22HLT01 QUMPHY

D7 - First report from the ethics advisor to 22HLT01

Organisation name of the lead participant for the deliverable: PTB

Due date of the deliverable: 30.06.2024

Actual submission date of the deliverable: 22.05.2025

Confidentiality Status: Classified C-UE/EU-C – EU CONFIDENTIAL under the Commission Decision No 2015/444

Deliverable Cover Sheet

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EURAMET. Neither the European Union nor the granting authority can be held responsible for them.

The project has received funding from the European Partnership on Metrology, co-financed from the European Union's Horizon Europe Research and Innovation Programme and by the Participating States.

Table of contents

Document history

Date	Version	Author versions	Change description
16.05.25	0.1	Jenny Venton	First draft
21.05.25	1.0	Jenny Venton	Incorporate ethics self-assessment details and scope of report details from project coordinator.
21.05.25	1.1	Jenny Venton	Minor edits and format referenences.
22.05.25	1.2	Nando Hegemann	Minor edits and formatting to comply with submission requirements

1 Introduction

1.1 Outline and scope of the report

QUMPHY is a timely project and is laying the groundwork for safe and equitable health measurements for all through benchmark datasets and a good practice guide for photoplethysmography (PPG) machine learning (ML) algorithms. While not a clinical project, QUMPHY is in a strong position to be able to incorporate and implement ethical considerations at the early research stage of uncertainty quantification. This could act as a flagship example of how to incorporate ethical considerations at the earliest stages of health ML research. This ethics report will address the ethical considerations and implications of using uncertainty quantification (UQ) methods for machine ML applications in the analysis of PPG signals. To this end it will focus on the following key areas:

1. **Informed Consent**: Addressing the ethical considerations related to obtaining informed consent from individuals whose PPG data is used for model training and validation.
2. **Regulatory Compliance**: Ensuring that the development and deployment of ML models for medical applications comply with relevant ethical guidelines and regulatory standards, such as those outlined by the European Commission and international bodies like the WHO and ITU.
3. **Bias and Fairness**: Evaluating the potential for bias in ML models, particularly concerning demographic subgroups such as sex, age, and skin tone, to ensure equitable performance and avoid discrimination.

4. **Data Privacy and Protection**: Ensuring compliance with data protection regulations, such as the General Data Protection Regulation (GDPR), and addressing the ethical handling of personal health data collected from PPG signals.
5. **Transparency and Explainability**: Discussing the importance of transparency in ML models, including the interpretability of model predictions and the communication of uncertainty estimates to end-users, such as clinicians and patients.

The ethics report will aim to provide a comprehensive overview of the ethical considerations taken in the project and suggest measures to further ensure the responsible and ethical use of ML models in the analysis of PPG signals for medical applications.

1.2 Involvement of the ethics advisor

The approach taken in this ethics report is to: help the consortium highlight potential areas of risk; to propose ways that QUMPHY output can be delivered safely, equitably and adhering to existing guidelines and regulations where applicable; and with as wide a reach as possible. The mandate of the ethics advisor as described in the grant agreement is to "…help the consortium with ensuring that a risk assessment plan is delivered for the development and use of AI […] as well for ethically correct access to the widely available data." The ethics advisor has played an important role in the project, actively participating in the months 9 and 18 project meetings. During these meetings, the ethics advisor led dedicated sessions to discuss and address ethical issues relevant to the project. The consortium maintains a continuous exchange with the ethics advisor, ensuring that their suggestions and recommendations are integrated into the project's framework. Additionally, the consortium proactively identifies potential ethical concerns and engages in discussions with the ethics advisor to develop effective mitigation methods. This ongoing collaboration ensures that the project adheres to the highest ethical standards and addresses any emerging ethical challenges promptly.

1.3 Project summary

PPG signals are rich in information and easy to measure passively without any physical or mental limitations of the subject. ML algorithms are commonly used to extract physiological parameters from PPG signals for diagnosis, avoiding the need for complex and costly clinical review. As of today, no regulations specifying how these ML algorithms have to be applied, how their performance has to be measured or how their associated uncertainties have to be specified exist. At the core of this project stands the development of measures to quantify the uncertainties associated with ML algorithms applied to medical problems, in particular the analysis and processing of PPG signals. To achieve this the following tasks are addressed: (i) benchmark datasets are generated using publicly available in vivo, and synthetic data (ii) different ML models and uncertainty quantification (UQ) methods are used to analyse the processing of the PPG signals and specify the associated uncertainty and (iii) a good practice guide with accompanying software repository showcasing the used models, methods and benchmarks is developed and will be made publicly available.

