Data Selfie

_ Design Process and Evaluation

Hang Do Thi Duc _ 22-8miles.com
MFA Design + Technology _ Parsons School of Design
December _ 2015

Thesis Concept

Data Selfie is a tool for Facebook users to personally track their activities on the platform and to see what algorithms might predict from that data. With this chrome browser extension, free for anybody to use, I intend to create an experience that evokes thoughts about the role and increasing importance of big data, machine learning and artificial intelligence in our everyday life. Larger questions can be asked regarding filter bubbles and consumption feedback loops that only enforces your own world views, as well as the concept of implicit bias as an emergent property of these rule sets. My central questions and also the innovative element of my project is: How would one might feel when seeing what *can* and possibly *already is being* tracked of oneself and more importantly, what is the reaction to this data analysis? It's not a secret that social media platforms, Facebook being the primary example, track us and sell that records to (for now, only) advertisers, but we never get to actually see our machine evaluated profile. Thus, the perspective that I will provide users is unique.

Research

I am also very curious about *my* reaction to my own data and its analysis. To implement my concept I am studying the relationship of online self presentation, data mining, predictive analytics, data/online identity and self tracking. I want to find out how such a data identity looks like. What can my consumption data say about me? Or what can Facebook know about me that I don't? Facebook is just

one example that I want to use to demonstrate, that we create different external appearances on different platforms. We can control these – publish, edit and delete. However there are appearances of us that only Facebook's computers see, because they create these new profiles or identities based on what we, the users, feed them just by interacting with content.

My research has been guided by these main questions: What can *I*, as a fresh, but passionate web developer track of myself? And what algorithms might I use to simulate the scenario happening behind the scenes at Facebook? Or how might I interpret this data or find correlations based on new rules that I determine? So far I have concentrated on the first and second question, the last one will be important as I explore the mentioned algorithmic rule sets. An inspiration to this projects was Nicholas Felton's *Annual Reports* that featured his personal offline and later online data, tracked by himself from 2005 to 2014. As an information designer, Felton visualized this data not necessarily to raise awareness about surveillance. Still, in light of Snowden's revelations about government surveillance in 2013, one can't help but situate his reports in this domain.

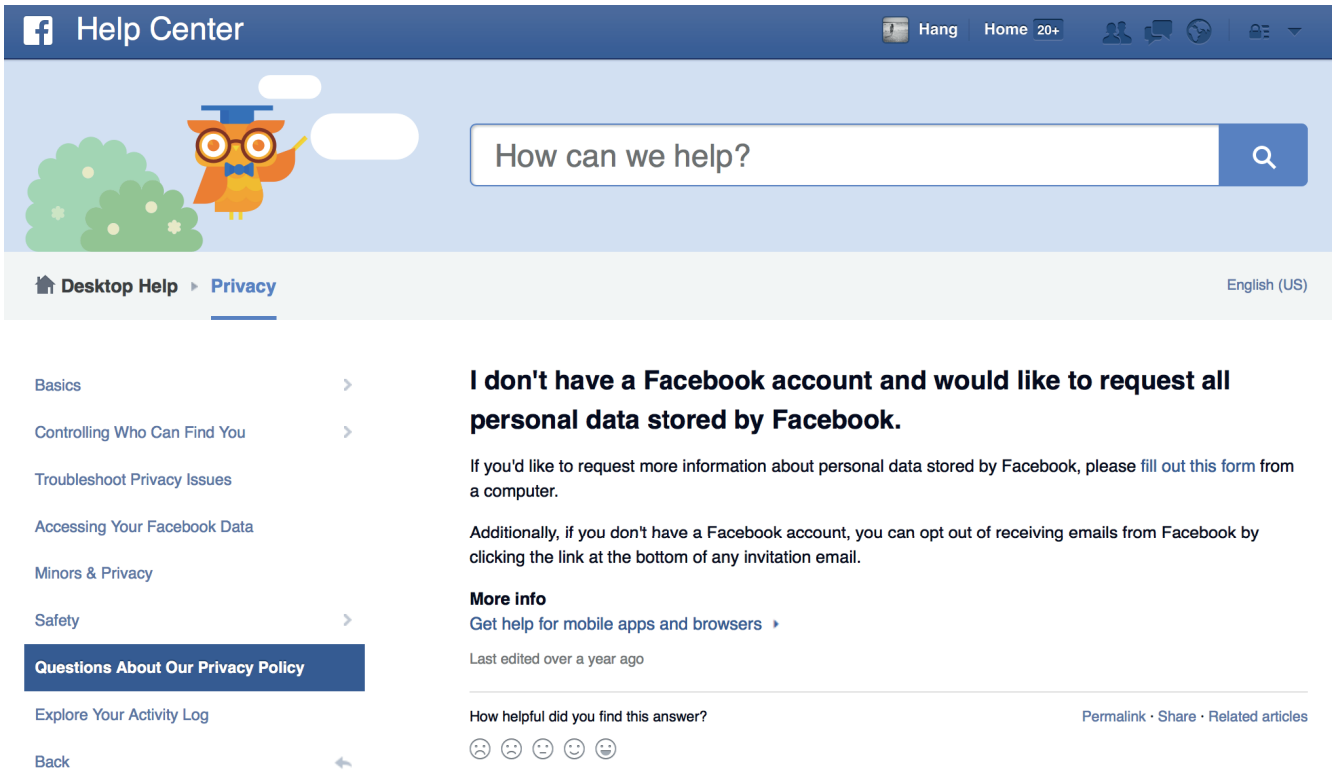
I also relate two art pieces by Evan Roth to my project. The first, *Since You Were Born* is a book that contains the website history over three months beginning with the birth of his daughter. In my view this addresses self tracking and also self reflection on that accumulated data. It is implied that this history, the specific content tell us something about the person and their current life. The second, *Internet Cache Self Portrait series* is a collage of images from the artist's internet browsing cache. "Faces of friends from social media exist side by side with corporate logos, mangled pieces of google maps, family photos and banner advertisements. [...] The prints, composed of memories that were never intended to be saved, are archiving the seemingly incidental into something permanent."[1] I think he mentions something interesting that I relate to the public debate about the right to be forgotten online.

