

# DECIDE-AI - Delphi 1st Round

New reporting guidelines for the Developmental and Exploratory Clinical Investigation of DEcision support systems driven by Artificial Intelligence (DECIDE-AI).

## Annex V-4

## PARTICIPANT INFORMATION SHEET

Delphi round 1 and 2

Version 1 - 27.11.20

Developmental and Exploratory Clinical Investigation of DEcision support systems driven by Artificial Intelligence (DECIDE-AI): development of new reporting guidelines through a Delphi process.  
CUREC Approval Reference: R73712/RE002

We'd like to invite you to take part in our research study. Before you decide, it is important that you understand why the research is being done and what it would involve for you. Please take time to read this information, and discuss it with others if you wish. If there is anything that is not clear, or if you would like more information, please don't hesitate to contact us.

What is the purpose of the study?

This study aims to develop new reporting guidelines to make the early clinical evaluation of AI driven algorithms more consistent, comprehensive and reproducible. A robust evaluation at this stage where algorithms are first used by clinicians is indispensable to bridge algorithm development to large-scale clinical trials. To this end, DECIDE-AI ambition is to improve the evaluation reporting along four main axes:

i) the algorithm performance when first used with humans in small-scale, representative clinical conditions, ii) the safety profile of the algorithm prior to its wider-scale utilisation, iii) the human factors (ergonomic) evaluation of the algorithm, and iv) the preparatory steps towards large-scale clinical trials. A more detailed description of the project can be found in the document "DECIDE-AI project presentation".

Why have I been invited?

You have been invited on account of your expertise and experience in the field of artificial intelligence and/or evaluation of clinical intervention. We believe that your expert opinion will contribute to develop comprehensive and well-informed reporting standards.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw consent at any point for any reason, and without any adverse consequences or penalty. You can exit the survey at any time before submitting your answers by closing the browser. If you need to interrupt your session but wish to continue participating, you can choose the option to save your progress and return later. If you do not wish to answer a question, you can simply fill the answer field with "NA" or choose the "I don't know" option.

What will happen to me if I decide to take part?

The research project is based on the published Delphi methodology. The main goal of a Delphi process is to reach consensus between experts through several rounds of feedback. The results of each round are presented during the following round to inform participants' decision and guide them toward consensus. The present Delphi adaptation will include two general rounds, followed by a consensus meeting (for a subset of participants). The present invitation and information sheet refer to the two general rounds only. Consent will be asked before each round through an online form.

Round 1 (online survey): you will be asked to answer open-ended questions about what you think should be reported when evaluating an artificial intelligence based algorithm for the first time in clinical settings. You will then be asked to score, on a 1-9 scale and according to their importance, a list of provisory items developed by the research team and reviewed by the DECIDE-AI steering group. You will also have the opportunity to add comments, propose new items and recommend additional experts to take part in the Delphi. This round should take you between 45 and 60 minutes, depending on the extent of your free text inputs. You can interrupt your session at any time and resume later.

Round 2 (online survey): you will be asked to score, on a 1-9 scale and according to their importance, a modified list of reporting items, updated based on the first-round results and feedback. You will again have the opportunity to provide open comments. This round should take you between 30 and 45 minutes, depending on the extent of your free text inputs. You can interrupt your session at any time and resume later.

Are there any possible risks from taking part?

This study is considered to be at very low risk of physical or psychological harm. Issues related to data breaches or loss of confidentiality cannot be totally excluded. However, the research team took this aspect very seriously and designed a data handling strategy to minimize this risk.

