of the partners involved in order to create an informative model and another model of consent.

The medical trials inherent in the project must be developed by physicians and submitted to the ethics committee responsible for ensuring that the laws and the integrity/safety of the participants/workers are respected. This collaboration must be based on voluntary participation. Moreover, in a concern for equity and non-discrimination, it will be necessary to ensure the heterogeneity of the subjects. The choice will therefore have to be made by taking into account parameters like gender, age, social background, ethnicity, mobility etc. The doctors involved in the project should clearly explain the various processes to the participants (preferably individually). They will also learn about the needs of the various participants in order to adapt the project to their requirements. All personal data will be made available, if requested, to the data subject. Each of the discoveries concerning a participant should be the subject of an interview and an appropriate protocol for treatment if necessary which is part of the Deliverable D10.1.

The medical protocol must be voted first on by the Ethic Advisory Board (see Proposal), then by the consortium. The ethics committee will have to verify that there will be no danger for either the physical or moral integrity or the personal and cultural identities of the participants. It will also ensure that individual freedom has been preserved, laws and human rights respected. Finally, at the end of the project, the committee, if necessary, will submit a report on the environmental and social impacts. In addition, the latter will collect each document submitted to the consortium in order to ensure transparency.

The participants in the BIONIC experiments will be on a volunteering basis.. There will be great effort to achieve, as much as possible, maximum heterogeneity (in terms of gender, age, ethnicity, social background, and mobility) of the workers participating in the BIONIC experiments.

Physicians involved in the project should clearly explain the various data collection and processing processes to the participants (preferably individually). They will also record any special needs / requirements of the participants.

8 POTENTIAL PROBLEMS

The impact on the health of participants should be studied once the BIONIC use cases have been launched. A bond of trust must be therefore established between the medical authorities and the participants. If certain risks exist (such as hazardous discoveries), the participants, through a previously completed form, will express their wish and agreement to be informed. Medical authorities, on the other hand, are committed to make correct diagnosis and refer the people to the appropriate specialists. Again, it is difficult to predict potential discoveries and thus they will need to be assessed on a case-by-case basis. Last but not least, it will be also necessary to ensure that the discovery of a health problem does not lead to the participant being fired from the company, which would be unethical and also contrary to the spirit of the BIONIC project. If necessary, lawyers will be consulted.

Another remaining problem is that of non-discrimination. Although the BIONIC project is in full agreement with this requirement, it may be difficult to take a sufficiently large sample of participants. Unfortunately, the working environment (for the BIONIC use cases) has some similarities with the clichés and is mainly composed of male subjects. If we manage to find some female participants, they will not correspond to the 50% of the entire sample, which will undoubtedly lead to a problem regarding the anonymization of personal data. Similarly, it will be very difficult to represent all ethnic groups. Therefore, the BIONIC project, although it will try its best, won't be able to be perfectly equalized on those questions. Nevertheless, all the categories will be represented.

In order to protect and also improve the health of the participants and to offer them adequate exercises to remedy their work-related disorders, the RRD partner has developed a gamification software. Although its development and usefulness has been approved in CA, the fact remains that the latter raises certain dangers. Initially, the participant may, through this "game", be infantilized, which would be contrary to the first ethical principle and the first ethical condition mentioned above. It is therefore important to ensure that the program is adapted to adults and does not treat them like children. Secondly, there is still a risk of addiction, linked to the phenomenon of "gambling", which may lead to solitary and marginal behaviour. It will also be necessary to ensure that participants are not able to "play" on a continuous basis, even if their health is improved. Finally, in the case of future commercialization, it will be necessary to ensure that no advertising alters the quality and purpose of the program.

A final point to be clarified is the equity of the programs used in the BIONC project. Indeed, computerization implies that each participant has a computer tool allowing him/her to receive and consult his/her personal data, as well as to be able to "play" to improve his/her health through the gamification program. However, if the participants come from a disadvantaged background, a solution must be found to restore equity of opportunity and equality for all. For example, a mobile phone whose purpose will only be to receive the data may be provided by the company to employees and therefore reused once the data has been erased or the employees have left.

9 CONCLUSION

In order to ensure the smooth implementation of the BIONIC project, it is essential to respect the various points set out above. The DPO and the ethics committee will be there to ensure that the consortium's altruism is in line with the European objectives of respect for human beings and the development of systems to improve their lives. This document will also be sent to all partners involved in the BIONIC project in order to be unanimously voted on by the CA.