1.4 Project ethics self-assessment

1. **Involvement of humans, human cells or tissue**
   In this project it is important to clarify that no humans, human cells, or tissues are directly involved in any experimental or data collection processes. The project focuses on the development and validation of ML models using existing datasets of PPG signals. These datasets are either publicly available or synthetically generated, ensuring that no new human data is collected specifically for this project. The analysis and methodologies employed are designed to work with pre-existing data, thereby eliminating the need for direct human participation or the use of human biological materials. This approach ensures that all research activities are conducted in compliance with ethical standards and regulations, prioritizing the use of non-invasive and pre-collected data to achieve the project's objectives.
2. **Personal data**
   In this project the usage of personal data is carefully managed to ensure compliance with ethical standards and data protection regulations. All data utilized in the project are either generated through computer simulations or derived from publicly available datasets that already conform to the EU data protection guidelines. This approach ensures that no new personal data is collected directly from individuals, thereby minimizing privacy concerns.

For datasets used within the project that have access restrictions, these are handled with the utmost care and will only be made available to the public in strict accordance with the data access guidelines of the respective dataset. This meticulous handling of data underscores the project's commitment to maintaining high ethical standards and protecting individual privacy.

3. **Animals**
   In this project no animals are involved in any experimental or data collection processes. The project's focus is solely on the development and validation of ML models using existing datasets of PPG signals, which are either publicly available or synthetically generated. The only exception to this involves the potential presence of emotional support animals, which may accompany individual consortium members in certain settings to provide comfort and relieve stress. However, these animals are not subjects of the research and are not involved in any data collection or experimental procedures. This approach ensures that the project adheres to ethical guidelines and regulations concerning animal welfare.

4. **Non-EU countries**
   This project does not utilize any facilities, materials, or experimental designs from non-EU countries. All datasets employed in the project strictly adhere to EU standards for medical data, ensuring compliance with the rigorous guidelines and regulations set forth by the European Union.

5. **Environment, health and safety**
   In this project, there is no risk of environmental damage, and consequently, no specific precautions are necessary to address such concerns. The experimental design and structure of the project have been carefully planned to ensure that they do not meet any criteria that could potentially harm the health or safety of the individuals involved. Additionally, the technologies employed in the project have been evaluated to confirm that they do not pose any undesirable side-effects that could endanger any of the persons participating in the project.

6. **Artificial intelligence**
   a. **Human agency and oversight:** All ML methods developed and investigated in the project are critically evaluated for their performance. The project's focus on uncertainty quantification for ML specifically emphasizes human agency and oversight, ensuring that humans can interpret the reliability of the outputs in conjunction with the associated uncertainties. The project is centred on establishing general procedures for uncertainty quantification and benchmarks ML models using specific, openly available, and anonymised datasets. Consequently, the results are not intended to inform decisions for direct clinical practice. The limitations of the employed models and the uncertainty quantification methods are thoroughly discussed and highlighted within the software and any publications produced by the project. It is crucial to note that no software or result of the project will attempt to exaggerate confidence or coerce, deceive, or manipulate individuals, as such actions are fundamentally opposed to the project's goals.

   b. **Privacy and data governance:** Privacy and data governance are handled with the utmost care and adherence to international, EU, and national laws regarding anonymity, ethical treatment of subjects, and data quality. All datasets utilised in the project are openly available and comply with EU standards. For datasets with restricted access, these are clearly noted, and such datasets will not be published without the explicit consent of the original data holders. To ensure transparency in data processing and augmentation for machine learning purposes, scripts used to process restricted access datasets will be made publicly available. This allows for the reproduction of the project's results while maintaining data privacy. The selection of data subsets for training purposes is conducted internally, with consultation from the ethics advisor, to prevent or reduce inherent bias against minority groups. Network

architectures, training procedures, and datasets are made publicly available to ensure reproducibility of the results. Furthermore, results are run multiple times to provide robust statistics for training procedures, ensuring the reliability and robustness of model outputs.

c. **Transparency:** Transparency is a cornerstone of the research methodology and dissemination strategy. All datasets utilized in the project are openly available, ensuring that the broader research community can access and verify the data used. For datasets with restricted access, these are clearly noted, and such datasets will not be published without the explicit consent of the original data holders. To further ensure transparency, scripts used to process restricted access datasets will be made publicly available, allowing for the reproduction of the project's results and providing insight into data processing and augmentation for machine learning purposes. Network architectures, training procedures, and datasets are made publicly available to ensure reproducibility of the results. Results are run multiple times to provide robust statistics for training procedures, ensuring the reliability and robustness of model outputs. All software developed in the project is tracked in an online repository with version control, allowing for traceability of the development process. An automatically generated online documentation is provided to simplify access to the software library, and scripts to reproduce any published results, including network architectures, training routines, and used hardware, will be provided.

