IF YOU'RE NOT PAYING FOR IT YOU ARE THE PRODUCT - A LESSON SERIES ON DATA, PROFILES, AND DEMOCRACY

Bettina Berendt, Gebhard Dettmar
KU Leuven, Department of Computer Science, Celestijnenlaan 200A, 3001 Heverlee, Belgium
Behörde für Schule und Berufsbildung, Felix-Dahn-Straße 3, 20357 Hamburg
bettina.berendt@cs.kuleuven.be, g.dettmar@web.de

We developed and performed (in school) a lesson series that covers the effects of Internet-tracking and the data analysis of the data collecting industry, Facebook, Google & Co., on a political system that ranks autonomy and privacy among its basic rights.

Various court judgments (including, in Germany, the Lüth Judgement by the German Constitutional Court) have established that these rights are positive fundamental rights that are indispensable for political participation, and that therefore bind not only state actors, but also actors such as companies.

Accordingly, the series introduces, step by step, data protection related issues such as tracking and targeted advertising, data mining and social exclusion (health insurance / credit frame as examples), as well as psychometry applied to Facebook likes ("Big 5 Personal Traits" plus IQ). Based on this, the series deals with the consequences for privacy, their function for the possibilities of participation within democracy and the legal guarantees attached to it by the BVerfG.

We report on genesis, concept and experience with the implementation of this series and will further discuss developments since the first iterations of the lesson series and sketch how these could be integrated, in particular the debate around filter bubbles and fake news (on Facebook and other social media).

THE LESSON SERIES IN DETAIL

The lesson series starts with the visualisation of internet tracking via browser-plugins, thus revealing the core business of the so called data leech industry: data collection, data evaluation, profile creation. These profiles are a valuable product which is sold to other companies. A data collection company has to inform its users about this in a privacy policy which is a part of the terms of service.

Here is an excerpt from Facebook's data use policy, status quo 2013[1]: "We do not share any of your information with advertisers (unless, of course, you give us permission). As described in this policy, we may share your information when we have removed from it anything that personally identifies you or combined it with other information so that it no longer personally identifies you.[2] […] We use the information we receive, including the information you provide at registration or add to your account or timeline, to deliver ads and to make them more relevant to you. This includes all of the things you share and do on Facebook, such as the Pages you like or key words from your stories, and the things we infer from your use of Facebook."

The striking term in this excerpt is "things we infer from your use of Facebook". The main purpose of this lesson series is to give a deeper understanding of the potential of machine learning algorithms in inferring things from a user's use of Facebook.

The social implications of data aggregation performed by machine learning algorithms are aptly described in an article by Lori Andrews in the New York Times. She reports the case of an Atlanta man returning from his honeymoon who found his credit limit lowered from 10,800 to 3,800$. A letter told him: "Other customers who have used their card at establishments where you recently shopped have a poor repayment history with American Express." The establishment he recently shopped in was a guitar shop. A term that was popular in the 1970s, "Redlining", is thus changing to "Weblining", in the author's words: "The term Weblining describes the practice of denying people opportunities based on their digital selves. You might be refused health insurance

1 Data Use Policy, https://www.Facebook.com/about/privacy/ retrieved 06/30/2013
2 To the issue Pseudoanonymizing personal data see Berendt et al. (2015)

based on a Google search you did about a medical condition. You might be shown a credit card with a lower credit limit, not because of your credit history, but because of your race, sex or ZIP code or the types of Web sites you visit."[3]

It is already here understandable that decisions entirely based on correlation are, from a user's point of view, totally intransparent and beyond comprehension.

At this stage of the series it is necessary to give further insight and hands on experience in the working methods of machine learning algorithms. We chose the Apriori-Algorithm for frequent itemset mining, as database we took a small sample of Facebook's Graph API.

User ID	Favorite athletes (field)	Education	Relationship Status
1	Basket Ball	Gutenberg-Gymnasium	+
2	Soccer Player	Helmut Schmidt-Gymnasium	-
3	Soccer Player	Helmut Schmidt-Gymnasium	-
4	Soccer Player	Gutenberg-Gymnasium	+
5	Soccer Player	Helmut Schmidt-Gymnasium	-

The resulting association rules (candidate frequent 2&3-itemsets):

Soccer -> HSG: support = 60%, confidence = 75%
Gutenberg-Gymn. -> +: support = 40%, confidence = 100%
Soccer & HSG -> -: support = 60%, confidence = 100%

In 2013, the Psychometric Centre of the University of Cambridge announced that "easily accessible digital records of behavior, Facebook Likes, can be used to automatically and 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."[4] Matter of fact, the Psychometric Centre tried to establish a tool, "designed for use by brands and agencies", PreferenceTool, that uses Facebook Likes to determine the big 5 personal traits (OCEAN) and intelligence. The meaning of the several attributes is described like this: "IQ […] is used extensively both in educational setting, to distinguish individuals with a high ability from those who will need more help, and work setting, to hire and promote employees. It is the best predictor of job success."[5]

There are two Facebook Likes which are extremely popular among students: Converse and Dunkin Donuts. The results for IQ and Conscientiousness, the most important attributes in education and work settings, are disastrous: People who liked these two items are "more intelligent than 25% of the population, putting them in the very low range." Students matching a result like this are obviously not apt for higher education and a promising job career. My student's indignation was accordingly.

