DARE UK (Data and Analytics Research Environments UK)

Name of Proposed WG: AI Risk Evaluation Group
Affiliated DARE UK Interest Group (if applicable): N/A

The WHY

Introduction

The increasing availability of neuroimaging and genomics data within Trusted Research Environments (TREs), has gained interest from researchers dedicated to the development of Artificial Intelligence (AI) models. These models hold significant promise for integration into clinical healthcare systems, presenting a transformative potential for patient care and medical practice.

However, as this field has progressed, it has introduced a heightened level of complexity in assessing the disclosure risk associated with the outputs generated by these AI models. The intricate nature of AI technology amplifies the challenges we face. Some models may inadvertently incorporate sensitive data, raising concerns about privacy and security. Additionally, the potential susceptibility of these models to attacks and other forms of breaches requires a thorough review of the safeguards in place to protect the integrity of the data and, most importantly, the privacy of the individuals who contributed to their training.

To address these critical concerns, we aim to develop comprehensive guidelines and recommendations on the responsible and ethical utilisation of AI trained on neuroimaging and genomic data within TREs. Our objective is to establish a robust framework that not only safeguards sensitive data but also ensures that the privacy of individuals involved in model training remains paramount. By bringing together a multidisciplinary team of experts in neuroimaging, data security, AI development, and clinical research, we aim to create a path forward that enables the development of these important health AI models within TREs.

This working group represents an evolutionary step forward, capitalising on the insights and groundwork laid by two cornerstone initiatives within the DARE UK program: the GRAIMATTER Sprint Exemplar Project and the SACRO Driver Project. Our specific focus now lies on the intricate realm of a range of AI models trained on neuroimaging and genomic data. The AI Risk Evaluation Group will look to build guidelines around the ethical framework, social contract, and operational approach TREs may take to enable AI model development and release.

Additionally, the Dementias Platform UK Data Portal has been carrying out a pilot project dedicated to developing an AI Privacy Risk Index. This endeavour is characterised by its inclusivity, with input and guidance gathered from a diverse range of stakeholders, including researchers, members of the public, policymakers, and data providers. This multifaceted approach, facilitated through comprehensive surveys, ensures a balanced perspective that is crucial in shaping a robust index. Clinical expertise will also be included from Cambridge and the DEMON network which are currently looking at developing clinical guidelines for the application of AI, especially for AI trained on complex data modalities such as imaging and genomics.

Building upon prior work in AI privacy and clinical applications, our project leverages existing knowledge and expertise, allowing us to progress efficiently and make meaningful advancements. This continuity in our approach maximises the translation of our project outputs into real-world outcomes and tangible impacts, particularly in the realm of responsible AI adoption in clinical practice.

The proposed working group will engage in an in-depth exploration of current best practices and regulatory frameworks surrounding AI implementation. We will produce a set of guidelines that serve as a beacon for researchers and clinicians establishing a clear roadmap for the responsible integration of AI technologies developed on complex data within TREs into clinical settings.

This will be achieved through a series of purposeful workshops, each tailored to gather insights and expertise from distinct stakeholder groups. Separate sessions will be dedicated to engaging with data providers, researchers, and the public, allowing us to discern their primary concerns and collaboratively develop robust strategies for their effective mitigation.

The AI risk evaluation Group will address the 'Capability and Capacity' theme and will look to standardise, centralise and unify processes enabling access to sensitive data for research across the UK where appropriate and feasible by assessing the risks and mitigation strategies available to enable researchers to develop AI models that can safely be removed from a TRE.

The WHAT

The AI risk evaluation group will undertake 4 workshops in the following areas:

Public/Patient Workshop: WHAT are the risks associated with AI model release?

- Overview: This workshop will bring together patients from the Great Minds project and members of the public to gather patient perspectives, concerns, and expectations regarding the use of AI. AI models will be trained on patients' data, so it is crucial that we understand how people feel about their data being used for such research. This will allow us to establish what risks people are most concerned about regarding various data modalities, AI methods, and collaborative/sharing scenarios.
- Activities: There will be a presentation to provide the public with the necessary background information on what a TRE is, what AI models are, how they work, why they are important, and what some of the potential risks may be. We will then have a discussion to gather people's opinions and they will be given a survey created for the 'AI Privacy Risk Index', to find out what they are most concerned about with their data being used in AI.
- Outcome: Patient perspectives on the risk of AI

Researcher Workshop: WHAT are the most effective mitigation techniques?

- Overview: This workshop will bring together experts in AI, neuroimaging, genomics and clinical applications to establish what the current landscape of AI methods being employed is, what privacy-preserving techniques people are aware of, whether they would be happy using these in their research and what they think the most effective mitigation strategies are to combat the risks.

- Activities:

    1. The workshop will provide a review of existing AI methodologies within the realms of neuroimaging and genomics. Experts who are developing these models with plans for implementation into clinical practice will share insights into the current methods being used and whether they have considered privacy-preserving methods.
    2. Evaluating privacy-preserving methods: participants will discuss and assess various methods available for mitigating privacy risks, and how comfortable/willing they would be for employing these methods.
    3. Identifying risks: participants will also be asked what they think are the privacy risks of implementing these AI models into clinical practice or employing methods such as federated

learning. They will also fill out the same survey for the 'AI Privacy Risk Index' as well as the DEMON network clinical guidelines survey.

