Privacy and Security in the Digital Age

The emergent ethical and security concerns arising from publicly available facial recongition technology

Introduction

The first quarter of the twenty-first century has been characterised by rapid advancement development and innovation. The maturation of deep-learning algorithms has brought into reality concepts that previously existed only in the imagination. However, with great power comes great responsibility. As powerful new technologies are developed, and existing technologies are refined, novel privacy and security concerns arise. Since the advent of social media, with the creation of MySpace in 2003, a growing number of individuals have uploaded both photographs of themselves, and sensitive personal information to the internet. Much of this personal information would have been uploaded by users of an age unable to fully comprehend the consequences of doing so. Unfortunately, the statement that "nothing gets deleted on the internet" has proven to be remarkably accurate.

Internet scrapers are automated algorithms that crawl the internet and save data to local storage. Almost half of internet traffic now comprises scraping bots (Conde and Svantesson 2024). Scrapers do serve legitimate purposes; search engines, for example, require scrapers to discover and index as many websites online as possible. These scrapers discover new websites by going from link to link and recording the metadata associated with each page. Similarly, archiving services such as the Internet Archive use scrapers to save a 'copy' of pages on the internet at different time-points for historical purposes. However, scrapers can also be used for more nefarious purposes. Personal data has value, and the internet contains a lot of it. Data brokers and other businesses are now scraping the internet specifically for personal or identifying information that can be sold to third parties or the public.

In the most consequential example of this phenomenon, ClearView and PimEyes are two companies that essentially offer a facial search engine. ClearView is only available to government services and private business, however PimEyes can be used by any member of the public for a fee. These companies have performed large-scale indiscriminate scraping of facial photographs and videos from numerous different internet sources, including social media, news sites, video hosting websites (such as YouTube), forums and pornographic websites. They then use deep-learning facial recognition algorithms to create a 'faceprint' of each individual. These algorithms measure features such as the spacing of the eyes, width and height of the nose, depth of the eye sockets, shape of the eyes, and shape of the mouth, to create a unique numerical code that corresponds to an individual's face. They then store each faceprint with its associated images. When a user uploads a picture of another person, the algorithm will generate a faceprint for the pictured individual, and match it to a faceprint within the system's database with high similarity. The website will then return all photographs that the pictured individual has posted online, in addition to the websites from which they were obtained, and potentially sensitive personal data on those websites.