10 ANNEXES

10.1 QUESTIONNAIRE FOR ETHIC VERIFICATION

TRUSTWORTHY AI ASSESSMENT LIST (PILOT VERSION)

1. Human agency and oversight

1.1 Fundamental rights:

111 Did you carry out a fundamental rights impact assessment where there could be a negative impact on fundamental rights? Did you identify and document potential trade-offs made between the different principles and rights?

112 Does the AI system interact with decisions by human (end) users (e.g. recommended actions or decisions to take, presenting of options)?

- Could the AI system affect human autonomy by interfering with the (end) user's decision-making process in an unintended way?
- Did you consider whether the AI system should communicate to (end) users that a decision, content, advice or outcome is the result of an algorithmic decision?
- In case of a chat bot or other conversational system, are the human end users made aware that they are interacting with a non-human agent?

1.2 Human agency:

121 Is the AI system implemented in work and labour process? If so, did you consider the task allocation between the AI system and humans for meaningful interactions and appropriate human oversight and control?

- Does the AI system enhance or augment human capabilities?
- Did you take safeguards to prevent overconfidence in or overreliance on the AI system for work processes?

1.3 Human oversight:

131 Did you consider the appropriate level of human control for the particular AI system and use case?

- Can you describe the level of human control or involvement?
- Who is the "human in control" and what are the moments or tools for human intervention?
- Did you put in place mechanisms and measures to ensure human control or oversight?
- Did you take any measures to enable audit and to remedy issues related to governing AI autonomy?

132 Is there is a self-learning or autonomous AI system or use case? If so, did you put in place more specific mechanisms of control and oversight?

- Which detection and response mechanisms did you establish to assess whether something could go wrong?
- Did you ensure a stop button or procedure to safely abort an operation where needed? Does this procedure abort the process entirely, in part, or delegate control to a human?

2. Technical robustness and safety

2.1 Resilience to attack and security:

211 Did you assess potential forms of attacks to which the AI system could be vulnerable?

- Did you consider different types and natures of vulnerabilities, such as data pollution, physical infrastructure, cyber-attacks?

212 Did you put measures or systems in place to ensure the integrity and resilience of the AI system against potential attacks?

213 Did you verify how your system behaves in unexpected situations and environments?

214 Did you consider to what degree your system could be dual-use? If so, did you take suitable preventative measures against this case (including for instance not publishing the research or deploying the system)?

2.2 Fallback plan and general safety:

221 Did you ensure that your system has a sufficient fallback plan if it encounters adversarial attacks or other unexpected situations (for example technical switching procedures or asking for a human operator before proceeding) ?

222 Did you consider the level of risk raised by the AI system in this specific use case?

- Did you put any process in place to measure and assess risks and safety?
- Did you provide the necessary information in case of a risk for human physical integrity?
- Did you consider an insurance policy to deal with potential damage from the AI system?
- Did you identify potential safety risks of (other) foreseeable uses of the technology, including accidental or malicious misuse? Is there a plan to mitigate or manage these risks?

223 Did you assess whether there is a probable chance that the AI system may cause damage or harm to users or third parties? Did you assess the likelihood, potential damage, impacted audience and severity?

- Did you consider the liability and consumer protection rules, and take them into account?
- Did you consider the potential impact or safety risk to the environment or to animals?
- Did your risk analysis include whether security or network problems such as cybersecurity hazards could pose safety risks or damage due to unintentional behaviour of the AI system?

224 Did you estimate the likely impact of a failure of your AI system when it provides wrong results, becomes unavailable, or provides societally unacceptable results (for example discrimination)?

- Did you define thresholds and did you put governance procedures in place to trigger alternative/fallback plans?
- Did you define and test fallback plans?

2.3 Accuracy

231 Did you assess what level and definition of accuracy would be required in the context of the AI system and use case?

- Did you assess how accuracy is measured and assured?
- Did you put in place measures to ensure that the data used is comprehensive and up to date?
- Did you put in place measures in place to assess whether there is a need for additional data, for example to improve accuracy or to eliminate bias?

232 Did you verify what harm would be caused if the AI system makes inaccurate predictions?

233 Did you put in place ways to measure whether your system is making an unacceptable amount of inaccurate predictions?

234 Did you put in place a series of steps to increase the system's accuracy?

2.4 Reliability and reproducibility:

241 Did you put in place a strategy to monitor and test if the AI system is meeting the goals, purposes and intended applications?