d. **Fairness, diversity and non-discrimination:** All datasets used in the project adhere to the FAIR principles, ensuring that they are Findable, Accessible, Interoperable, and Reusable. A significant focus of the project is on investigating potential biases against demographic subgroups within these datasets, with specific activities aimed at examining biases related to sex, age, and skin tone. Considerable thought is given to the limitations of the produced results with respect to their application to demographic subgroups, and these concerns are discussed with the ethics advisor to ensure a comprehensive understanding of potential biases and limitations. The outputs of the project include discussions on the limitations of the data when applied to certain minority groups, highlighting areas where biases may exist or where data may not be fully representative. All machine learning applications developed in the project are trained on a variety of data to reduce bias within the constraints of the given datasets. Additionally, the uncertainty quantification methods employed are investigated for their potential to reflect biases against minority groups, ensuring that the project's findings are both robust and ethically sound.

e. **Societal and environmental well-being:** From a societal perspective, the project focuses on improving the reliability and trustworthiness of machine learning models used in medical diagnostics through the analysis of PPG signals. By enhancing the accuracy and uncertainty quantification of these models, the project aims to improve healthcare outcomes, particularly in the early detection and management of conditions such as hypertension and diabetes. This can lead to better health management and potentially reduce the burden on healthcare systems, thereby positively impacting societal well-being. Environmentally, the project does not involve any activities that could harm the environment. The research is conducted using existing datasets and computational models, which do not require physical experimentation or the use of hazardous materials. The project's reliance on data analysis and machine learning models ensures that there is no direct environmental impact, such as pollution or resource depletion. Thus, the project aligns with principles of sustainability and ethical research, ensuring that both societal and environmental

well-being are preserved and potentially enhanced.

f. **Accountability:** Accountability is realized through a structured approach that ensures responsibility, transparency, and oversight throughout the development and operation of machine learning models. The project involves multiple stakeholders, including developers, researchers, and clinicians, who are collectively responsible for the functionality and outcomes of the ML systems developed. This accountability is supported by the project's commitment to transparency, as evidenced by the open availability of datasets, network architectures, training procedures, and software repositories. These resources are made publicly accessible to ensure that the processes and methodologies used can be scrutinized and validated by external parties. Furthermore, the project incorporates rigorous oversight mechanisms, including regular consultations with an ethics advisor and stakeholder committee, to ensure that the ML applications are developed and operated in an ethically sound manner. The project's focus on uncertainty quantification also enhances accountability by providing a clear understanding of the reliability and limitations of the ML models' predictions. This comprehensive framework will help developers and operators of the ML software applications explain how and why the systems exhibit particular characteristics or result in certain outcomes, thereby fulfilling the criteria for accountability.

2 Overall feedback

2.1 Feedback and Recommendations

General comments:

- Be clear about what the end goal of the project outputs are. For example, is this for research only or is the good practice guide and code framework to be used to evaluate medical devices?
- Where existing regulations (EU AI Act for example) have been considered, indicate this alongside any outputs and specify which aspects have been considered. This may help reassure end users who want to use the developed frameworks to support an application for regulatory approval, and improve acceptability of this work.
- There are many existing frameworks for reporting on or evaluating medical ML, and criteria for healthcare datasets that QUMPHY could benefit from. These resources are a good starting point:
    - Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD+AI): https://www.bmj.com/content/385/bmj-2023-078378
    - Prediction model risk of bias assessment tool (PROBAST+AI): https://www.bmj.com/content/388/bmj-2024-082505
    - Standards for data diversity, inclusivity and generalisability (STANDING Together): https://www.datadiversity.org/
- Note that factors such as older age, lower educational status and lower household income were associated with lower odds of wearable usage (Marvasti et al., 2024) meaning there will be less representation in the data used for QUMPHY for these groups.
- Patient, practitioner and end user feedback is increasingly important at all stages of medical software development. Include any views gathered from patients or end users on what kinds of uncertainty are important and what should impact uncertainty. If not relevant or not possible then comment on this.

