

# IEEE P7003 Working Group Meeting Minutes 6<sup>th</sup> January 2022 / 13:00 PM UTC - 15:00 PM UTC Teleconference

#### 1. Call to Order

The meeting was called to order at 13:05 UTC

## 2. Roll call and Disclosure of Affiliation

The list of attendees is attached. A quorum was reached and noted.

## 3. Approval of January Agenda

**Motion to approve the agenda for the January meeting.** The agenda was approved without change.

# **4. IEEE Patent Policy** (Call for Patents)

The call for patents was raised; no one raised any concerns or any comments for consideration.

## 5. IEEE SA Copyright

The copyright policy was presented.

## 6. Approval of 2<sup>nd</sup> December meeting minutes

Motion to approve the minutes from the  $2^{nd}$  December 2021 meeting. The motion to approve the minutes from the  $2^{nd}$  December 2021 meeting was approved.

### 7. Announcements

#### 8. Structural review of P7003 work

I. Review of Jan update of master working document New master document updated from May 2021 → January 2022. Access to this has been given to everyone who previously had access, for suggesting rather than direct editing so we can see changes. Please can people also comment why they have made a change if it is non-trivial.

 $\underline{https://docs.google.com/document/d/1RNgU7dkT\_W4XIPhlLsZnIfWsm4Fq4GZCo4TvBwghauI\_/edit?usp=sharing}$ 

This version includes the most recent version of all the drafts and cleaned up removed the taxonomy sections, replacing with definitions, and Conceptual Insights about algorithmic bias. Comments from previous version are maintained; some have been accepted/resolved if they are simple or typos etc. Previous versions still exist so you can go back if you think something is

missing. Each section heading has a link to the section draft document, and indication of last update.

We are still looking at updating the figures as a visual way to understand where in the document to look for content. Anyone willing to help with this please get in touch.

Proposal to go through one section each month in depth for an hour. It is increasingly difficult to review one section without the context of the other sections, they need to be reviewed in the context of the entire standard.

i. List of terminology shifts between sections Due to different sections being worked on independently there may sometimes be inconsistencies in terminology, eq algorithmic systems vs AIS vs AI/S. Clare sent some useful documents to refer to: https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead1e\_glossary.pdf; https://ieeexplore.ieee.org/browse/standards/readingroom/page?refinements=ReadingRoom:Additional%20Reading&ranges =2021 2021 Year&utm medium=aem&utm source=ieeesa&utm cam paign=ais-2021. It was suggested that at this point we should look for making sure all the content and use cases are there and coherent and then work on the definitions. However, it may be beneficial to come to consensus on basic terminology and it might be easy to do that now. Some terms change the sense of the sentence they are within so it may be useful for some of them to be agreed on. Do we know which terms or more or less contentious? Differences between actual definition in the document and how the word is actually used in the standard?

It's unclear what terms should be used in some cases as there is no common consensus between or within other standards. We may also need to consider if some of the previously included definitions are still needed. For example, the way stakeholders are discussed has changed.

ii. Proposal for illustrative use-case

We need some use cases that are used throughout the standard. These should be fairly short pieces of text to say how you would apply a particular normative requirement to that use case. Anyone who has experience of use cases in different standards? It can be tricky because it depends on the domain of the use case. The challenge is picking a good representative mix of domains eg medical, transport, cybersecurity, financial, ML. Extremely useful if it can be done. Need domain specificity but cannot be too specialized, needs to be relevant and laid out clearly and economically, with consideration of how much prior knowledge can be assumed. Suggestions include AV, credit fairness, face recognition, hiring, things you can talk about without necessarily going too deep into how it is done. Could we align with other standards using use cases? Areas we can talk about in terms of ML and other types of algorithms. Hiring and credit seen as good options.

Most use cases seem to assume a human user but a non-human situation is not discussed. Machine-to-machine use cases should be

considered. These were included at the start but they were sidelined as ultimately concerned with impact on people and environment. Mathana suggested that it's maybe an issue of due diligence, requirements to input data from non-trusted sources, eg industrial QA/QC processes, creating a normative threshold for confidence of data input. Mark and Shashidha to look through the document and suggest places where this m2m might be included.