#### How will my data be used?

The data we will collect that could identify you will be: your name, your affiliation, your main professional geographical location, your stakeholder group, your experience with AI/clinical evaluation and your professional email address. Your answers will be dissociated from these data using the REDCap software and data analysed in a de-identified manner. We will take all reasonable measures to ensure that your answers remain confidential. Your personal data and answers will be retained in the REDCap DECIDE-AI project database. REDCap is a secure web application, developed by a multi-institutional consortium initiated at Vanderbilt University. Access to the REDCap DECIDE-AI project database is password protected and for authorized users only. Additionally, your personal data and de-identified questionnaires' answers will be stored on a password-protected university network drive. Any linkage between your personal data and questionnaires' answers as well as your professional email addresses will be deleted one month after the publication of the study results. De-identified questionnaires' answers and consent records will be stored for at least three years on a password-protected university network drive. Your name, affiliation, stakeholder group, experience with AI/clinical evaluation and main professional geographical location will remain part of the study outputs (see below).

#### Who will have access to my data?

The University of Oxford and REDCap are the data controller with respect to your personal data and, as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that we perform in the public interest. Further information about your rights with respect to your personal data is available from <https://compliance.admin.ox.ac.uk/individual-rights>. The data you provide will be de-identified before it is shared outside the research team. As required by the Delphi methodology, anonymous (and often aggregated) answers will be presented to the other participants during the second round, to members of the steering group and to members of the consensus group. Anonymous quotes and anonymous aggregated answers may be used in academic publications.

#### Will my taking part in the study be kept confidential?

In your capacity as invited expert, and for the transparency of the process, your name and affiliation will be disclosed in the final publication. However, your answers will be kept confidential and no linkage between your personal information and answers will be shared outside the research team.

#### Will I receive compensation for taking part?

There will be no financial or in-kind compensation for taking part.

#### What will happen if I don't want to carry on with the study?

You may withdraw consent at any point for any reason, and without any adverse consequences or penalty. If you withdraw your consent up to ten working days after the end of a Delphi round, all data collected during this round will be deleted. Otherwise, your personal information will be deleted but the rest of your answers will remain part of the analysis.

To withdraw consent after completing one or both of the surveys, please contact the primary researcher or Principal Investigator (details below).

#### What will happen to the results of this study?

The results of this study will be published in a peer-reviewed journal and advertised through social media platforms. Participants will have full access to the results on request to the research team. This project will be written up for a DPhil degree.

#### Who is organising and funding the study?

The Principal Researcher is Baptiste Vasey and the Principal Investigator is Prof Peter McCulloch, who are affiliated to the Nuffield Department of Surgical Sciences at the University of Oxford. The project is carried out in collaboration with the DECIDE-AI Steering Group

No specific founding was acquired for the project. Funding is available if required from the IDEAL Collaboration research group general funds. BV is supported by a Berrow Foundation Lord Florey scholarship.

#### Who has reviewed this study?

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee [reference number].

#### Who do I contact if I have a concern or I wish to complain?

If you have a concern about any aspect of this study, please speak to Baptiste Vasey ([baptiste.vasey@nds.ox.ac.uk](mailto:baptiste.vasey@nds.ox.ac.uk)) or their supervisor Prof Peter McCulloch ([peter.mcculloch@nds.ox.ac.uk](mailto:peter.mcculloch@nds.ox.ac.uk)/+44 (0)1865 740870), and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Medical Sciences Interdivisional Research Ethics Committee;

Email: [ethics@medsci.ox.ac.uk](mailto:ethics@medsci.ox.ac.uk);

Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

CONSENT

First name: \_\_\_\_\_

Middle name: \_\_\_\_\_

Last name: \_\_\_\_\_

Please note that you may only participate in this survey if you are 18 years of age or over. ☐ yes ☐ no

I certify that I am 18 years of age or over:

I have read the information above and agree to participate with the understanding that the data (including any personal data) I submit will be processed accordingly: ☐ Yes, I agree to take part ☐ No, I don't wish to take part

## INTRODUCTION

Thank you very much for accepting our invitation to participate as an expert in the present Delphi process.

