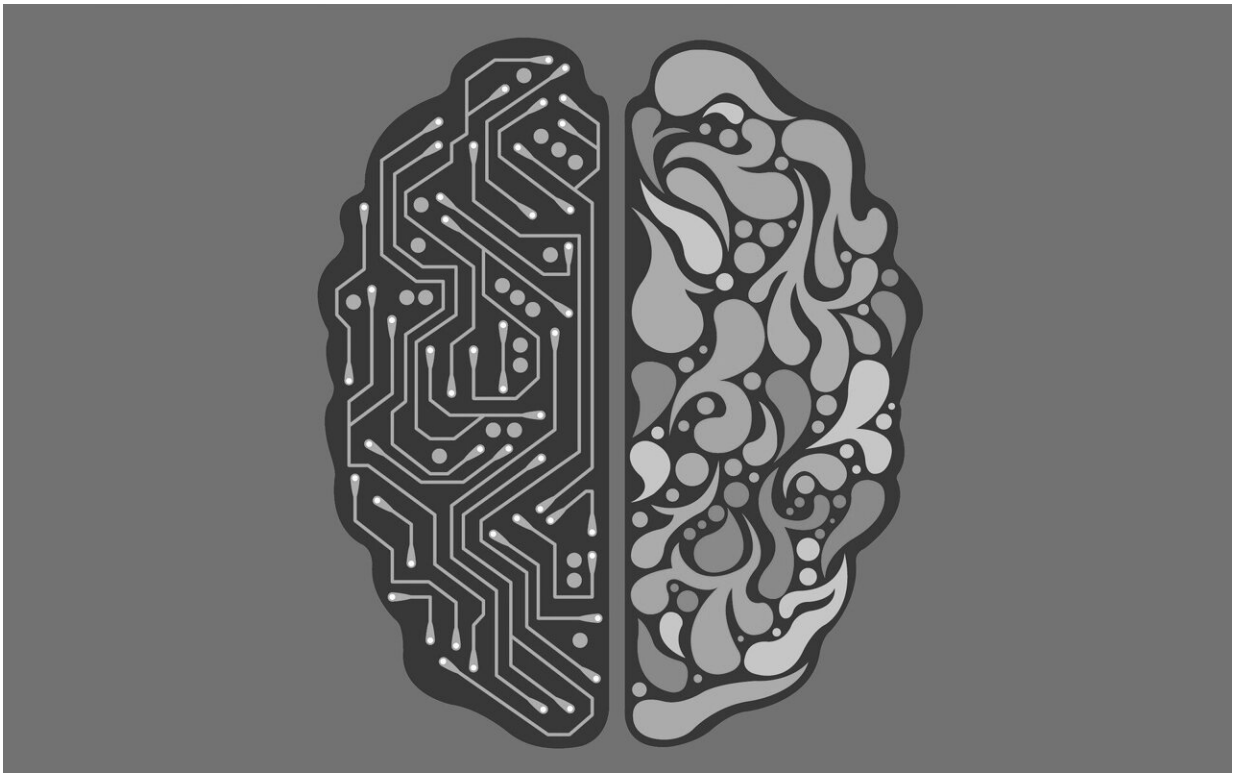
New AI model can help prevent damaging and costly data breaches

November 8 2022



Credit: Pixabay/CC0 Public Domain

Imperial privacy experts have created an AI algorithm that automatically tests privacy-preserving systems for potential data leaks.

This is the first time AI has been used to automatically discover

vulnerabilities in this type of system, examples of which are used by Google Maps and Facebook.

The experts, from Imperial's Computational Privacy Group, looked at attacks on query-based systems (QBS)—controlled interfaces through which analysts can query data to extract useful aggregate information about the world. They then developed a new AI-enabled method called QuerySnout to detect attacks on QBS.

QBS give analysts access to collections of statistics gathered from individual-level data like location and demographics. They are currently used in Google Maps to show live information on how busy an area is, or in Facebook's Audience Measurement feature to estimate audience size in a particular location or demographic to help with advertising promotions.

In their new study, published as part of the 29th ACM Conference on Computer and Communications Security, the team including the Data Science Institute's Ana Maria Cretu, Dr. Florimond Houssiau, Dr. Antoine Cully and Dr. Yves-Alexandre de Montjoye found that powerful and accurate attacks against QBS can easily be automatically detected at the pressing of a button.

According to Senior Author Dr. Yves-Alexandre de Montjoye: "Attacks have so far been manually developed using highly skilled expertise. This means it was taking a long time for vulnerabilities to be discovered, which leaves systems at risk.

"QuerySnout is already outperforming humans at discovering vulnerabilities in real-world systems."

The need for query-based systems

Our ability to collect and store data has exploded in the last decade. Although this data can help drive scientific advancements, most of it is personal and hence its use raises serious privacy concerns, protected by laws such as the EU's General Data Protection Regulation.

Therefore, enabling data to be used for good while preserving our fundamental right to privacy is a timely and crucial question for data scientists and privacy experts.

QBS have the potential to enable privacy-preserving anonymous data analysis at scale. In QBS, curators keep control over the data and therefore can check and examine queries sent by analysts to ensure that the answers returned do not reveal private information about individuals.

However, illegal attackers can bypass such systems by designing queries to infer personal information about specific people by exploiting vulnerabilities or implementation bugs of the system.

Testing the system

The risks of unknown strong "zero-day" attacks where attackers capitalize on vulnerabilities in systems have stalled the development and deployment of QBS.

To test the robustness of these systems, in a similar way to penetration testing in cyber-security, data breach attacks can be simulated to detect information leakages and identify potential vulnerabilities.

However, manually designing and implementing these attacks against complex QBS is a difficult and lengthy process.

Therefore, the researchers say, limiting the potential for strong unmitigated attacks is essential to enable QBS to be usefully and safely

implemented whilst preserving individual rights to privacy.

QuerySnout

The Imperial team developed a new AI-enabled method called QuerySnout which works by learning which questions to ask the system to gain answers. It then learns to combine the answers automatically to detect potential privacy vulnerabilities.

By using machine learning, the model can create an attack consisting of a collection of queries that combines the answers in order to reveal a particular piece of private information. This process is fully automated and uses a technique called 'evolutionary search' which enables the QuerySnout model to discover the right sets of questions to ask.

This takes place in a 'black-box setting' which means the AI only needs access to the system but does not need to know how the system works in order to detect the vulnerabilities.

Co-First Author Ana-Maria Cretu said, "We demonstrate that QuerySnout finds more powerful attacks than those currently known on real-world systems. This means our AI model is better than humans at finding these attacks."

Next steps

Presently, QuerySnout only tests a small number of functionalities. According to Dr. de Montjoye, "The main challenge moving forward will be to scale the search to a much larger number of functionalities to make sure it discovers even the most advanced attacks."

Despite this, the model can enable analysts to test the robustness of QBS

against different types of attackers. The development of QuerySnout represents a key step forward in securing individual privacy in relation to query-based systems.

More information: Confrence: www.sigsac.org/ccs/CCS2022/

Provided by Imperial College London