Intended end user and impacted people views

- Standardised approaches and benchmark datasets for PPG are very timely. The Annex states that “The implemented algorithms will support device manufacturers, clinicians and end users in deciding on medical treatment plans for certain diseases such as diabetes and hypertension”. As the outputs of QUMPHY are intended for future medical use, and include ML tools, there is a responsibility to incorporate wider views such as that of clinicians, patients, physiologists, and other experts in the intended medical application domains.
- Guidelines describe how it is critical that teams that design, develop and test ML systems reflect the diversity of end users and people impacted by the ML system, not only in terms of gender culture or age but also in terms of professional backgrounds and skill sets (Ethics Guidelines for Trustworthy AI, 2019). Describe how are these views accounted for in the development of QUMPHY’s technical work.
- Clearly describe results of any findings or information gathering of patient/end user views on what is important to capture in terms of uncertainty quanitification, and what their views are on what should impact that uncertainty.

2.2 Essential recommendations

- How will scope of application, and limitations, be communicated to end users? For example, specify which devices the method(s) was validated for and the population demographic used to develop the data (i.e. healthy/patient, age, sex, ethnicity)
- Be clear about what the benchmark datasets can and can’t be used for. Are they for ML developers generally, or specifically for PPG ML developers. As this will be publicly available resources from standard development laboratories there is a risk it could be used to validate PPG AI in settings the resources are not developed for. State the domain of each UQ method, in terms of device type, population demographic, laboratory/hospital community setting, disease state, exercise etc
- Alongside the code framework and good practice guide, be clear about what it can and can’t be used for and what scenarios it is applicable for.
- Include some comment clarifying the position of this work on pulse oximetry (PO) and the known impact of skin tone on PO performance.
- Include a summary in the deliverable outlining feedback gathered from practitioners and clinicians about what kind or type of uncertainty is useful and important for them and what they want an uncertainty value to tell them.

3 Work package and task feedback

3.1 Work package 1

- Different kinds of uncertainty are discussed. Comment on how these developed uncertainty methods will report on accuracy and uncertainty for different groups of people (i.e. age, sex, ethnicity) or different settings (i.e. wrist worn, health tracker, hospital etc), or if they will.

Task 1.1:

- State whether the ML models chosen are models that are currently used for PPG or will feasibly be used in real world applications.
- How detailed are the common evaluation frameworks and are they based on any existing evaluation frameworks? Are the common evaluation frameworks to be used internally or will this also feed into the framework being developed to be used by others in Task 2.3?

Task 1.2

- Describe how the validation of uncertainties on the prototypical datasets will translate to benchmark datasets. State any methods or assumptions made to do this.

Task 1.3

- Describe how we can verify that the uncertainty comparison on the protoypical datasets holds when translated to different datsets. For example, if the findings have used signals from healthy populations, how will the findings translate to a disease population.
- Make clear why skin tone classification is being used. For example, 'to understand whether, from an algorithm's perspective, there is a difference between signals from different skin tones'.
- The common evaluation framework being developed to distinguish in-distribution and out-of distribution data sounds very promising and perhaps a key of this kind of work. More detail on in-distribution data and the links to uncertainty this would be helpful.
- Activity 1.3.4. Understanding the dependence of model performance and uncertainty on factors such as sampling rate, pre-processing, signal or label noise, length of signals etc are important findings. Will the findings cover how these factors interact? For example, if signal length gives uncertainty x to a measurement, and pre-processing gives uncertainty y, how these combine. Is there a risk of amplifying uncertainties if an end user combined these? State what combinations (if any) the outputs are valid for.
- Activity 1.3.5. Detecting whether a sample is out of distribution is an important problem for uncertainty in medical applications. Describe how the project will determine whether epistemic uncertainty is the most appropriate for detecting out of distribution samples.
- Activity 1.3.6. Describe the limitations of the Fitzpatrick skin type scale as a skin tone classification system, and describe how this might impact the findings on skin tone in this activity.
-

3.2 Work package 2

Task 2.1

- State how the scope of the benchmark datasets will be communicated. For example, in a publication, on a website with the datasets, alongside the framework. Scope here means demographics, characteristics and disease status of the population in the benchmark datasets, as well as device types and setting where the data was collected.
- Describe how social and health system priorities have been accounted for when creating the benchmark datasets (Panch T et al, 2020)
- Comment on how the generalisabilty of the benchmark datasets to other data will be quantified.
- WFDB format is suggested for the benchmark datasets. Explain the rationale behind this in the context of datasets formats used by device manufacturers or in a clinical setting.
- Clearly state the classification or regression problem the benchmark datasets are applicable for. For example, is the atrial fibrillation (AF) benchmark dataset benchmarking uncertainty for identifying AF in a signal, or the potential for a sinus rhythm signal to become AF.