The DECIDE-AI guidelines aim to improve the reporting of the first-with-humans/early-stage clinical evaluation of artificial intelligence (AI) driven decision support systems in order to optimally connect algorithm development to large-scale clinical trials. This intermediary evaluation step has four key objectives:

to confirm the system performance when used with humans in small-scale actual clinical settings, to assess the safety profile of the system prior to its wider-scale utilisation, to conduct the human factors (ergonomic) evaluation of the system, to collect the background information necessary for the design of further large-scale (randomised controlled) clinical trials. Please make sure you have read the project description and objectives (see document attached) before proceeding to the questions.

In this first round, we would like to capture, through open-ended questions and from a wide range of stakeholders, the different opinions about what should be reported when an AI driven algorithm is first tested for its actual impact on human decision-making. In addition, you will be invited to score and comment on a provisory list of reporting items generated by the DECIDE-AI steering group. Your answers to this round will be used to modify and update the provisory items list. In the Delphi's second round, you will score each reporting item of the updated items list according to their relevance. During this second scoring exercise, you will see each item's average score from the first round to inform you about the current state of consensus between all the participants.

[Attachment: "DECIDE-AI Project presentation.pdf"]

---

**DEFINITION** Early-stage clinical evaluation refers to the formative, small-scale, evaluation of an algorithm with human users in actual clinical settings, with the intention to collect preliminary data about the impact of the algorithm on human decision-making. In the context of this Delphi, the chronological stage of evaluation and the emphasis on the human-computer interaction are more important than the study type used for this evaluation (which will depend on the main purpose of the tested algorithm).

In the context of this Delphi, algorithm refers to a clinical decision support system based on artificial intelligence/machine learning and whose main objective is to be interactively used as an adjunct to human intelligence in order to support clinician decision making (i.e the final decision is made by the human user). This decision support system can be of any modalities (detection, diagnostic, prognostic, therapeutic, ...).

User refers to the human using the algorithm and making the decision. In most of the current literature, this person is a medical doctor, but they can also be an allied health professional or a patient.

Human factors (or ergonomics) is defined as "the scientific discipline concerned with the understanding of interactions amongst humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance". International Ergonomics Association

"A reporting guideline is a simple, structured tool for health researchers to use while writing manuscripts. A reporting guideline provides a minimum list of information needed to ensure a manuscript can be [...] understood by a reader, replicated by a researcher, used by a doctor to make a clinical decision, and included in a systematic review. [...] Whether presented as structured text or a checklist, a reporting guideline [...] presents a clear list of reporting items that should appear in a paper and explains how the list was developed." EQUATOR Network The DECIDE-AI Steering Group is aware of other reporting guidelines (being) developed around artificial intelligence in healthcare and is committed to guidelines harmonisation.

---

**INSTRUCTION**

---

This is Round 1, which is broken down into 3 main tasks:

**Generating key themes:** In section 1 to 4, you will be asked open-ended questions designed to elucidate general themes about what should be evaluated during the early-stage clinical assessment of an AI-based algorithm, whose main intended use is to influence human decision-making. These themes will inform the content of the items list for the Delphi's second round and the discussion of the consensus group. They can also guide your answers to the two following tasks. **Scoring items:** In section 5, you will be asked to score each item of the provisory items list on a scale from 1 to 9 (1-3: not important; 4-6: important but not critical; 7-9: important and critical). In other words, not important items should not be included in the guidelines; important but not critical items should be reviewed case by case by the consensus group; important and critical items should be included. You will also have the option to enter comments about each item, and/or propose modifications (e.g. re-word an item, split one item in several items, delete part of an item, etc). **Proposing new items:** At the end of each subsection you will have the opportunity to propose additional reporting items. When doing so, please be as specific as possible with the wording and bear in mind that the item should apply to all types of AI-based algorithms designed to support human decision-making. Please provide a justification/explanation detailed enough, which we will share during the second round and the consensus meeting. If you need to interrupt your session, please use the "Save and Return Later" button (at the bottom of each page). If you close the tab or window directly, you will have to start over from the beginning.

