



UvA-DARE (Digital Academic Repository)

Speculative data selfies

van der Vlist, F.; Helmond, A.

Publication date

2017

Document Version

Final published version

License

CC BY

[Link to publication](#)

Citation for published version (APA):

van der Vlist, F. (Author), & Helmond, A. (Author). (2017). Speculative data selfies. Web publication or website, Internet Policy Review.

<https://policyreview.info/articles/news/speculative-data-selfies/449>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



Speculative data selfies

01 Mar 2017 by *Fernando van der Vlist, Anne Helmond*
on *Data collection*

This short piece examines [Data Selfie](#), an open-source Chrome browser extension that collects and analyses data about your behaviour on facebook.com. The extension has been developed by [DATA X](#) (Hang Do Thi Duc and Regina Flores Mir) and was originally announced [in April 2016](#) as an iPhone app and a Chrome browser extension. The latter was released in January 2017 and has since gained traction from technology blogs (e.g., [Fast Co. Design](#), [The Next Web](#), [Mashable](#), and [Big Think](#)) and has already been installed by over 70,000 users ([Chrome Web Store](#)).

In this piece we briefly describe how Data Selfie works, what data it does and does not collect, what kind of data profiles or "selfies" it generates, and what type of data awareness raising strategy it employs.

WHAT IS DATA SELFIE AND HOW DOES IT WORK?

As the developers describe the extension on the project's website,

Data Selfie is a browser

extension that tracks you while you are on Facebook to show you your own data traces and reveal how machine learning algorithms use your data to gain insights about your personality. ([Data Selfie, Home](#))

Once the user has installed the browser extension, it runs in the background as you visit and use facebook.com on your desktop computer. Meanwhile, users see a simple counter in the bottom-left corner of the browser interface that tracks the number of seconds they are spending on Facebook. This counter is reset each time the user refreshes the page or visits other users' Timelines or Pages. In addition, users can view their data "selfie" from within the browser and can export, import, and delete their data.

The extension tracks both explicit and implicit forms of participation on facebook.com. While "explicit participation" is driven by user motivation and entails clicking, liking, sharing, and posting, "implicit participation" refers to the capture of social interactions and user activities, which "are channelled and controlled by design" (Schäfer, 2011, p. 51).

Specifically, the extension tracks:

clicks on likes in your newsfeed, clicks on newsfeed links to external sites, duration spent on different posts and the specifics of those posts (authors, images and text) in your newsfeed, anything you type, and time spent on Facebook overall.

(Data Selfie, FAQ)

Both types of data are captured from Facebook's "front-end", or the user interface where participation takes place (Stalder, 2012) as opposed to Facebook's "back-end", or its data infrastructure, some of which can be accessed programmatically via Facebook's APIs. Data Selfie's developers declare they do not utilise Facebook's APIs. Instead,

What we do is look at the rendered front-end (the DOM) of the Facebook page in your browser – the HTML elements and their content – and with JavaScript the browser extension can detect when you scroll, when you type and when you click something. **(FAQ)**

Yet many of Facebook's advertisers and marketing partners do have programmatic access to Facebook's backend and use this to create detailed profiles.

WHAT IS A DATA "SELFIE"?

Data Selfie's data is aggregated and presented in the form of a dashboard composed of multiple data plots, ranked lists, and predictions based on different types of captures – impressions ("looked") and explicit actions ("liked", "link clicked", and "typed") – which are colour-coded accordingly. It is presented as white text in a monospaced font on a black background, resembling a typical classic terminal. Some activity-based data like top friends, top pages, and top likes are updated in real time, whilst other data first needs to be processed by machine learning APIs – namely Apply Magic Sauce and IBM Watson ([FAQ](#)). After the data is processed, predictions are updated within the dashboard at a regular interval. The dashboard thus provides a fragmented view of the collected data in more than one sense.