- Did you test whether specific contexts or particular conditions need to be taken into account to ensure reproducibility?
- Did you put in place verification methods to measure and ensure different aspects of the system's reliability and reproducibility?
- Did you put in place processes to describe when an AI system fails in certain types of settings?
- Did you clearly document and operationalise these processes for the testing and verification of the reliability of AI systems?
- Did you establish mechanisms of communication to assure (end-)users of the system’s reliability?

3. Privacy and data governance

3.1 Respect for privacy and data Protection:

311 Depending on the use case, did you establish a mechanism allowing others to flag issues related to privacy or data protection in the AI system’s processes of data collection (for training and operation) and data processing?

312 Did you assess the type and scope of data in your data sets (for example whether they contain personal data)?

313 Did you consider ways to develop the AI system or train the model without or with minimal use of potentially sensitive or personal data?

314 Did you build in mechanisms for notice and control over personal data depending on the use case (such as valid consent and possibility to revoke, when applicable)?

315 Did you take measures to enhance privacy, such as via encryption, anonymization and aggregation?

316 Where a Data Privacy Officer (DPO) exists, did you involve this person at an early stage in the process?

3.2 Quality and integrity of data:

321 Did you align your system with relevant standards (for example ISO, IEEE) or widely adopted protocols for daily data management and governance?

322 Did you establish oversight mechanisms for data collection, storage, processing and use?

323 Did you assess the extent to which you are in control of the quality of the external data sources used?

324 Did you put in place processes to ensure the quality and integrity of your data? Did you consider other processes? How are you verifying that your data sets have not been compromised or hacked?

3.3 Access to data:

331 What protocols, processes and procedures did you follow to manage and ensure proper data governance?

- Did you assess who can access users' data, and under what circumstances?
- Did you ensure that these persons are qualified and required to access the data, and that they have the necessary competences to understand the details of data protection policy?
- Did you ensure an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed?

4. Transparency

4.1 Traceability:

411 Did you establish measures that can ensure traceability? This could entail documenting the following methods:

- ➢ Methods used for designing and developing the algorithmic system:
- Rule-based AI systems: the method of programming or how the model was built;
- Learning-based AI systems; the method of training the algorithm, including which input data was gathered and selected, and how this occurred.
- ➢ Methods used to test and validate the algorithmic system:
- Rule-based AI systems; the scenarios or cases used in order to test and validate;
- Learning-based model: information about the data used to test and validate.
- ➢ Outcomes of the algorithmic system:

- The outcomes of or decisions taken by the algorithm, as well as potential other decisions that would result from different cases (for example, for other subgroups of users).

4.2 Explainability:

421 Did you assess:

- to what extent the decisions and hence the outcome made by the AI system can be understood?
- to what degree the system's decision influences the organisation's decision-making processes?
- why this particular system was deployed in this specific area?
- what the system's business model is (for example, how does it create value for the organisation)?

422 Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?

423 Did you design the AI system with interpretability in mind from the start?

- Did you research and try to use the simplest and most interpretable model possible for the application in question?
- Did you assess whether you can analyse your training and testing data? Can you change and update this over time?
- Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?

4.3 Communication:

431 Did you communicate to (end-)users – through a disclaimer or any other means – that they are interacting with an AI system and not with another human? Did you label your AI system as such?

432 Did you establish mechanisms to inform (end-)users on the reasons and criteria behind the AI system's outcomes?

- Did you communicate this clearly and intelligibly to the intended audience?
- Did you establish processes that consider users' feedback and use this to adapt the system?
- Did you communicate around potential or perceived risks, such as bias?
- Depending on the use case, did you consider communication and transparency towards other audiences, third parties or the general public?

433 Did you clarify the purpose of the AI system and who or what may benefit from the product/service?

- Did you specify usage scenarios for the product and clearly communicate these to ensure that it is understandable and appropriate for the intended audience?
- Depending on the use case, did you think about human psychology and potential limitations, such as risk of confusion, confirmation bias or cognitive fatigue?

434 Did you clearly communicate characteristics, limitations and potential shortcomings of the AI system?

- In case of the system's development: to whoever is deploying it into a product or service?

- In case of the system's deployment: to the (end-)user or consumer?

5. Diversity, non-discrimination and fairness

5.1 Unfair bias avoidance:

511 Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?

- Did you assess and acknowledge the possible limitations stemming from the composition of the used data sets?
- Did you consider diversity and representativeness of users in the data? Did you test for specific populations or problematic use cases?
- Did you research and use available technical tools to improve your understanding of the data, model and performance?
- Did you put in place processes to test and monitor for potential biases during the development, deployment and use phase of the system?