We really appreciate your valuable contributions and insight. Please feel free to contact us (baptiste.vasey@nds.ox.ac.uk) at any point should you have further questions.

---

**TIME MANAGEMENT** To facilitate time management here is the detailed content of the first round:

4 open-ended questions 54 items to score (+ optional comments) 7 opportunities to add item (optional) 4 questions on participants personal information 3 questions on personal suggestions (optional) We recommend you to plan at least 30 minutes for the items list section.

---

**LET'S START** Please consider an AI-based algorithm whose intended use is to influence human decision-making. The algorithm has already been evaluated on a dataset of clinical data and its performance in silico (i.e via computer simulation) has been judged satisfactory to proceed to the next step of implementation. It is now time to test this algorithm when used with human clinicians for the first time, at small-scale and in actual clinical settings. The results of this early-stage clinical evaluation (formative evaluation) will inform if and how the evaluation should progress to large-scale clinical trials (summative evaluation, for example through randomised controlled trials). Its finding will also provide the rationale for early stage model improvements.

## 1. ALGORITHM PERFORMANCE

When using an AI-based decision support algorithm for the first time with human clinicians in actual clinical settings and at small-scale, which aspects of the system should be considered and what data collected to ensure that the algorithm performs as well when used interactively with humans as it did in the development phase?

Example: The assisted human performance (as defined by the clinical application) should be evaluated and compared to the algorithm in silico (i.e. through computer simulation) performance. Because human operators don't always follow the algorithm's recommendations, a system which is efficient in theory can perform poorly when human clinicians have the last word on the decision-making. The choice of performance metrics to be reported is context-dependent but should reflect the metrics used in the model development studies.

---

## 2. PATIENTS SAFETY

---

When using an AI-based decision support algorithm for the first time with human clinicians in actual clinical settings and at small-scale, which aspects of the system should be considered and what data collected to ensure that the algorithm is safe for patients?

Example: It is important to analyse any instances where the system gives an erroneous recommendation. The occurrence rate of such errors as well as their potential implication for patients should be reported as they are necessary to assess the safety profile of the CDSS.



### 3. HUMAN FACTORS

When using an AI-based decision support algorithm for the first time with human clinicians in actual clinical settings and at small-scale, what usability and human factors aspects of the algorithm should be evaluated?

Example: The users' subjective workload when using the algorithm should be evaluated and reported using the NASA Task Load Index (TLX) score.

#### 4. PREPARING LARGE-SCALE TRIALS

When using an AI-based decision support algorithm for the first time with human clinicians in actual clinical settings and at small-scale, which aspects of the system should be considered and what data collected to inform the design of a future large-scale (randomised controlled) clinical trial?

Example: The evolution of trust in the algorithm should be assessed. This would inform the training/familiarisation time necessary for users before starting measurements in a large-scale trial. The number of overrides of the algorithm's recommendation over time should be reported as well as the time point where it reaches stability.

## 5. PROVISORY ITEMS LIST

Below is the provisory items list developed from literature searches and discussions within the steering group and with external experts.

To avoid biasing the participants' opinion, the research team is not providing any arguments to justify the items at this stage. However, neutral explanations are provided for some items (marked with "explanation available"), when the proposed concepts might not be common knowledge for all of the involved stakeholders. These explanations are only informative and do not form part of the item.

Could you please grade each item (there are 54 in total) according to their importance:

1-3: not important 4-6: important but not critical 7-9: important and critical In addition, you will have the possibility to:

propose modification to items propose new items If you propose the modification of an item or the addition of a new item, please explain why in detail.

Display answers to question 1 to 4?

☐ Yes  
☐ No

By clicking yes, you can see your answers to questions 1 to 4 as reference to inform your feedback on the items list.

