

**IEEE P7003 Working Group
Meeting Minutes
28th February – 1st March 2019
Face-to-Face and teleconference**

1. Call to Order

The meeting was called to order at 15:31 UTC 28th February 2019 and 10:03 UTC 1st March 2019.

A quorum has been established and noted.

2. Roll call and Disclosure of [Affiliation](#)

The list of attendees present is attached.

3. Approval of face-to-face meeting Agenda

Motion to approve the meeting agenda from 28th February – 1st March 2019.

The agenda was approved as submitted without objection.

4. [IEEE Patent Policy](#) (Call for Patents)

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

5. Approval of 7 February meeting minutes

Motion to approve the minutes from February 7th, 2019. The minutes from the February 7th, 2019 meeting were approved without objection.

6. Summary

As we were joined by the New York meeting and online participants, Ansgar summarized the presentations given during the morning session. They were dedicated to providing the background context to the standard both within IEEE and situated within other initiatives. The slides are uploaded to the shared Google Docs, and were also circulated via email. The talks were also video recorded and will be made available. This was followed by an introduction by Takeshi on the fundamentals of industry standards, and what kind of focus will ensure that our standard will actually be used. Takeshi has been involved in many, including those that no one ever uses.

7. Purpose of the P7003 standard

It is important that we are thinking about producing a standard that provides clear actionable points, so although it is an ethics standard that engages thought and isn't a mechanistic tick-box exercise, it does need to be something that someone can go

through and confirm they have addressed the requirements to be compliant with the standard. It needs to use language and points which allow us to say that something has or has not been achieved.

The standard is being developed in a new area, an ethics related standard not a compatibility standard so we need to consider how to get industry to want to engage with this kind of standard. There is a need to encourage people to engage with the standard to ensure their system, tool, etc meets ethical requirements, is for the benefit of humanity, and is trustworthy (ie not just that it works). This could relate to potential regulatory requirements, where if one is compliant, you can show you comply with regulations. We cannot control this, but the other path to usage is that companies want to be able to communicate that their system is something that is trustworthy to their users - people feel they are confident in using the system. This can be done by the standard becoming associated with other companies that already have a strong brand recognition in terms of trustworthiness, for example big established auditing companies. Discussion of how we get these involved, and that they don't necessarily need to be commercial, and how we write the standard so that it's appealing to these organisations.

Due to the broad nature of the standard and the context dependence of various aspects of bias and ethics in general, it may not make sense to always require compliance with everything that is in the standard. What does that mean and how does that affect the branding? Suggestion of optional versus compulsory sections, and how this might affect potential certification issues. Potentially, if a company has considered something but it's not relevant, it can be scored as not relevant so there is no negative effect. It should be included in the public documentation on what has been done to apply the standard - that a factor has been considered and through reasonable analysis of the issue the conclusion is that this is something that does not need to be considered further. This is maybe more about certification side rather than the standard side.

Along a similar line of thought, there was discussion about the idea that initial (and recurring) risk assessment may be used to identify if the system under consideration poses high/medium/low bias risk and how that might impact the way in which compliance with subsequent requirements of the standard is assessed.

8. Scope of P7003

Ansgar read out the original scope of the standard as written in the proposal. It specifies that it's not about getting rid of all forms of bias - it's about negative bias. It often makes it sound like the problem is bad input data but we are going further than that to say it can occur at all stages of the design and deployment cycle. Do we therefore need to emphasize a focus on outcomes rather than how the problems occur? "Direct negative impact on natural persons based on group membership" may more fully capture what we are trying to say, but at the same time we're not saying it's necessarily bad but needs to be called out and investigated.

Still need to consider the whole process through identifying harm and mitigating it, which is not just an outcome based objective.

Question of whether we need a legal frame for what bias is? Bias is implicit to different parts of legislation, and the standard contains a legal section. Focus on legislative criteria may be difficult because it's so different (and not necessarily our idea of ethical) across the board. Instead of focusing on the legislative link,

potentially reference human rights as a global accepted framework. It could mention compliance with international law, principals of equality and non-discrimination etc. Discussion about value-based design: whether you have a justification around your process and the way your decision system comes to its outputs. Which values are built in?

Action: Ansgar to draft a new version then circulate for comments

9. High level mapping of P7003 document

Another element to the discussion is making sure that we can feel confident the standard is delivering what we want. What is the actual content that we want to have in each of the sections in order to say that it is completed and performs the task that it needs to? What do we need in earlier sections that allows us to build on it for later purposes? As such, during the meeting a new Google doc was used for the two days, to put down bullet points to indicate the key things that each section needs to deliver on to be minimally complete. This doc has been circulated to group members via email ([P7003 section content checklist](#)).

The high level mapping diagram aims to help users to understand the position of the different sections relative to each other, and which parts are more relevant depending on the design stages of the system. It is a quick sketch of what it could look like and needs agreement and modification. Ansgar explained the current outline and then each section was looked at in terms, to identify if it was in the correct position and/or where it should be moved to.

Most of the top level bias requirements are sections that will need to be referred back to multiple times during the rest of the process, and this needs to be made clear. It was flagged that it is important to think about longitudinal risk factors, and how the standard can be used over time. If you use it at the beginning, at the design or right before live, you might not come across bias, the replication of use is how bias shows up. Outcomes evaluation for example should include all the upstream work including legal, psych, taxonomy etc.