- Outcome: Recommendations of best practices for mitigating privacy risks in AI and effective privacy-preserving techniques which they would be happy to use.

Data Provider Workshop: WHAT is the risk appetite of data providers?

- Overview: Data providers are ultimately the ones who own and control the data being used to train these AI models, so understanding their risk appetite is imperative to knowing what risks matter and what level of mitigations are necessary. Therefore, this workshop will be a discussion and evaluation of the risks and mitigations established in the previous workshops.
- Activities:
    - Presentation of results from the two previous workshops and discussions on the data providers views on those risks and mitigations. Whether those mitigation recommendations are suitable, if those risks are valid.
    - Understand what the barriers are which might make data providers hesitant for AI models to be trained on their data. Whether there's any restrictions relating to their ethics, or the rules for data sharing which may prevent it etc.
    - Establish risk appetite from the cohorts to get an idea of what kind of AI research they would allow to be developed with their data.
- Outcome: Risk appetite of data providers

Workshop: Developing guidelines and recommendations for TREs on assessing AI risk and implementing privacy-preserving methods

- Overview: This final workshop will bring together all members from the previous workshops to collaboratively create guidelines and recommendations for evaluating the risk to privacy in AI research and development.
- Activities: Results from the 3 previous workshops will be presented to the whole group and there will be a session on defining the final AI risks and mitigations. The group will also assess current methods for assessing AI model disclosure risk and their potential usefulness in being used in TREs to assess AI.
- Outcome:
    - Comprehensive guidelines for AI risk assessment within TREs
    - Recommendations for privacy-preserving methods for researchers to use

The WHO

The community group proposed will be composed of a multidisciplinary team of experts in neuroimaging, data security, AI development, and clinical research. Dementias Platform UK is a mature research platform with expertise in Neuroimaging and Genomic data. AI model development is already underway in DPUK utilising Neuroimaging and Genomic data and DPUK are working closely with data providers to ensure that the models can be assessed prior to release into clinical settings.

Professor John Gallacher (Oxford University) will co-chair the group and brings a wealth of expertise across all areas of TRE governance. Dr Timothy Rittman is the PI of the QMIN-MC dataset. QMIN-MC is a study

which collects real-world neuroimaging and cognitive data from NHS memory clinics from different sites in the UK. The data has been collected specifically for AI/ML analysis studies and is already being used in creating diagnostic tools to be implemented within the NHS. Dr Rittman brings the perspective of the data provider and as a clinician. Lewis Hotchkiss is the neuroimaging research officer at DPUK and has spent the past year developing the infrastructure to be able to support neuroimaging data collection and AI research. He works with key members from the DEMON network and other universities to develop AI models on data within the DPUK data portal.

The group is well placed to deliver this work as key members are already working on similar areas. Dr Rittman is leading a project in the DEMON network on establishing guidelines for the use of AI in memory clinics, working with a variety of international members to establish recommendations. Lewis Hotchkiss is developing an AI Privacy Risk Index to assess AI models which use multi-modal data within TRE's and takes into account different privacy-preserving techniques and methods for sharing AI models. Both of these ongoing projects, from the clinical side, and the TRE side, provide the groundwork for this community group.

The HOW and WHEN

Four workshops will be held over the funding period. These workshops will be both in-person and online to allow for maximum attendance.

DPUK have fostered a community of engaged Data providers and researchers over the years and many of these members have already agreed to be members of the AI Risk Evaluation Group. If successful the AI Risk Evaluation Group will also invite participants from other affiliated networks such as the British Neuroscience Association, DEMON Network, HDRUK Alliance and other TRE's. The DEMON network has a subgroup looking at developing guidelines for AI for clinical applications and have already expressed an interest in collaborating with us.

DPUK are also very active in participant recruitment to trials and through the great minds register participants of cohort studies will be invited to take part in the public workshop. The SAIL Databank Public consumer Panel will be engaged with to determine the public perception of AI risk and how they would feel about risk mitigations. A session has already been booked with panel and the results will feed into this work and members of the consumer panel have already agreed to be part of the bigger public workshop as part of this proposal. Participants of the public workshop will also be invited from other public engagement panels such as HCRW and ARUK.

	Timeline				
Activity	**Nov-23**	**Dec-23**	**Jan-24**	**Feb-24**	**Mar-24**
Public/Patient workshop					
Researcher workshop					
Data Provider workshop					
Guideline development workshop					
Guidelines/reports written					

Guidelines/Frameworks will be developed as part of the working group and will be implemented across DPUK to enable risk to be identified, assessed and for risk mitigation tools to be provided to researchers. The experiences of DPUK could then be written up and disseminated to other TRE's to allow them to utilise the lessons learned and to build on the guidelines already set out.

Potential members

FIRST NAME	LAST NAME	EMAIL	(Co-)Chair / Member
Simon	Thompson	Simon@chi.swan.ac.uk	Co-chair
Lewis	Hotchkiss	Lewis.hotchkiss@chi.swan.ac.uk	Co-chair
John	Gallacher	John.gallacher@psych.ox.ac.uk	Co-chair
Timothy	Rittman	tr332@medschl.cam.ac.uk	Co-chair

** Note: please do not hesitate to point out gaps in the current DARE UK set of strategic themes and/or recommendations that the programme should consider as it continues to evolve these. Community feedback and input is welcomed.*