Performance  
[performance\_freetext]

Patients safety  
[safety\_freetext]

Human factors  
[hf\_freetext]

Preparing large-scale trial  
[trial\_freetext]

TITLE

	1	2	3	4	5	6	7	8	9	I don't know
1. Identify the study as early clinical, exploratory or first-with-human assessment of an artificial intelligence or machine learning based clinical decision support algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Do you have any comments on the item(s) in the Title section? (Please always specify the item number when commenting.)

Do you want to propose any new item(s) to the Title section?

☐ Yes  
☐ No

Title - New item 1

---

Title - New item 1: what should the wording of the item be?

---

---

Title - New item 1: please provide a justification.

---

---

Title - New item 2

---

---

Title - New item 2: what should the wording of the item be?

---

---

Title - New item 2: please provide a justification.

---

---

Title - additional new item(s)

---

---

Title - additional new item(s): please word and justify any additional new item(s) you would like to propose.

---

---

#### ABSTRACT

	1	2	3	4	5	6	7	8	9	I don't know
2. Provide a structured summary of the study, including mention of: [will be completed according to the outcomes of the Delphi]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

Do you have any comments on the item(s) in the Abstract section? (Please always specify the item number when commenting.)

---

---

Do you want to propose any new item(s) to the Abstract section?

☐ Yes  
☐ No

---

---

Abstract - New item 1

---

---

Abstract - New item 1: what should the wording of the item be?

---

---

Abstract - New item 1: please provide a justification.

---

---

Abstract - New item 2

---

Abstract - New item 2: what should the wording of the item be?

Abstract - New item 2: please provide a justification.

Abstract - additional new item(s)

Abstract - additional new item(s): please word and justify any additional new item(s) you would like to propose.

## INTRODUCTION

	1	2	3	4	5	6	7	8	9	I don't know
3. Describe the target conditions and the patient population that would benefit from the algorithm, including information on the target conditions' prevalence and their impact on the healthcare system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Describe the intended use of the algorithm, including its position in the care pathway and the conditions under which it would be used, the impact in terms of patient care it intends to achieve and the current state of the art practice.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Refer to the algorithm's development and validation studies and name the algorithm, including the version number. State the algorithm's expected performance from development and validation studies.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Identify the dataset used to develop the algorithm and provide information on its relevance to the test environment, including the target conditions' prevalence when appropriate.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. Describe the current stage of development of the algorithm in terms of the essential questions which have been and remain to be answered about it (both from a scientific and a regulatory perspective).

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

8. State the study objectives.

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Do you have any comments on the item(s) in the Introduction section? (Please always specify the item number when commenting.)

\_\_\_\_\_

Do you want to propose any new item(s) to the Introduction section?

☐ Yes  
☐ No

Introduction - New item 1

Introduction - New item 1: what should the wording of the item be?

\_\_\_\_\_

Introduction - New item 1: please provide a justification.

\_\_\_\_\_

Introduction - New item 2

Introduction - New item 2: what should the wording of the item be?

\_\_\_\_\_

Introduction - New item 2: please provide a justification.

\_\_\_\_\_

Introduction - additional new item(s)

Introduction - additional new item(s): please word and justify any additional new item(s) you would like to propose.

\_\_\_\_\_

## METHODS

	1	2	3	4	5	6	7	8	9	I don't know
9. Provide a reference to any study protocol.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Specify the primary and secondary outcome measures.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Describe the study design using standard methodological terminology.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Describe precisely how users and patients were selected. If only a subgroup of users took part in the human factors evaluation, describe how these were selected.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. Justify the sample sizes (for both users and patients).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. Identify the hardware and software platforms used during the study. Describe the data needed by the algorithm as inputs, the data provided by the algorithm as outputs and the minimal computational resources needed. (explanation available)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. Describe how the algorithm was used, at which stage of the decision-making process and who held the responsibility for the final clinical decision.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. Describe precisely the environment in which the algorithm was tested, including the availability of the algorithm's input data and which additional clinical information (i.e. not provided by the algorithm) was accessible to the users.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. Describe how the patient data were acquired (including from which sources), how they were processed and how missing or low-quality data were handled.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. State what measures were taken to protect patient privacy and data security. (explanation available)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