The notion of the data "selfie" is itself embedded in a much longer history of related notions about quantified selves developed since the early 1990s. For example, Mark Poster wrote about an

“additional self” created in databases (1990), and others about a “data image” (Lyon, 1994), a “digital persona” (Clarke, 1994), and a “data-double” (Haggerty & Ericson, 2000). Whilst the term “selfie” is commonly used to refer to self-portrait photographs the term is used here to refer to visual data self-portraits and constitutes a way “to see ourselves through technology” (Rettberg, 2014), or as quantifiable selves.

WHAT DATA DOES DATA SELFIE NOT CAPTURE?

Although the extension collects much data already, it still only represents the tip of the iceberg. There is a lot of data that it currently does not collect, including data from mobile devices, external sources, and historical data.

A first type of data that is missing comes from mobile devices. The extension collects diverse data from your interactions on facebook.com, however as Facebook has stated in its latest quarterly results, “the vast majority of monthly and daily usage now occurs on mobile devices” ([Q4 '16 Earnings Transcript](#)). Data Selfie does not collect data about user interactions while using any of Facebook’s mobile apps. Although the developers acknowledge this limitation, they claim “there is no way for

us to access Facebook’s native mobile application” (FAQ). In fact, they argue “that is a good thing”, given the far-reaching privacy implications associated with data collecting extensions. For the time being, it is still harder to capture data generated in mobile environments in contrast to desktop, which may indeed be a good thing.

A second source of missing data is generated via third-party integrations of Facebook’s services on external websites and apps. Over recent years, Facebook’s platform has significantly expanded across the web and mobile apps via social plugins such as Like buttons, comment plugins, and Facebook Login (Gerlitz & Helmond, 2013; Helmond, 2015). As a result, Facebook data is not only being produced on its own website and apps anymore, but is also generated via external websites and apps. This effectively means that the extension misses out on a vast amount of data generated by Facebook users.

A third type of missing data is historical Facebook data. Although it is possible to import data from previous sessions, the extension only starts tracking once installed on a user’s device. This means that its results, projections, and predictions are not based on historical

Facebook user activities. This is important to acknowledge since the accuracy of predictions generally improve significantly as more historical user data becomes available. The developers have announced plans to support importing copies of user's own Facebook data (see [Downloading Your Info](#)).

HOW DOES DATA SELFIE GENERATE DATA PROFILES?

Data Selfie thus only provides partial insights into the data collection practices of Facebook, which utilises a large variety of data sources for creating profiles and projections. Moreover, Data Selfie's data profiles and projections (i.e., "predictions", "orientations", and "preferences") are not based on Facebook's predictions but on IBM Watson's. Watson's results are then used to provide "a sense of what Facebook could know about your personality" ([FAQ](#)). Since we are presented with Watson's predictions and not Facebook's it remains unknown what Facebook, and its advertisers and marketing partners, actually know about you.

Considering the broader implications of data profiling, Data Selfie does provide intriguing and detailed insights about data profiling in general and how

personal profiles can be generated from Facebook data. In a related study, Michael Kosinski et al. have famously demonstrated that Facebook Likes can be used to

accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. (2013, p. 1)

Some of Watson's other results are based on sentiment analysis, "the computational treatment of opinion, sentiment, and subjectivity in text" (Pang & Lee, 2008, p. 1) employed to classify a text's polarity (e.g., positive or negative). However, there are still many challenges and limitations to the use of such analytical techniques (e.g. Pang & Lee, 2008). This can result in uncanny data selfies. For example, one of the author's selfies was showing a slight positive sentiment towards the entity "Donald Trump" whilst the author's feed was filled with critical articles about Trump.

Even though sentiment analysis' numerical outcomes are precise they are often inaccurate. As a result, users are made aware of the limitations of these analytical techniques for classification and prediction. Yet regardless of their accuracy, the outcomes of sentiment analysis (usually positive or negative numerical values) can be used for calculating predictions, including detailed targeting for serving relevant ads. Consequently users may see irrelevant ads or content occasionally.

