Predicting sensitive personal attributes from [*only*] Facebook Likes
University of Cambridge 2013

The study was based on data from 58,466 users from the United States.
Based on an average of 170 likes, the study was able to predict a handful of personal attributes with varying majority accuracy.

Predicted attribute	Prediction accuracy
Ethnicity – "Caucasian vs. African American"	95%
Gender	93%
Gay?	88%
Political views – "Democrat vs. Republican"	85%
Religious views – "Christianity vs. Islam"	82%
Lesbian?	75%
Smokes cigarettes?	73%
Drinks alcohol?	70%
Uses drugs	65%
Single or in a relationship?	67%
Were the parents still together at 21?	60%

Notes: "Only a "few users were associated with likes explicitly revealing their attributes". For example, "less than 5% of users labeled as gay were connected with explicitly gay groups" such as "Being Gay", "Gay Marriage" or "I love Being Gay." Predictions rely on less obvious, but more popular Likes such as"Britney Spears" or "Desperate Housewives" – which proved to be weak indicators of being gay"(15). Further, it should be noted that modern social media companies seldom stop at surface level data collection, such as Facebook likes, and instead incorporate data such as location, engagement frequency and time, device hardware and attributes, mouse movement, network information, third-party partner data, and more.