19. Describe the control group in sufficient detail to allow replication.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20. State the predefined statistical analysis plan and any additional exploratory analyses performed (state if the chosen approach accounts for both user and patient variability).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. State any pre-specified subgroup analyses and their rationale.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. Define the algorithm safety requirements, how these were established and how compliance to these requirements was evaluated. (explanation available)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. Describe how algorithm recommendation/output errors were defined and how they were identified. (explanation available)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. Describe any attempts to familiarise users with the algorithm, including any training received.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25. Describe the human factors tools, methods or frameworks used to evaluate usability, situation awareness and any other relevant human factors considerations. Justify this choice. (explanation available)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
26. Describe any attempt to understand the user acceptance of the algorithm as well as user deviations from the algorithm's recommendations or intended use.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27. Describe any involvement of patients in understanding their opinion on the algorithm and how the algorithm's outputs could influence their care.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



28. Describe the methodology used to collect and analyse data for the health economic assessment of the algorithm's use. (explanation available)

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

---

#### Additional explanations

Item 14: Computational resources include, but are not limited to, the time, processing power, memory storage or bandwidth needed to complete a computational task.

Item 18: Amongst other, debated definitions, patient privacy refers to the freedom from intrusion in the patient confidential data and the right of the patient to make decision about the acquisition, storage and sharing of their confidential data. Data security is the protection against unauthorized access, modification or theft of data. Breach of data security can have consequence on privacy but also on the system overall functionality.

Item 22: Safety requirements are generated by domain experts or derived from regulatory requirements and are informed by the system's risk analysis (AMLAS 2021). They should ensure an acceptable level of risk compared to the potential benefits of the system. Several methods and tools are available to support risk management (e.g. FMEA or SHERPA). For example, a safety requirement could be "the number of false negatives should not exceed X % of the total cases" or "the algorithm should never recommend drug X to patients with condition Y".

Item 23: For example, errors can be defined in comparison with a reference standard, as a failure to detect an event within a given timeframe or as a therapeutic option vetoed by the clinical team.

Item 25: A description of usability and situation awareness is given in the results section.

Item 28: "Health economics is about using resources efficiently to improve the population's health. Health economic [assessment] forms an integral part of the public health guidance development process" (NICE). The most appropriate way to conduct a health technology assessment and the level of economic analysis required at each implementation stage are context dependent. However, evidence standards, like for example the NICE Evidence Standards Framework for Digital Health Technologies (section B) exist to guide this process.

---

Do you have any comments on the item(s) in the Methods section? (Please always specify the item number when commenting.)

---

---

Do you want to propose any new item(s) to the Methods section?

☐ Yes  
☐ No

---

Methods - New item 1

---

Methods - New item 1: what should the wording of the item be?

---

---

Methods - New item 1: please provide a justification.

---

---

Methods - New item 2

---

Methods - New item 2: what should the wording of the item be?

---

Methods - New item 2: please provide a justification.

Methods - additional new item(s)

Methods - additional new item(s): please word and justify any additional new item(s) you would like to propose.

## RESULTS

	1	2	3	4	5	6	7	8	9	I don't know
29. Describe the user population baseline characteristics (number, number of centres, specialty, seniority, previous experience with digital support, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. Describe the patient population baseline characteristics (number, number of centres, age, sex, ethnicity if relevant, comorbidities, prevalence of the target conditions, etc.).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
31. Report on the proportion of intended users who had exposure to the algorithm (implementation reach), on the number of instances the algorithm was used (implementation dose) and on the users' compliance with the intended implementation (implementation fidelity).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
32. Report on the prespecified outcomes for the algorithm-assisted users (both overall and at an individual user level) as well as for the control group.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

33. If applicable, report on the prespecified outcomes which would have been observed had all the algorithm's recommendations been strictly followed. (explanation available)

☐☐☐☐☐☐☐☐☐☐☐

34. Describe any instances where the algorithm gave an erroneous recommendation/output. Report their rate of occurrence and detail their potential impact on patient care.