Similar to Facebook's profiling practices, Data Selfie also groups users based on similarities. Facebook's detailed targeting interface enables advertisers to create custom audiences or alternatively to "define your audience by including or excluding demographics, interests and behaviors" ([Ads Manager](#)). Data Selfie similarly uses group characteristics based on the familiar "Big 5" personality traits model – which is not without its problems (e.g., Boyle, Matthews, & Saklofske, 2008) – to identify a user's likeliness (i.e., probability) to purchase a certain type of product or to have a certain kind of preference. Crucially however, Facebook also integrates external data sources from its partners, including offline consumer behaviour data and lifestyle attribute datasets

(Acxiom).

DATA SELFIE'S AWARENESS RAISING STRATEGY

The extension aims "to create awareness in society about the erosion of privacy" (FAQ). This kind of erosion exceeds an individual's privacy settings where users can decide what to share and with whom. Instead, it addresses how user activities can be aggregated and transformed into valuable profiles.

As a privacy awareness project, Data Selfie is part of broader genre that includes tracker and ad blockers, privacy plugins, and alternative mobile browsers (e.g., Adblock Plus, AdNauseam, Disconnect, Ghostery, Go Rando, Lightbeam, and Privacy Badger). These projects typically raise awareness by blocking trackers, exposing infrastructures or mechanisms, and obfuscating or confusing algorithms – for example by automating ad clicks or camouflaging feelings for sentiment analysis. These are all different forms of "data activism" that address the socio-political consequences of datafication (Milan & Van der Velden, 2016).

Data Selfie is not only successful in raising awareness about the collectability of massive social data, but also, and in

particular, in demonstrating the artificiality and boundless calculability of a user's behavioural profile and its value to the platform and its advertisers. Its most effective contribution, in our view, lies in displaying and confronting users with social media companies' practices of data collection, calculation, and potential profiling by means of a speculative data awareness raising strategy.

References

Boyle, G. J. (2008). Critique of the Five-Factor Model of Personality. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment: Personality Theories and Models* (pp. 295–312). London: SAGE Publications.

Clarke, R. (1994). The Digital Persona and Its Application to Data Surveillance. *The Information Society*, 10(2), 77–92.
doi:[10.1080/01972243.1994.9960160](https://doi.org/10.1080/01972243.1994.9960160).

Gerlitz, C., & Helmond, A. (2013). The Like Economy: Social Buttons and The Data-Intensive Web. *New Media & Society*, 15(8), 1348–1365.
doi:[10.1177/1461444812472322](https://doi.org/10.1177/1461444812472322).

Haggerty, K. D., & Ericson, R. V. (2000). The Surveillant Assemblage. *The British Journal of Sociology*, 51(4), 605–622.
doi:[10.1080/00071310020015280](https://doi.org/10.1080/00071310020015280).

Helmond, A. (2015). The Platformization of The Web: Making Web Data Platform Ready. *Social Media + Society*, 1(2).

doi:[10.1177/2056305115603080](https://doi.org/10.1177/2056305115603080).

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private Traits and Attributes Are Predictable From Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. doi:[10.1073/pnas.1218772110](https://doi.org/10.1073/pnas.1218772110).

Lyon, D. (1994). *Electronic Eye: The Rise of Surveillance Society*. Minneapolis: University of Minnesota Press.

Milan, S., & Van der Velden, L. (2016). The Alternative Epistemologies of Data Activism. *Digital Culture & Society*, *2*(2), 57–74. doi:[10.14361/dcs-2016-0205](https://doi.org/10.14361/dcs-2016-0205).

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135. doi:[10.1561/1500000001](https://doi.org/10.1561/1500000001)

Poster, M. (1990). *The Mode of Information: Poststructuralism and Social Context*. Chicago: University of Chicago Press.

Rettberg, J. W. (2014). *Seeing Ourselves Through Technology: How We Use Selfies, Blogs and Wearable Devices to See and Shape Ourselves*. New York: Palgrave Macmillan.

Schäfer, M. T. (2011). *Bastard Culture!: How User Participation Transforms Cultural Production*. Amsterdam: Amsterdam University Press.

Stalder, F. (2012). Between Democracy and Spectacle: The Front-End and Back-End of the Social Web. In M. Mandiberg (Ed.), *The Social Media Reader* (pp. 242–256). New York: New York University Press.