There is full funding for a PhD position at the University of Glasgow to work on machine learning algorithms for protecting sensitive information in information retrieval (IR) systems.

IR systems typically assume that all information should be findable. However, the success of modern IR technologies is limiting the amount of information with societal value that is actually made available to be searched, due to the risk that a document collection may contain sensitive information. For example, government documents that must be released to the public through Freedom of Information laws, email collections and evidence in legal cases each require an extensive manual review before being made available to be searched.

The successful candidate will build on the current literature to develop novel AI methodologies and machine learning algorithms, for example to:

- Identify the information that needs to be protected in the text of documents.
- Develop techniques to prevent information leakage from search and recommendation systems, e.g. through the use of games theory or GAN neural networks.
- Formalise and automate the process for automatically protecting sensitive information.
- Develop a framework to quantify the protection that is provided by an IR System.

For example, building on the observation that GANs have been shown to be successful in developing robust algorithms by means of learning adversarial examples, the applicant may propose to leverage game-theory and penetration testing strategies to develop an evaluation framework that can quantify, and automatically improve, the levels of protection that an IR system can provide for information producers.

The work will be evaluated using a dataset containing real sensitivities that has been developed with our partners in an ongoing related project.

The successful candidate will be based within the IR Group in the University of Glasgow. The Glasgow IR Group is often considered one of the founders of modern IR. From its outset, the Glasgow IR group has focused on improving the effectiveness of IR systems, such as inventing new logic & probabilistic retrieval models, the development of adaptive query expansion techniques, interactive multimedia models and the Divergence From Randomness framework, as well as leading research into quantum, expertise search and search result diversification models. In recent years, the IR group has been at the forefront of the research, development and application of emerging machine learning and deep learning technologies for very large corpora and data streams in search and recommendation use-cases.

In the first instance, interested candidates should contact Dr. Graham McDonald (graham.mcdonald@glasgow.ac.uk) with a current CV and a motivating paragraph outlining why you would like to undertake this project. In the second instance, as part of the application process, the applicant will submit a formal application including a research proposal of 3-4 pages.
Information about the PhD application process can be found here: https://www.gla.ac.uk/schools/computing/postgraduateresearch/prospectivestudents

Application Deadline: 29th June 2020

Funding is available to cover tuition fees for UK/EU applicants for three years, as well as paying a stipend at the Research Council rates.