Of course that is not in Facebook's interest. Their business model is based on precise personal data. We can never certainly say what is tracked of us and what is done with that data. Below is a screenshot

---

[1] Roth, Evan. (http://www.evan-roth.com/work/internet-cache-self-portrait/)

that shows that even if you don't have an account (anymore?), Facebook *knows you*.

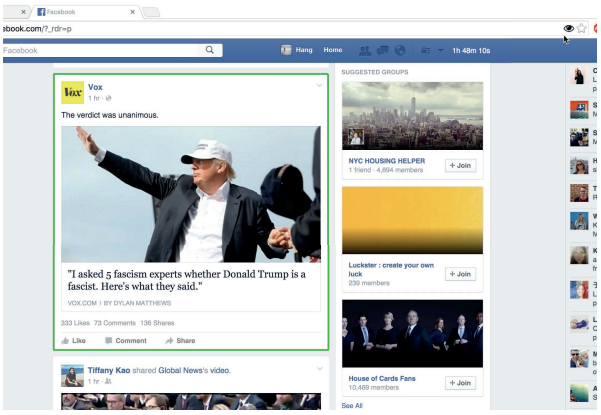


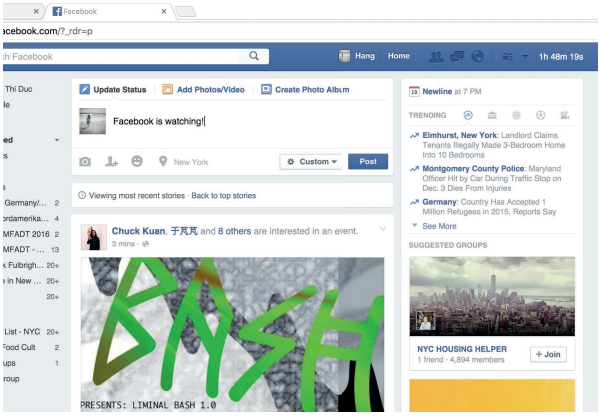
Exploration Through Making

As I have arrived at this concept fairly late in the semester I have not explored as much compared to other students. So far, I have explored my domains and questions through one proof of concept prototype that I have iterated once. First, I started to think of the possible form for this personal data tracking tool. I decided to create a chrome extension. Chrome is the most used browser and extensions have been popular as critical design pieces.[2] Coding in this format was the first challenge that I had to overcome, since I had no experience with this beforehand. I also wanted to make sure that the data that is collected of oneself won't exist in any cloud data base. For this I experimented with the Chrome local storage which has a capacity of 5MB.