512 Depending on the use case, did you ensure a mechanism that allows others to flag issues related to bias, discrimination or poor performance of the AI system?

- Did you establish clear steps and ways of communicating on how and to whom such issues can be raised?
- Did you consider others, potentially indirectly affected by the AI system, in addition to the (end)-users?

513 Did you assess whether there is any possible decision variability that can occur under the same conditions?

- If so, did you consider what the possible causes of this could be?
- In case of variability, did you establish a measurement or assessment mechanism of the potential impact of such variability on fundamental rights?

514 Did you ensure an adequate working definition of “fairness” that you apply in designing AI systems?

- Is your definition commonly used? Did you consider other definitions before choosing this one?
- Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness?
- Did you establish mechanisms to ensure fairness in your AI systems? Did you consider other potential mechanisms?

5.2 Accessibility and universal design:

521 Did you ensure that the AI system accommodates a wide range of individual preferences and abilities?

- Did you assess whether the AI system usable by those with special needs or disabilities or those at risk of exclusion? How was this designed into the system and how is it verified?
- Did you ensure that information about the AI system is accessible also to users of assistive technologies?
- Did you involve or consult this community during the development phase of the AI system?

522 Did you take the impact of your AI system on the potential user audience into account?

- Did you assess whether the team involved in building the AI system is representative of your target user audience? Is it representative of the wider population, considering also of other groups who might tangentially be impacted?
- Did you assess whether there could be persons or groups who might be disproportionately affected by negative implications?
- Did you get feedback from other teams or groups that represent different backgrounds and experiences?

5.3 Stakeholder participation:

531 Did you consider a mechanism to include the participation of different stakeholders in the AI system's development and use?

532 Did you pave the way for the introduction of the AI system in your organisation by informing and involving impacted workers and their representatives in advance?

6. Societal and environmental well-being

6.1 Sustainable and environmentally friendly AI:

611 Did you establish mechanisms to measure the environmental impact of the AI system's development, deployment and use (for example the type of energy used by the data centres)?

612 Did you ensure measures to reduce the environmental impact of your AI system's life cycle?

6.2 Social impact:

621 In case the AI system interacts directly with humans:

- Did you assess whether the AI system encourages humans to develop attachment and empathy towards the system?
- Did you ensure that the AI system clearly signals that its social interaction is simulated and that it has no capacities of "understanding" and "feeling"?

622 Did you ensure that the social impacts of the AI system are well understood? For example, did you assess whether there is a risk of job loss or de-skilling of the workforce? What steps have been taken to counteract such risks?

6.3 Society and democracy:

631 Did you assess the broader societal impact of the AI system's use beyond the individual (end-)user, such as potentially indirectly affected stakeholders?

7. Accountability

7.1 Auditability:

711 Did you establish mechanisms that facilitate the system's auditability, such as ensuring traceability and logging of the AI system's processes and outcomes?

712 Did you ensure, in applications affecting fundamental rights (including safety-critical applications) that the AI system can be audited independently?

7.2 Minimising and reporting negative Impact:

721 Did you carry out a risk or impact assessment of the AI system, which takes into account different stakeholders that are (in)directly affected?

722 Did you provide training and education to help developing accountability practices?

- Which workers or branches of the team are involved? Does it go beyond the development phase?
- Do these trainings also teach the potential legal framework applicable to the AI system?
- Did you consider establishing an 'ethical AI review board' or a similar mechanism to discuss overall accountability and ethics practices, including potentially unclear grey areas?

723 Did you foresee any kind of external guidance or put in place auditing processes to oversee ethics and accountability, in addition to internal initiatives?

724 Did you establish processes for third parties (e.g. suppliers, consumers, distributors/vendors) or workers to report potential vulnerabilities, risks or biases in the AI system?

7.3 Documenting trade-offs:

731 Did you establish a mechanism to identify relevant interests and values implicated by the AI system and potential trade-offs between them?

732 How do you decide on such trade-offs? Did you ensure that the trade-off decision was documented?

7.4 Ability to redress:

741 Did you establish an adequate set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact?

742 Did you put mechanisms in place both to provide information to (end-)users/third parties about opportunities for redress?

11 REFERENCES

1. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
2. ETHICS GUIDELINES FOR TRUSTWORTHY AI, European Commission, High-Level Expert Group on Artificial Intelligence, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419