Task 2.2

- End users may have different application problems, models or datasets. In the good practice guide, include how end users can extrapolate these findings to their own work and how different application problems, models or datasets might impact the uncertainty quantification output.

Task 2.3

- As the code framework is to support industry in attaining certification and regulatory approval, describe how the code framework has taken the EU AI Act into account.
- State whether the framework is intended to be an end to end framework for evaluating medical machine learning for PPG, or for part of the process (i.e. for model development but not training data collection). If the framework is for part of the process, describe how considerations for the rest of the development pathway will be communciated to end users. When evaluating medical AI it is important to consider the wider influences and implications.

4 Other ethics issues

At this point there are no additional ethics issues that need to be addressed by the project.

5 Potential misuse of results

Potential misuse of the results and ML models in general is carefully considered and mitigated through several integrated measures. The project focuses on developing and validating ML models for the analysis of PPG signals, which are crucial for medical diagnostics and monitoring. To prevent misuse, the project emphasizes transparency and accountability by making all datasets, network architectures, training procedures, and software repositories openly available. Furthermore, this project ensure transparency in the scope, application domain and limitations of the benchmark datasets, good practice guide and code frameworks made publicly available. This transparency ensures that the processes and methodologies can be scrutinized and validated by external parties, reducing the risk of unethical use. It also ensures the outputs will not be used for purposes they are not validated for which could cause unintended harms. Additionally, the project incorporates rigorous oversight mechanisms, including regular consultations with an ethics advisor and stakeholder committee, to ensure that the ML models are developed and operated ethically. The focus on uncertainty quantification further enhances the reliability and trustworthiness of the ML models' predictions, providing a clear understanding of their limitations and potential biases. By discussing the limitations of the data and models, particularly concerning minority groups, the project aims to prevent any discriminatory or harmful applications of the results. Moreover, the project's commitment to open science practices, such as publishing results in open-access journals and presenting findings at international conferences, ensures that the broader research community can access and build upon the project's outcomes responsibly. These comprehensive measures underscore the project's dedication to ethical research practices and the responsible use of ML in medical applications.

6 Approval of ethics advisor for first ethics report

This report is approved by the project's external ethics advisor

Dr. Jenny Venton
The Royal Surrey NHS Foundation Trust
Egerton Road, Guildford, Surrey, GU2 7XX, UK

Date & signature: Jenny Venton, 21st May 2025

7 References

European Commission (2019). *Ethics guidelines for trustworthy AI.* Brussels: High-Level Expert roup on Artifical Intelligence. https://digital-st

European Commission (2021). *Ethics By Design and Ethics of Use Approaches for Artificial Intelligence*. Brussels: Directorate-General Research & Innovation, Research Ethics and Integrity Sector. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-

EU Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (OJ L 119, 4.5.2016, p. 1), https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1399

EU Directive 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA (OJ L 119, 4.5.2016, p. 89). https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1399642418437&uri=CELEX:32006L0024

European Commission (2021). *Ethics and data protection*. Brussels: Directorate-General Research & Innovation, Research Ethics and Integrity Sector. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

Council of Europe, European Court of Human Rights, European Data Protection Supervisor, European Union Agency for Fundamental Rights, (2018). *Handbook on European data protection law : 2018 edition.* Publications Office of the European Union. https://data.europa.eu/doi/10.2811/58814

European Commission (2021). *EU Grants: How to complete your ethics self-assessment vers. 2.0*. (2021-07-13), https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/how-to-complete-your-ethics-self-assessment_en.pdf

Marvasti, T. B., Gao, Y., Murray, K. R., Hershman, S., McIntosh, C., & Moayedi, Y. (2024). Unlocking Tomorrow's Health Care: Expanding the Clinical Scope of Wearables by Applying Artificial Intelligence. *Canadian Journal of Cardiology*, *40*(10), 1934–1945. https://doi.org/10.1016/J.CJCA.2024.07.009/ATTACHMENT/F1A3F39A-E190-44FA-8490-7216F98C1582/MMC1.PDF

Panch, T., Pollard, T.J., Mattie, H. *et al.* "Yes, but will it work for *my* patients?" Driving clinically relevant research with benchmark datasets. *npj Digit. Med.* **3**, 87 (2020). https://doi.org/10.1038/s41746-020-0295-6

Proposal for a Regulation laying down harmonised rules on artificial intelligence (2021-04-21), https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

European Commission (2020). *White Paper on Artificial Intelligence: a European approach to excellence and trust (COM(2020) 65 final).* Brussels. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en