☐☐☐☐☐☐☐☐☐☐☐

35. Report on the compliance with the safety requirements.

☐☐☐☐☐☐☐☐☐☐☐

36. Describe any instances where users decided to override the algorithm's recommendation or to follow an erroneous recommendation.

☐☐☐☐☐☐☐☐☐☐☐

37. Report on the evolution of users' trust in the algorithm (evolution of the overrides of the algorithm's recommendation with time) and on the learning curves (evolution of the users' performance with time).

☐☐☐☐☐☐☐☐☐☐☐

38. Report the number of users involved in the human factors evaluation, their characteristics and the use cases examined. (explanation available)

☐☐☐☐☐☐☐☐☐☐☐

39. Report on the usability evaluation, including time to task completion and display interface evaluation, using method-specific metrics. (explanation available)

☐☐☐☐☐☐☐☐☐☐☐

40. Report on the situation awareness evaluation and on the users' perspective on the algorithm's interpretability. (explanation available)

☐☐☐☐☐☐☐☐☐☐☐

41. Report on the outcomes of any other human factors evaluation, including the user acceptance of the algorithm and any induced changes in the care pathway.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
42. Summarize all changes made to the algorithm or its hardware platform between the prototype used at the beginning of the study and its final version. Report the timing of these modifications and the changes in outcomes observed after each design-evaluation cycle. (explanation available)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
43. Report the patients' opinion on the algorithm and whether they would accept their care being influenced by it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
44. Report the results of the health economic assessment of the algorithm's use and identify any trade-offs in the care pathway.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

#### Additional explanations

Item 33: In other words, the hypothetical performance of the “stand-alone” algorithm, without human intervention, in the study environment.

Item 38: A use case is “a specific situation in which a product or service could potentially be used” (Oxford Languages), or is used, to achieve a goal.

Item 39: Usability is defined as the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO 9241-11). In the context of medical device usability testing can include, but is not limited to, the evaluation of user experience, time to task completion, display interface, cognitive load or potential system malfunctions. System malfunctions can be different from erroneous recommendation/output (e.g. failure to produce a recommendation/output at all).

Item 40: Situation awareness is the perception of elements in one subject’s environment within a defined time and space, the comprehension of their meaning and the projection of their status in the near future (Endsley 1988). It can apply to both the algorithm, which should stay aware of changes in the patient data or received care, and the users, who need to integrate information from multiple sources for the index patient and might be simultaneously using different decision-support tools for different patients. (Sujaan et al. 2019). Arguably, issues around the algorithm’s interpretability can fit into this category as interpretability contributes to the comprehension of a situation’s meaning and communication about a system’s “normal” operating conditions supports situation awareness.

Item 42: During formative evaluation, rapid modification-evaluation cycles and their thorough recording are part of the process. In each cycle the algorithm or its platform will be modified according to user feedback gathered as the algorithm encounters an increasingly wider range of users and clinical situations.

---

Do you have any comments on the item(s) in the Results section? (Please always specify the item number when commenting.)

---

Do you want to propose any new item(s) to the Results section? ☐ Yes ☐ No

Results - New item 1

Results - New item 1: what should the wording of the item be?

Results - New item 1: please provide a justification.

Results - New item 2

Results - New item 2: what should the wording of the item be?

Results - New item 2: please provide a justification.

Results - additional new item(s)

Results - additional new item(s): please word and justify any additional new item(s) you would like to propose.

