JUTA
POPIA PORTAL

ISSUE NO 1 • TO SCRAPE OR NOT TO SCRAPE PERSONAL INFORMATION • JUNE 2022

TO SCRAPE OR NOT TO SCRAPE PERSONAL INFORMATION?

1. OVERVIEW

The Information Commissioner's Office (ICO) recently fined Clearview AI Inc £7,552,800 for collecting images of people from the UK from the internet and creating an online database with the images.

The ICO also issued an enforcement notice ordering Clearview to stop processing the personal information of UK residents that is available on the internet and to delete said information from its systems.

This must serve as a wake-up call for organisations scraping the web for personal information that they think is 'publicly available.'

Photo: Pexels

2. WHY DID I GET FINED FOR COLLECTING IMAGES ON THE INTERNET AND CREATING A DATABASE?

Clearview collected more than 20 billion images of people's faces and information from the internet (including social media websites) to create an online database.

Clearview provides a service that allows its customers (including police) to upload an image (input image) of a person to Clearview's app. The app checks for matches against all the images in the database. The images, metadata and URLs on the Clearview database constitute personal information.

When Clearview obtains images from the internet, stores them in their database, and matches input images with their database, they process personal information.

Organisations try to justify scraping the internet by arguing that the information is already public or that the data subject deliberately made the personal information public.

Clearview alleged that individuals voluntarily put their images on the internet, which means they can process this information as they please. The ICO reiterated that third parties could upload images of the individuals, and individuals can upload photos of themselves and restrict the image's privacy immediately or later.

3. WHAT DO YOU NEED TO CONSIDER?

As a general rule of thumb (with some exceptions), section 12(1) of POPIA states that organisations must collect personal information directly from the data subject.

However, organisations do not have to collect information directly from the data subject if the personal information is contained or derived from public records or has deliberately been made public by the data subject.

Firstly, the internet is not public record. Secondly, if the responsible party scrapes the internet for images, the responsible party will need to prove that the data subject deliberately made the personal information public. This is difficult for the responsible party, as the evidence to prove this is often in the data subject's possession.

It is a good idea to avoid collecting personal information if there is any protection mechanism to prohibit the general public from accessing personal information.

Most importantly, you need to ensure that you have the correct legal justification when processing personal information or special personal information (a picture of your face will fall under biometric information).

You also must ensure that you make the data subject aware that you are processing their personal information. You also need to conduct a personal information impact assessment before you start scraping the internet for personal information and special personal information.

4. WHAT YOU SHOULD READ NEXT

Chapter 10 on collecting and creating personal information provides in-depth explanations of all the exceptions under POPIA to the 'direct collection rule' for personal information.