Health AI Research Scenario: AI-Based Cancer Genetics Test

Someone close to you has recently been diagnosed with cancer. Their doctor orders a biopsy of the tumour to analyze the cancer cell DNA and several other tests to help her determine the best options for treatment. The data generated by these tests could be used in artificial intelligence (AI) research studies.

AI research could help doctors and patients have better cancer treatment options in the future. Researchers are using AI to analyze very large amounts of data to identify patterns that tell them where a tumour originated in a patient's body. In some cases, AI can help identify where a tumour came from faster or more accurately than other methods. If doctors know where a tumour started in the body, they might recommend different treatments. For example, in the future, doctors might use an AI test to determine that the cancer found in someone's ovary originated in the colon, and the doctor might recommend a colon cancer drug for treatment as a result.

AI is just one of the methods that is being used to study cancer and improve treatments. There is no guarantee that the AI study will provide benefits for the person who is close to you, or to anybody else.

If your friend or loved one does decide to provide their data to the research study, identifying information like their name, phone number, address, and health card number would all be removed before the data are made available to researchers. All people with access to the data commit to not attempting to re-identify any person in the dataset. The risk of re-identification would be very low, but is never zero, particularly when genetic information is involved, because every person's DNA is unique.

[Discussion prompts, e.g., Do you think it is appropriate to use health data for AI research in this way? Why? Why not?]

Health AI Research Scenario: AI-Based App to Help Older Adults Aging at Home

A group of private and public sector organizations want to use artificial intelligence (AI) research to develop an app that helps older adults self-manage chronic diseases like diabetes so that they can age independently in their own homes.

The research team includes AI scientists from universities and hospitals, physician researchers, app developers from a small company and staff from a not-for-profit organization that provides home care services for seniors. The research team plans to use AI to identify patterns in large amounts of data that help them predict when older adults are most likely to run into trouble living at home. They want to use the results of the AI research study to develop an app that provides advice and directs seniors to services before problems become serious.

The data they are using for the AI research study includes:

- information that older adults have already entered into websites and apps themselves, e.g., information they typed on social media platforms that members of the public can read
- data that the home care services not-for-profit organization gathers, e.g., how much help someone needed with bathing
- physician notes from family doctor's offices, e.g., notes about how a person's chronic condition appears to be affecting their mental health and well-being

In all cases the people who have data included in the study were informed that their data may be used for research, but they may not be fully aware of it. For example, people may have clicked "I accept" to terms on a website, or signed a form with fine print without reading all the terms. In other cases, people have been in an office that has a poster on the wall stating that data may be used for research, but they didn't notice the poster or completely process the fact that their data would be used for a research study like this.

If things go as planned, the AI research studies will lead to an app that will help older adults with self-management and direct them to healthcare services when they need them. For example, the app might send medication reminders to an older adult with diabetes and mild depression if the person's condition is stable, or suggest that they make an appointment to see their doctor if their condition is worsening. There is no guarantee that the AI research study will lead to an app being developed, or that the app will be effective.

[Discussion prompts, e.g., Do you think it is appropriate to use health data for AI research in this way? Why? Why not?]

Health AI Research Scenario: Sample Dataset with Lab Test Results for AI

A group of not-for-profit research institutes is planning to create a health dataset that can be used for artificial intelligence (AI). The research institutes have large datasets that are strictly controlled for use in approved research studies. Their goal is to create a small sample of one of these datasets that can be accessed with fewer restrictions than the larger, controlled health dataset.

The research institutes will start with a large dataset that has laboratory test results and basic information about hospital visits for over 10 million people that live in Ontario. They will take a random 2% sample to create the "sample dataset" which means that it would include data for about 200,000 people out of the 10 million people. It also means that out of every 100 people who had lab tests performed in Ontario, two people will have their data included in the sample dataset. However, there is no way for the people who are included in the sample dataset to know that their data are being used. Your data could be in the sample dataset, but you wouldn't know it.

The sample dataset will have all identifying information removed, including names, dates, geographic information and any details about services that might allow someone to re-identify an individual. In addition, the research institutes will remove all the data for people with rare conditions because those people might be easier to re-identify. The sample dataset will include data for people that are healthy and data for people who have common chronic diseases, like diabetes and high blood pressure. An external organization will certify that the sample dataset is de-identified before anyone is permitted to work with the data. Once the sample dataset is certified as "de-identified", it will be put in a controlled environment where researchers, students and companies can access it once they have completed training. They will not be able to download the data.

With AI, it is sometimes hard to know what the benefits of data analysis will be in advance. The benefits of the sample dataset could be:

- Researchers use the sample dataset to learn about new relationships and patterns in the lab test data. For example, they might have a new discovery which shows that a patient is at risk when two different lab tests increase at the same time.
- A large number of students work the sample dataset to learn and improve their computer programing skills.
- Companies use the sample dataset to develop apps that patients use to track their own laboratory test results over time.

It is possible that none of these benefits of the sample dataset are realized, or that the sample dataset has other benefits that the research institutes haven't thought of.

[Discussion prompts, e.g., Do you think it is appropriate to use health data for AI research in this way? Why? Why not?]