Discussion of 'fairness' and measurement, and distinction from bias, as well as considerations surrounding environmental and workforce impacts.

Since much of the bias discussions for both use cases is focused on aspects that impact the 'quality' of the outputs produced by the algorithmic system it would be good to identify a third use cases that bring out bias aspects related to the way the algorithmic systems' performance metrics etc. are conceptualized. This could include aspects such as sustainability, labor conditions.

Action: Between now and next call start thinking about the hiring and credit scoring use cases and possible third.

II. How to approach development of an ontology for P7003 (Randy) Not covered in this session as Randy not in attendance.

## 9. Updated Outline Discussion

i. Requirements

Focused around clarification as to what the values of the organisation are that go into the intent of the use of the AI system that is being developed, deployed or used, eg mapping accountability. This still needs a fair amount of work, some sections are very brief outlines.

Discussion of the term 'justification' and appropriateness of choices. The section requires recognition that an unbiased system is impossible, and requests to clarify POV so that people understand what bias is considered acceptable for example. Are we talking about 'bias within bias thresholds' where those thresholds need to be defined and adhered to? Discussion also of scaling of use, which could potentially be a change in the context of use, and has an impact on a lot of different sections. Potentially related to provenance and traceability. Might be something in P7000 we can cite or P2863 on organizational governance.

ii. Stakeholder Identification

Minor updates relative to what was there before.

Also separated out informative parts of Stakeholder Identification which is placed as an Informative annexe

iii. Risk and Impact Assessment

Very minor updates relative to what was there before.

Risk mitigation section is very much tied to other sections and so should probably be reviewed at a later date. Needs eyes on it as it was written a couple of years ago.

iv. Data Representation

Team sent an updated version. Please have a look at point out any holes. It's easy to lose focus and hard to avoid context drift. Do not feel you need to add more text if it is not necessary.

Also separated out informative content which is placed in the Informative annexe.

- v. System evaluation
  - Some old text from previous versions has been removed, and it's now more focused. Would be good to look at it in detail soon.
  - Also separated out for the Informative annexe
- vi. Conceptual insights about causes of bias [name to be improved]
  Temporary new title for taxonomy, retaining the understanding of where bias comes from.
- vii. Legal frameworks
  - Has not changed from previous version.
- viii. Human Factors
  - Hasn't been updated since August 2020, needs significant work to simplify for easy digestion by non-academic audiences and reduce diagrams
- ix. Cultural aspects
  Somewhat revised. Needs someone to go through it, email Clare.

## 10. Any Other Business

## 11. Dates/times for Future Meetings

- Thursday 3<sup>rd</sup> February @ 2200 UTC
- Thursday 3<sup>rd</sup> March @ 1300 UTC
- Thursday 7<sup>th</sup> April @ 2200 UTC
- Thursday 5<sup>th</sup> May @ 1300 UTC
- Thursday 2<sup>nd</sup> June @ 2200 UTC

# 12. Adjourn

The meeting was adjourned at 15:07 UTC

#### **Attendees:**

Last Name	First Name	Employer/Affiliation	Voting
Carneiro	Magnovaldo	Univesp - Virtual University of Sao Paolo	
Chaudhuri	Abhik	TATA Consultancy Services	
Chung	Edmon	DotAsia	
Clifton	Chris	Purdue University	X
Courtney	Patrick	Tec-connection	X
Dongre	Shashidhara	L&T Technology Services	
Dowthwaite	Liz	University of Nottingham	X
Duarte	Tania	Independent	
James	Clair	Independent	X
Koene	Ansgar	University of Nottingham	X
Leppala	Jussi	Valmet	X
Loughran	Roisin	Dundalk Institute of Technology	X
Padget	Julian	University of Bath	X
Perey	Christine	PEREY Research & Consulting	

Prasad	MS	Amity University, India	
Rivas	Pablo	Baylor University	
Shaw	Trish	Beyond Reach	
Stender	Mathana	Independent	
Underwood	Mark	Synchrony Finance	
Von Struensee	Susan	Independent	
Weger	Gerlinde	Independent	X
Whitaker	Jessica	Howard University	X