DISCUSSION

	1	2	3	4	5	6	7	8	9	I don't know
45. Discuss if the obtained results support the intended purpose of the algorithm in real world healthcare settings, including how the outcomes would translate into patient benefit, or if an alternative use could be more appropriate.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
46. Explain what was learned about the reasons for human deviation from the algorithm's recommendations or intended use, and what this tells us about achieving better alignment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

47. Discuss the algorithm's errors and identify any underlying pattern or algorithmic bias. Explain how these can be mitigated.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
48. Discuss what the results suggest about the safety profile of the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
49. Discuss the human factors results and comment on the evolution of the algorithm/hardware platform design. Discuss the need for additional technical requirements or product design improvement before large-scale summative evaluation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
50. Comment on the evolution of users' trust in the algorithm and on the learning curves. State when they reached a stable state.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
51. Highlight any performance difference in user or patient subgroups and discuss the merits of limiting further evaluation to a specific group of users or patients.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
52. Discuss the feasibility and appropriateness of large-scale summative evaluation in light of the obtained results.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

Do you have any comments on the item(s) in the Discussion section? (Please always specify the item number when commenting.)

---



---

Do you want to propose any new item(s) to the Discussion section?

☐ Yes  
☐ No

---

Discussion - New item 1

---

Discussion - New item 1: what should the wording of the item be?

---



---

Discussion - New item 1: please provide a justification.

---



---

Discussion - New item 2

---

Discussion - New item 2: what should the wording of the item be?

---

---

Discussion - New item 2: please provide a justification.

---

---

Discussion - additional new item(s)

---

---

Discussion - additional new item(s): please word and justify any additional new item(s) you would like to propose.

---

---

#### STATEMENTS

	1	2	3	4	5	6	7	8	9	I don't know
53. Disclose the source of funding for the study and authors' relevant conflicts of interest.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
54. Disclose code and data availability.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

---

Do you have any comments on the item(s) in the Statements section? (Please always specify the item number when commenting.)

---

---

Do you want to propose any new item(s) to the Statements section?

☐ Yes  
☐ No

---

---

Statements - New item 1

---

---

Statements - New item 1: what should the wording of the item be?

---

---

Statements - New item 1: please provide a justification.

---

---

Statements - New item 2

---

---

Statements - New item 2: what should the wording of the item be?

---

---

Statements - New item 2: please provide a justification.

---

---

Statements - additional new item(s)

---

---

Statements - additional new item(s): please word and  
justify any additional new item(s) you would like to  
propose.

---



**6. PERSONAL INFORMATION**

What is your main affiliation?

---

In which country do you mainly work?

---

Which of the following groups of stakeholders would describe you best?

Select any which are applicable.

- ☐ Administrator/other management position in hospital
- ☐ Allied health professional
- ☐ Clinician
- ☐ Engineer/Computer scientist
- ☐ Entrepreneur
- ☐ Epidemiologist
- ☐ Ethicist
- ☐ Funder
- ☐ Human factors specialist
- ☐ Implementation scientist
- ☐ Journal editor
- ☐ Methodologist
- ☐ Patients' representative
- ☐ Payer/Commissioner
- ☐ Policy maker/official institutions representative
- ☐ Private sector representative
- ☐ Psychologist
- ☐ Regulator
- ☐ Statistician
- ☐ Trialist
- ☐ other

please specify

---

Could you please briefly describe your level of experience/type of expertise with one or several of the following:

---

artificial intelligence/machine learning; clinical evaluation/technology implementation; evaluation of human-computer interaction; other relevant.

---

**7. OPTIONAL QUESTIONS**

---

Do you have any other comments on the content of the Delphi first round, or feedback you would like to share with the Steering Group?

---

---

Do you have any references (scientific publications, regulatory documents, reports, etc.) that you would like to highlight in the context of DECIDE-AI?

---

---

Are there other relevant experts you would like to recommend for this Delphi? Please provide name, title and an affiliation. We will contact them if their professional contact details are publically available online. You can recommend up to three experts here. Due to the timeline of the Delphi, we cannot guarantee to contact the recommended experts if their names are submitted after March the 7th 2021.

---