

By continuing to use this site you consent to the use of cookies on your device as described in our [cookie policy](#) unless you have disabled them. You can change your [cookie settings](#) at any time but parts of our site will not function correctly without them.

---

[Sign In](#)[Subscribe](#)

---

## Facebook

# With big data comes big responsibility

The fact that personal records reveal so much could deter online services users, says Michal Kosinski

MARCH 14, 2013 by: Michal Kosinski

Big data is subject to much hype. The ability to manipulate vast swaths of information at warp speed is transforming businesses across the world. But as with all technologies there are risks. All users of technology should take note: with **big data** (<http://lexicon.ft.com/Term?term=big-data>) comes big responsibility.

A recent [study I jointly led \(http://www.pnas.org/content/early/2013/03/06/1218772110.full.pdf+html?sid=1b67721e-32d0-4efa-a486-67c018015f53\)](http://www.pnas.org/content/early/2013/03/06/1218772110.full.pdf+html?sid=1b67721e-32d0-4efa-a486-67c018015f53) concludes it is possible to create instantaneous and surprisingly detailed psycho-demographic user profiles – containing statistically valid information about an individual’s race, personality and IQ scores, happiness, substance use, sexuality, political views and religious beliefs – using only publicly available Facebook Likes.

Once you have the data, building a model that updates automatically is simple. We used Likes and personal attributes extracted from 58,000 Facebook profiles and measured by our own questionnaires – and we had no reason to think the results were not representative.

Predictions did not depend on small sets of obvious Likes. While an interest in science is associated with higher intelligence, a preference for curly fries or a fondness for Morgan Freeman’s voice are similarly informative. Aggregating thousands of such data can lead to powerful inferences at the individual level.

Facebook is just the start. Likes are one example of a generic class of digital record from which predictions can be made, such as tweets, emails, web searches, browsing histories, credit card transactions, and purchases made both on and offline.

Like any great technology, these predictive powers can be **used for good – or not** (<http://next.ft.com/content/09c8172c-8a45-11e2-bf79-00144feabdco>).

Fast and automated psychological assessment could revolutionise recruitment. Why not assess, with consent, millions of applicants before inviting a final few for interview? This

could save time and money for recruiters and applicants. Why not automatically adjust products and services to fit an individual's profile? Imagine the FT targeting online articles on the basis of personality and mood. A "nights out in London" search could yield different results for open-minded extroverts and conservative introverts.

There is, of course, another side. While personalised advertising might be viewed as beneficial to both users and advertisers, tipping the balance of power towards the industry might enable manipulation of customers. An emotionally unstable user could be pushed towards buying unnecessary insurance based on their psychological make-up. Predictability of certain traits may even prove dangerous for individuals. It is already possible for a user's sexual orientation or religion to be exposed, compromising their safety – and not just in less liberal countries.

The realisation that playlists, shopping records and Likes can be used to reveal so much could deter many from using online technology. I do not believe such digital exclusion is a good thing – for individuals or economies. There is huge potential in predicting individual traits and preferences. I am no policy maker but I believe we should design policies and tools to minimise associated risks. Two principles should guide us: transparency and control.

First, we need to help users understand which of their personal data are out there, how they are being used and how they might be used. Second, we need to enable users to take full control of their data and decide how it is to be used. Both aspects may have technological solutions but also require user awareness and appropriate legal structures.

Users should have full control over data that can be used to make these inferences. It is widely accepted now that your data are stored and governed by third parties, such as corporations and governments. But does it have to be this way? Imagine a social network or online shop that does not store Likes or purchase records. These are safely stored in your computer or personal cloud account. Predictions could still be made, but under an individual's control, allowing people to approve resulting personal inferences.

I love Facebook. It is a truly great technology that brings people together. I want to help ensure we can keep using it in the knowledge that personal information is safe and secure.

*The writer is a researcher at Cambridge university's Psychometrics Centre. He co-authored the study on private traits with David Stillwell, also of the Psychometrics Centre, and Thore Graepel at Microsoft Research*

---

Print a single copy of this article for personal use. Contact us if you wish to print more to distribute to others. © The Financial Times Ltd.

