International Journal of Population Data Science

Journal Website: www.ijpds.org

Enabling AI Research within a Trusted Research Environment

Lewis Hotchkiss[1,2], Emma Squires[1,2], and Simon Thompson[1,2,3,4]

[1]Dementias Platform UK Data Portal, Swansea, United Kingdom

[2]Swansea University, Swansea, United Kingdom

[3]SeRP, Swansea, United Kingdom

[4]SAIL, Swansea, United Kingdom

We aimed to develop an AI governance framework for Trusted Research Environments (TREs) to assess the risks associated with releasing AI models trained on secure data. Through stakeholder engagement, risk assessment methodologies, and technological innovations, we have enabled the responsible development and deployment of AI models within a TRE.

We hosted a series of workshops with researchers, data owners and the public to identify risks, barriers, and appropriate mitigations, to develop a comprehensive AI risk assessment framework. Insights from discussions and questionnaires informed the creation of methodologies to evaluate AI model risk before release. Additionally, we developed technical solutions for secure AI model hosting and data federation for external validation/fine-tuning, as well putting processes in place to create AI-ready data. These developments collectively support AI innovation while maintaining strong privacy and governance safeguards to keep data secure.

The AI governance framework has been successfully implemented within the Dementias Platform UK (DPUK) Data Portal, establishing a robust approach for supporting AI model research within a TRE. The framework enables risk assessments, ensuring AI models meet governance standards before being considered for release. Additionally, the development of secure AI model hosting and federated learning capabilities allows models to be externally validated and trained without exposing sensitive data. Pipelines for generating AI-ready datasets have further streamlined AI research workflows within the TRE. Early feedback from stakeholders highlights the framework's effectiveness in balancing innovation with privacy. Future work includes refining risk assessment methodologies, expanding technical infrastructure, and extending adoption.

Our framework and supporting technologies provide an approach for responsible AI research in TREs. By integrating risk assessment methodologies, secure model hosting, and AI-ready dataset pipelines, we enable AI innovation while ensuring compliance with governance controls. This work strengthens privacy-preserving AI research and supports scalable AI development within secure environments.

https://doi.org/10.23889/ijpds.v10i3.3096