The completely intransparent character of purely correlational profiling and its implications for privacy have been made clear so far. We are now dealing with the function of privacy within democracy and the legal consequences.

In 1983 the Constitutional Court of Germany defined Informational Self-determination as a fundamental right in the "Census Judgement". The Court ruled against a census that government considered necessary for labor market-, health-, apartment's policy etc., mainly because of the planned processing of personal information and the collation with population registers. The possibility of Data linking and uncontrolled personality profiling was considered as a threat to privacy and democratic participation. Since then, Informational self-determination is a basic pillar of German and European data protection law which constrains public and private actors. The Ruling was pioneering and contains a fundamental description of privacy's function within democracy: "[The data] can, in addition – in particular, when integrated information systems are constructed – be combined with other data collections into a partial or complete personality model,

3 Andrews (2012), http://www.nytimes.com/2012/02/05/opinion/sunday/Facebook-is-using-you.html
4 Kosinski (2013), p.1.
5 see http://preferencetool.com/ for all attributes.

without the individual being able to control this model‘s correctness or use. [...]. The right to informational self-determination would be impossible to exercise in a societal order and underlying legal order in which citizens can no longer know who knows what, when and in which context about them. Being uncertain about whether deviant behaviour is continuously registered and persistently stored, used and transferred as information, people will try not to draw attention to themselves through such behaviours.”[6]

Since it is has been made clear so far that methods predicting attributes like IQ by evaluating Facebook Likes are incompatible to the right of informational self-determination the question is how a careful study of Facebook’s Data Use Policy is at all sufficient for giving a so called “informed consent” and whether giving an informed consent can be sufficient for the abandonment of fundamental rights. In german context the Lüth-Judgement is relevant here: it defines the fundamental rights as not only negative rights (“defence“) of the citizen against government, but also as positive rights (“claim“), therefor as indispensable and undetachable. They have a radiation effect, an indirect third effect (“mittelbare Drittwirkung”) that radiates in all areas of law. Thus, the freedom of contract in private law is affected, this radiation binds private actors. There have been objections to this point of view.

We discuss these objections at the end of a series in a role play: a Facebook user complains about FB‘s profiling of his user account. Both freedom of contract and right to data protection are fundamental rights.

1. The position of the Constitutional Court is: With privacy-invading private contracts correlates a dispension of fundamental rights, thus forcing the citizen to resign from his right of e.g. freedom of speech - the so called chilling effects - which is unconstitutional.
2. The position of Jurists like Böckenförde is: the self-responsible citizen is incapacitated by an interpretation of fundamental rights that transform the Constitutional Court of Germany into some kind of Constitution-Areopagus that dictates society its values in form of a value-oriented constitution.

The discussion touches an important debate in jurisprudence: the problem of freedom among unequal citizens which has been described (by Garaudy) as “the freedom of a free fox in a free henhouse”.

CONCLUSION

As a result of this lesson series the students have further to discuss the value of privacy in a society that increasingly ignores these rights.

A desideratum in this lesson series is the implications of algorithmic decision making with regard to: a) biased data, echo chambers, filter bubbles, fake news and “digital mass persuasion” and b) Big Data for monitoring educational systems with consideration of students' privacy, educational equity and efficiency, student tracking, assessment and skills.

REFERENCES

Berendt, B., Littlejohn, A., Kern, P., et al. (2017). Big Data for monitoring educational systems. Luxembourg: Publications Office of the European Union.

Berendt, B., Dettmar G. et al. (2016). Datenschutz im 21. Jahrhundert - Ist Schutz der Privatsphäre (noch) möglich? In J. Gallenbacher (Ed.), Informatik allgemeinbildend begreifen. INFOS 2015; 16. GI-Fachtagung Informatik und Schule. Darmstadt, Germany, 20-23 September, 2015 (pp. 33-42). Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn. http://www.infos15.de/GI_Proceedings_Band-249_incl.pdf.

Berendt, B. & Coudert, F. (2015). *Privatsphäre und Datenschutz lehren - Ein interdisziplinärer Ansatz. Konzept, Umsetzung, Schlussfolgerungen und Perspektiven. [Teaching privacy and data protection - an interdisciplinary approach. Concept, implementation, conclusions and perspectives.]* In Neues Handbuch Hochschullehre. [New Handbook of Teaching in Higher Education] (EG 71, 2015, E1.9) (pp. 7-40).

Berendt, B., Dettmar, G., Demir, C., Peetz, T. (2014). Kostenlos ist nicht kostenfrei oder: If you're not paying for it, you are the product. (LOG IN, 178/179), (pp. 41-56).

6 See full text german version here: https://openjur.de/u/268440.html

http://people.cs.kuleuven.be/%7Ebettina.berendt/Papers/berendt_dettmar_demir_peetz_2014.pdf

Kosinski, M., Stillwell, D. and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior, Proceedings of the National Academy of Sciences (PNAS), http://www.pnas.org/content/110/15/5802.full

Andrews, L. (2012): Facebook is using you, in: New York Times, Feb. 4, 2012, http://www.nytimes.com/2012/02/05/opinion/sunday/Facebook-is-using-you.html.

Dreier, H. (1993). Dimensionen der Grundrechte. Von der Wertordnungsjudikatur zu den objektiv-rechtlichen Grundrechtsgehalten (Schriftenreihe der Juristischen Studiengesellschaft-Heft 23), (pp. 9-64).