Public Data: Technical

What is “Public Data”?

- Data collected for commercial purposes
  - Subjects have sometimes nominally consented, e.g., via privacy policies
  - Other data is purchased, often with consent being solicited
- Access or use is purchased by law enforcement

Many Types!

- Location (already discussed)
- License plate readers
- Facial recognition
- DNA
- Data brokers
- Credit scores?

Automated License Plate Readers (ALPR)

- Many private companies are using ALPRs
  - Locate cars for repossession
  - “Protect” neighborhoods or gated communities
  - Privately operated toll roads
  - Contractors operating red light cameras

Police Use of ALPRs

- Some police cars have license plate readers
  - Spot cars on a “hot list”
  - Build up records of what cars are where, when?
- Find particular, wanted individuals
- Police data isn’t as comprehensive as the private sector’s—so they buy the data

Overriding Legal Issue

- Should legal process—a judge's authorization—be needed to acquire such data?
- Is this like *Knotts*—movement in public—or is this like *Carpenter*—too much collection and use of data not voluntarily surrendered?

Facial Recognition

- Law enforcement often wants to map a face to a name
- Example: surveillance video of a crime
- Today's machine learning-based algorithms can do facial recognition at scale —but where does the data come from?

Image Recognition at Scale, 2014

(https://xkcd.com/1425/)

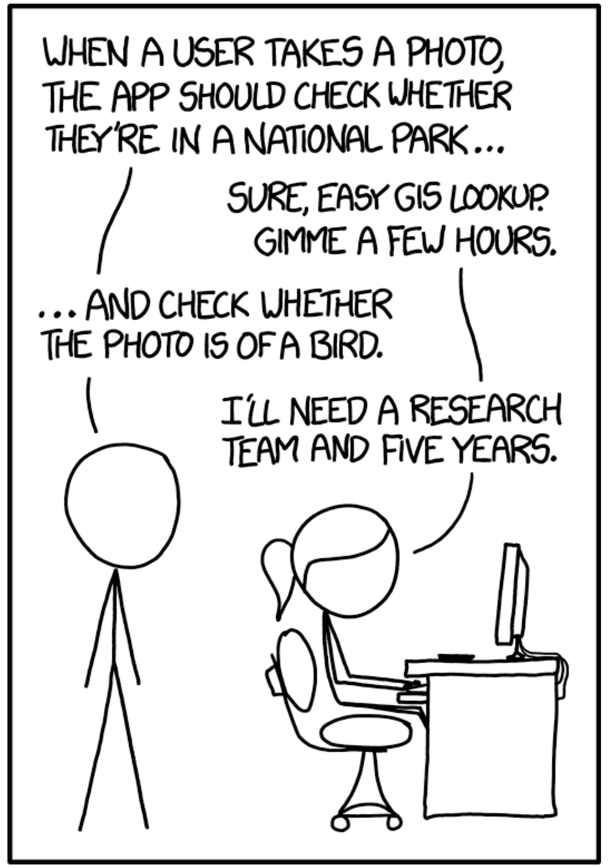

Image Recognition Today

Google identifies the type of bird

Image size:
2071 × 3102

Find other sizes of this image:
All sizes - Large

Possible related search: ***barred owl***

www.allaboutbirds.org › guide › Barred_Owl

Barred Owl Identification, All

Barred Owls are mottled brown and white over

underparts are mostly marked with vertical brov

Facial Image Sources

- Mug shots
  - Not a privacy issue, though correctness matters
  - Facial recognition algorithms don’t work as well on women and on darker-skinned faces
- Driver license database
- Purchased
- “Scraped” from websites

Facebook's Terms of Service

> 3. You may not access or collect data from our Products using automated means (without our prior permission) or attempt to access data you do not have permission to access.

- Are these terms enforceable?
- Is "scraping" a violation of the Computer Fraud and Abuse Act (18 U.S.C. §1030)
- (Case currently before the Supreme Court—oral arguments were Nov. 30)

DNA Matching

- DNA-matching is a powerful forensic technique
  - Note—many different types, with different properties
  - (Must be done carefully—risk of sample contamination)
- Public DNA sequencing is also popular—learn of your ancestry, find relatives, perhaps learn of medical risks
  - There are large databases of voluntarily submitted DNA samples, e.g., *23andMe*'s
- What happens when we combine the two?

DNA and Family

- People inherit 50% of their DNA from each parent
  - They share *approximately* 50% with siblings
  - They share *approximately* 12.5% with first cousins
- DNA can control—or influence—physical features (*phenotypes*)
  - Relatives sometimes physically resemble each other

Combining the Two

- Police can recover forensic DNA from a crime scene
- They can look for exact matches in law enforcement DNA databases—or they can search for partial matches in commercial databases
- *This lets them find relatives*
- In other words, the privacy decision of someone uploading their DNA to a commercial site affects others

Phenotype Reconstruction

- Some companies are trying to go a step further—using the DNA to construct an image of someone's face
- Dubious science—there's too much unknown about the genetics of faces
- Even skin color, a much simpler issue, isn't well understood except for the extremes
- Combine the errors in facial recognition with the errrors from dubious DNA interpretation, and violate privacy besides…

Data Brokers

- Not the same as credit reporting agencies—those are regulated
- Data brokers collect—and analyze and sell—thousands of data points on as many people as they can
  - Find public records, buy data from others, merge
  - Merging accurately is hard without a common, stable database key
- Yes, the data can contain errors—and unlike with credit agencies, there is no obligation to provide access or opportunity for correction

Uses

- Raw data is analyzed via ML; individuals are sorted into categories
  - Used for marketing—and by political campaigns
  - (Political use goes back to the late 1950s!)
- Knowledge-based authentication
- Credit scores
  - Note: credit scores *are* used in setting auto insurance rates
- Does law enforcement use this data? Unknown—but the theft of data on 150M people has been attributed to a foreign intelligence agency

What is the Real Problem?

- Use by law enforcement?
- Use by law enforcement without a warrant?
- Collection of the data by private parties, without adequate notice or consent?
- Sale to more or less anyone?

Private Abuses of Public Data

Dr. Anthony Fauci, to the *New York Times*:

> It was the harassment of my wife, and particularly my children, that upset me more than anything else. They knew where my kids work, where they live. The threats would come directly to my children's phones, directly to my children's homes. How the hell did whoever these assholes were get that information? And there was chatter on the internet, people talking to each other, threatening, saying, "Hey, we got to get rid of this guy. What are we going to do about him? He's hurting the president's chances." You know, that kind of right-wing craziness.

Daily Bird

(Great blue heron, Central Park, January 22, 2020)