Consultation response: Project ExplAIn

phg foundation
making science work for health

Submitted to
Information Commissioner's Office

Submitted by
Alison Hall
alison.hall@phgfoundation.org

Johan Ordish
johan.ordish@phgfoundation.org

January 2020

The PHG Foundation is currently undertaking an extensive report, funded by the Wellcome Trust on Transparency and explanation in Black Box Medicine we have focused on the consultation questions that are most relevant to our own perspective.

General Responses

- We commend the Information Commissioner's Office for facilitating dialogue around how to meet the obligations for transparency and explanation. As the guidance suggests, meeting the obligations for transparency and explanation goes further than simply satisfying relevant legal obligations, but involves fostering trust and confidence across a diverse range of stakeholders
- In our view, the draft guidance is a good attempt to demonstrate that different explanations serve different purposes, and that in a given context a variety of different explanations are necessary. There is a wealth of detail in the guidance and we are pleased that the communication challenges of the decision recipient receiving multiple explanations throughout the process, is addressed in step 7. The emphasis on layering explanations, and ensuring that there is a continuing dialogue rather than a one way process are key, and we note that in the healthcare context, similar discussions have been had about the nature of consent to care. Delivering robust and appropriate explanation will requires a high level of staff engagement and expertise and sufficient resources, for which an institutional commitment is needed
- Our focus is on decision making in healthcare. Here, healthcare professionals have a substantial role in implementation of AI systems, and in the foreseeable future, are liable for their use. More weight could be given in the guidance to the role of professional guidance in framing the obligations of implementers for offering explanation. An example is the General Medical Council, Duties of a Doctor guidance which includes for example, the requirement for a doctor to communicate the risks associated with a treatment or intervention.

Specific responses to the Project ExplAIn consultation

Q1: the guidance provides a comprehensive grounding in how to explain AI enabled decisions to individuals. We recommend that the ICO consider making the following adaptions:

- The inclusion of a human in the decision making loop is presented as being a binary decision. In fact, the role and position of the human within the decision-making pathway is critical in determining whether there is meaningful intervention
- The human may consider 'other information' but this is not necessarily the case (p.5)
- There is some conflation of the ideas of 'accountability', 'responsibility' and 'liability' for the decision. More clarity about these terms would be helpful (p. 6)

Q3: The summaries of the GDPR are clear. The infographics are effective and informative. In the health sector, the Data Protection Act 2018 is relevant and we suggest that relevant provisions relating to transparency should be included.

Q6: Clarifying that different explanations serve different purposes is welcome. We agree that it might be helpful to developers to think through how explanations might serve these different objectives. However, we suggest that it is not always easy to distinguish between the different types of explanation. Our work on explanation focuses on the transparency aspects, but we recognise that some explanations serve multiple purposes, especially when building trust for multiple audiences.

Q7: The GDPR refers to explanations addressing meaningful information about the logic involved in the case of automated decision-making (e.g. Articles 13-15. This is mentioned in terms of satisfying the principle of transparency (p 31) but it would be helpful if the guidance addressed in more detail, what this means in practice, and how this requirement is captured within each of the explanation types described in the guidance (p 19).

The discussion of a 'rationale explanation' assumes that this is post-hoc (i.e. that led to a decision). Does the guidance exclude a rationale explanation that is proffered before a decision is made? If so, on what basis is this excluded?

We don't find the definition of 'explanation' on p.20 very helpful or informative. In particular, it does not address the requirements of an explanation specifically required by the GDPR, although it does speak to the more general obligation to provide information in an intelligible and easily accessible form using clear and plain language (GDPR Article 12).

Part 2

Q13: in the data section, the example is given of a distinction between social and demographic data being more inherently biased than biological or physical data (p 36). This distinction seems rather simplistic. Other types of biological data such as genetic/genomic data can, depending on type, nature and context, be stigmatising/ identifying and historical methods of collection mean that variant databases are biased and incomplete (p36). We would welcome a more nuanced discussion of this topic, recognising that both physical/biological and social demographic data can be both sensitive and biased, depending on context.

Q14: the tables setting out the strengths and weaknesses of different models and explainers are very helpful, as are the references.

Q24 Applying the guidance to a specific example is very illuminating. It shows the level of institutional commitment that is needed to deliver a satisfactory explanation, both to satisfy health services, professionals and managers that the use of an algorithmic tool within a system is safe, effective and trustworthy enough to be relied upon within a patient pathway, and also to provide an additional layer of explanation to patients (information recipients). In the roundtable workshops that we held, as part of the Wellcome Trust project, we explored the views of developers, clinicians and patients to different hypothetical uses of AI in health. We developed a tool to assist developers in assessing the various dimensions of their AI model, and assess the risks and benefits involved which we would be happy to discuss in more detail with ICO.

As mentioned in Q13, the distinction made between biological data and social/demographic data seems rather simplistic. In healthcare, many signs and symptoms (phenotype) are the product of both biological factors (genotype) and social factors. Whilst some physical data may be non identifying, and non sensitive this is not always the case. However describing some categories of data is being exceptionalist and always deserving of special protection may be equally misguided. A nuanced context specific approach is often required, and we suggest that the guidance should reflect this.

In general, we suggest that:

- More emphasis be put on the patient pathway. Will the AI tool be used as a first or second line test? Will the text be used for screening of healthy individuals for early detection of tumours before more detailed assessments are done, or is the imaging to stage and guide the diagnostic process? Will the imaging be used as an adjunct to other phenotypic assessments (e.g. other types of imaging; biopsy etc). The nature of the explanations required to be given will be very different depending on the anticipated patient pathway involved, and the degree of professional involvement.

- More emphasis be put on the role of professional organisations to develop codes of conduct. Multidisciplinary teams could be used to develop explanation sheets in the first instance, with strong patient involvement through PPI after early drafts have been developed.

Part 3

Q27: we note that the burden of complying with the policies and procedures described in the guidance could be considerable, and beyond the means of some smaller developers. Going forward, it will be important to develop collaborative approaches to policy development.

Q31: as a health policy think-tank, we will not be developing AI tools and implementing them ourselves. However, our current remit includes consideration of transparency and explanation in black box medicine in our projects.

Q32: As described above, the PHG Foundation is a third sector organisation.

Q33: Health

Q34: ICO blog

Subject to agreement from the funder, we would be happy to share outputs from our research project once the report has been finalised. We are happy to be contacted if there are further questions or queries about this response or our other work in this area

Contact: intelligence@phgfoundation.org