There was a question about UX/UI considerations, and whether the outcomes evaluation relates to what is actually presented to end-users. This also led to discussion of the differences between outcomes evaluation and process evaluation. Suggestion to put in reference to other standards that deal with UX/UI bias considerations, for example P7010. Similar discussion of acknowledging and referencing different standards that are more in scope with this.

It was proposed that the section on "resilience against manipulation" should be removed since cybersecurity issues such as system hacking are either out of scope (we do not have the relevant expertise to address them), or can be considered part of other sections (e.g. Assurance of representativeness of [testing/training] data).

Discussion regarding the role of the "Documentation & Transparency" section revealed that this section should probably be split into two sections:

"Documentation", which focuses on providing guidance on HOW to document the compliance with requirements indicated in the other sections of the standard (WHAT needs to be documented will be specified within each of those sections).

"Transparency and Accountability", which addresses the issue of which parts of the documentation need to be accessible by whom (general public; regulatory authorities; auditors; internal ethics board etc). A key challenge will be how to resolve the balance between IP rights and public scrutiny. It is anticipated that there will be a range of documentation transparency levels and that the transparency

requirements will depend on the level of risk and impact of the intended/current use of the algorithmic system under consideration.

The "person categorization" section was considered to have too much overlap with either the "processing evaluation" or the "representativeness of data" section if it were maintained as originally pitched. Instead a related, and more important, issue was identified as the need for guidance on identifying and understanding the needs of the stakeholder who will be affected by the system. It was therefore proposed to re-label the section as "Stakeholder Analysis" and move it to a much earlier stage in the project process, just before "Risk Assessment".

For a more detailed discussion of the outcomes related to each section in this discussion, see the google doc referenced above.

10. Any other business

It was noted that many of the sections rely heavily on each other and it is therefore important that the subgroups working on them have good communication. As such, new 'teams' have been created that encompass 1 or more sections of the standard. Each team has a leader who is responsible for the overall sections, and is expected to pull the content together into the standard. The new teams are detailed at the end of the [google doc](#).

A slack workspace <https://ieeep7003.slack.com> was created to facilitate team working and minimize e-mail clutter.

Case Studies: After a re-evaluation of the intended use of the Use Cases it was decided to: a) move the current case studies into a separate document for internal use only, since it would not be appropriate for a published standard to point to specific organizations as examples of bad behavior (there can be legitimate uses for "name and shame" but standards are not the appropriate place – the audience that would benefit is not the audience that reads standards documents); b) consider the current case studies as reference points for inspiration for our work; c) assign a team to create a fictional use case example(s) that can illustrate how to apply the recommendations that are given in each section of P7003.

Action: We need someone to check through the current version of the document to identify key terms that require definitions, and make sure that we are consistent when using terminology.

Action: Ansgar to circulate new teams with request for working group members to a) confirm if they are still working on the sections they have been previously named against, and b) add their names to any other sections they can contribute to.

Action: Ansgar to create a new "master-document" (in G-doc format) that will form the basis of the next phase of work on P7003.

11. Future Meetings

Thursday, 4th April, 1:00pm – 2:30pm (UTC)

Thursday, 2nd May, 3:00pm – 4:30pm (UTC)

Thursday, 6th June, 5:00pm – 6:30pm (UTC)

12. Adjourn

The meeting was adjourned at 18:00 UTC on February 28th and 16:34 UTC on March 1st

Attendees:

Last Name	First Name	Employer/Affiliation	Day 1	Day 2	Voting
Bond	Raymond	Ulster University	X		
Brown	Shea	BABL AI Inc.	X	X	
Carrier	Ryan	ForHumanity		X	
Chaudhuri	Abhik	Tata Consultancy Services	x	x	X
Clifton	Chris	Purdue University	X	X	X
Costley	Jennifer	NY Academy of Sciences	X		
Courtney	Patrick	Tec-connection		X	X
Dean	Sheila	Independent	X	X	
Dechesne	Francien	U. Leiden	X		X
Dowthwaite	Liz	U Nottingham	X	X	X
Egawa	Takashi	NEC	X	X	
Gagwa	E Arthur	Centre for Intellectual Property and Information Technology Law, Strathmore University	X		
Gardner	Allison	Keele University	X	X	X
Gautam	Sumit	LG Electronics		X	
Gilburt	Ben	SOPRA STERIA	X	X	
Hailey	Vicky	VHG-The Victoria Hailey Group Corporation	X	X	
Hatada	Yohko	EMLS RI	X	X	X
Ilieva	Snezhana	PwC	X	X	
Jimenez Gomez	Carlos	Self		X	
Kapetanios	Epaminondas	U. Westminster	X	X	
Koene (Chair)	Ansgar	U Nottingham	X	X	X
Kostopoulos	Lydia	Digital Society Institute	X		
Levesque	Maroussia	Self	X	X	
Mandal	Sunakya	Self	X		X
Nadel	Lawrence	NIST		X	X
Ng	Vivian	U. Essex	X	X	
Rovatsos	Michael	University of Edinburgh	X	X	X
Saucedo	Alejandro	Institute of Ethical AI	X	X	

Shaikh	Fareed	self		X		
Smith	Adam	Piccadilly Group	X	X		X
Stender	Mathana	Self	X	X		
Tan	Caryn	Accenture	X	X		
Underwood	Mark	Synchrony	X			
Wickenden	Luke	Self	X	X		
Bahn	Christy	IEEE-SA (staff)				