Subject: Feedback on explainability NIST (Email feedback)
Date: Wednesday, October 14, 2020 at 8:09:05 AM Eastern Daylight Time
From: Paola Di Maio
To: NIST Explainable AI, W3C AIKR CG
Attachments: nist feedback.txt, Inbox (12,949) - paoladimaio10@gmail.com - Gmail.mp3, NIST COMMENTS.xlsx

Dear NIST
cc W3C AI KR CG

Thank you for the opportunity to provide feedback on the draft principles for Xplainability

I paste below and attach (as text, xls and mp3 narration) some comments, which I would be grateful if they could be taken into account and possibly addressed.

Looking forward to progress towards a standard for explainability
Keep us informed, thank you

Best regards
Paola Di Maio, PhD

FEEDBACK FOR NIST ON EXPLAINABILITY
Draft NIST IR 8312

from PAOLA DI MAIO, Expert and Co-chair W3C AI KR CG
13 October 2020

PREAMBLES
a) before explainability can be addressed in the context of AI, AI should be better understood/defined. The reality is that we may not yet have AI after all
b) In addition to the distinction between narrow and general AI, the distinction between closed vs open system AI is also necessary. This particularly applies to the point of Knowledge limits in the draft.

GENERAL COMMENTS ON THE PRINCIPLES IN THE DRAFT

1. EXPLANATION type mismatch among the principles
for example explanation, is a noun, while meaningful is an adjective, would be advisable to have some consistency in the naming conventions?
2. MEANINGFUL explanation is described as a principle that mandates an explanation for AI, and meaningful is described as a principle that the explanation is meaningful, but it does not describe criteria/parameters for meaningfulness. This does not seem up to standard. Looks to me that meaningful is a qualifier for explanation (1)
3. EXPLANATION ACCURACY - same as above, this does not seem a principle more like a qualifier for principle 1. Looks to me that 2 and 3 are qualifiers for 1. however they should be better defined

4. KNOWLEDGE LIMITS - this is new (ie. unheard of) Is there a reference for such a notion? Where does it come from? who may have come up with such an idea?
Intelligence can be said to overcome knowledge limits, ie, given limited knowledge, an intelligent process relies on logical inferences deduction, abduction to achieve a conclusion. Reasoning with limited knowledge is a defining characteristic of intelligent systems. Furthermore in open systems, knowledge is not limited, by contrast, it is continually updated with new knowledge. To consider limited knowledge for intelligent systems/AI is a contradiction in terms. A knowledge limit applies to closed database systems not to AI.

OTHER points
=======
- In addition to meaningful and accurate, explanations should also be timely, accessible, updatable etc

- (symbolic) Knowledge Representation (KR) is a mechanism for explainability should be emphasized

- this work possibly leads to a standard for explainability? would be needed, please keep me up to date

Best regards