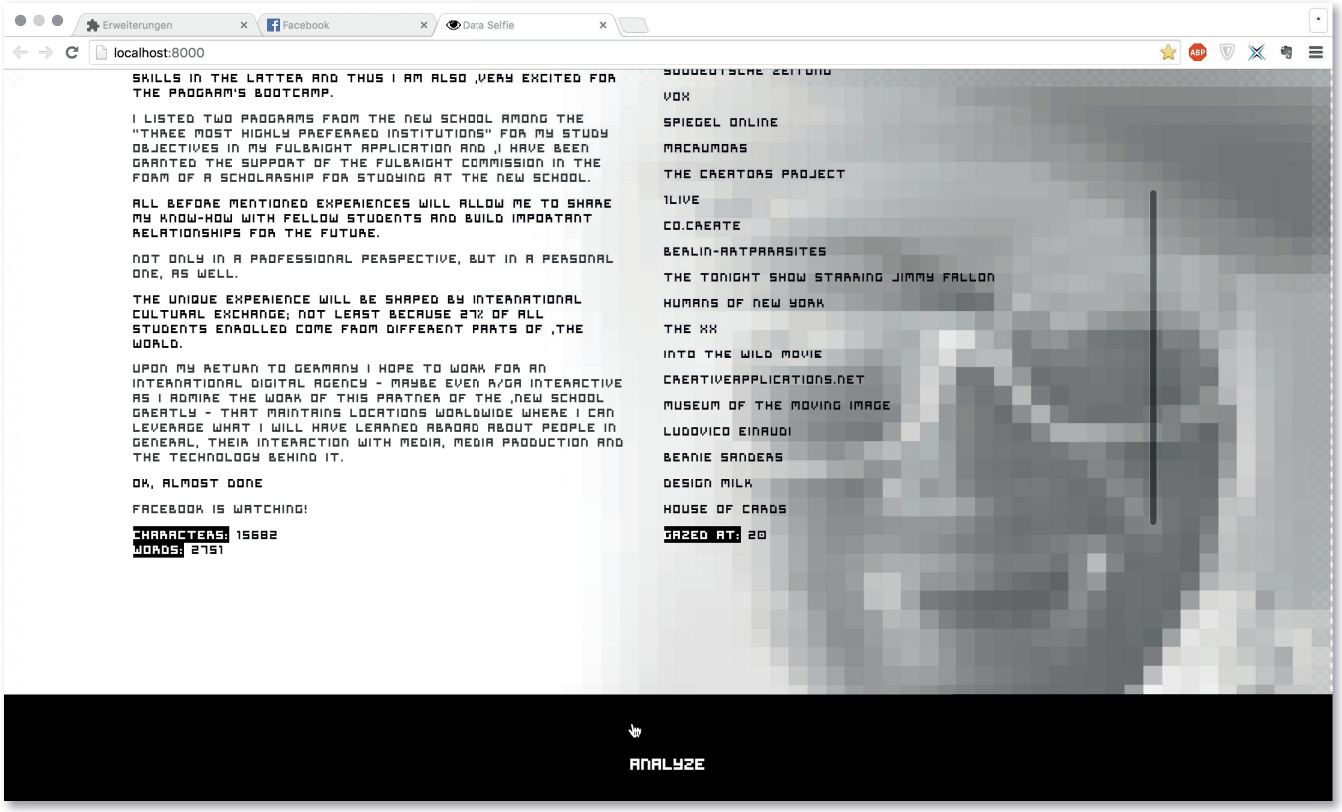
[2] For example there is an extension called *Rehumanize* that replaces the word "refugee" with "human" in your browser. Or *The Ethical Ad Blocker*, that, when it detects advertising on a website, blocks the entire website. The user doesn't see ads, but they also don't get to use free content.

What I actually track now is what one types on Facebook, what posts stay longer in the viewport (highlighted with a green border) and the time one spends on facebook.com. In my experiment the data that was collected while I developed this took up about 1% of the Chrome local storage.

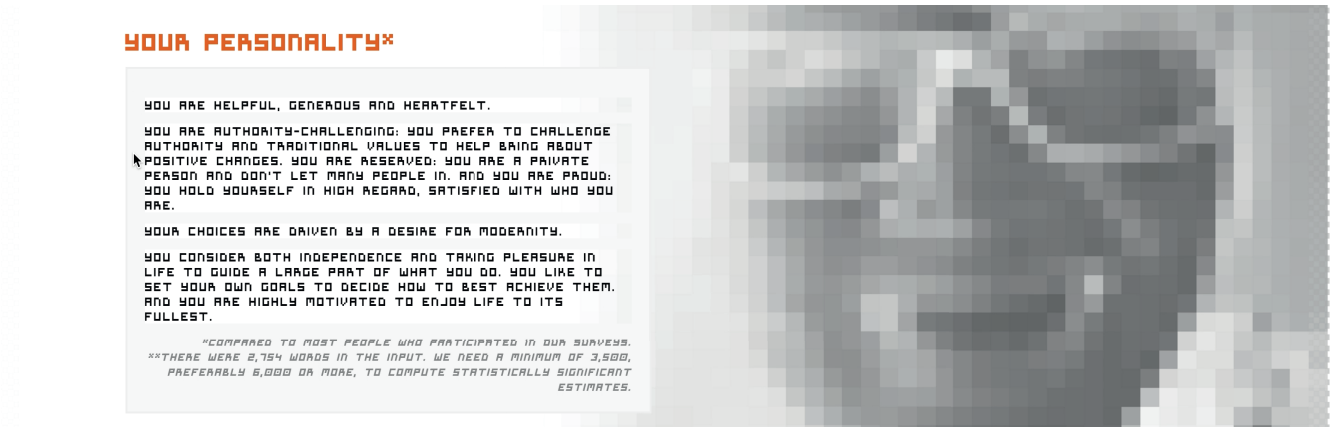




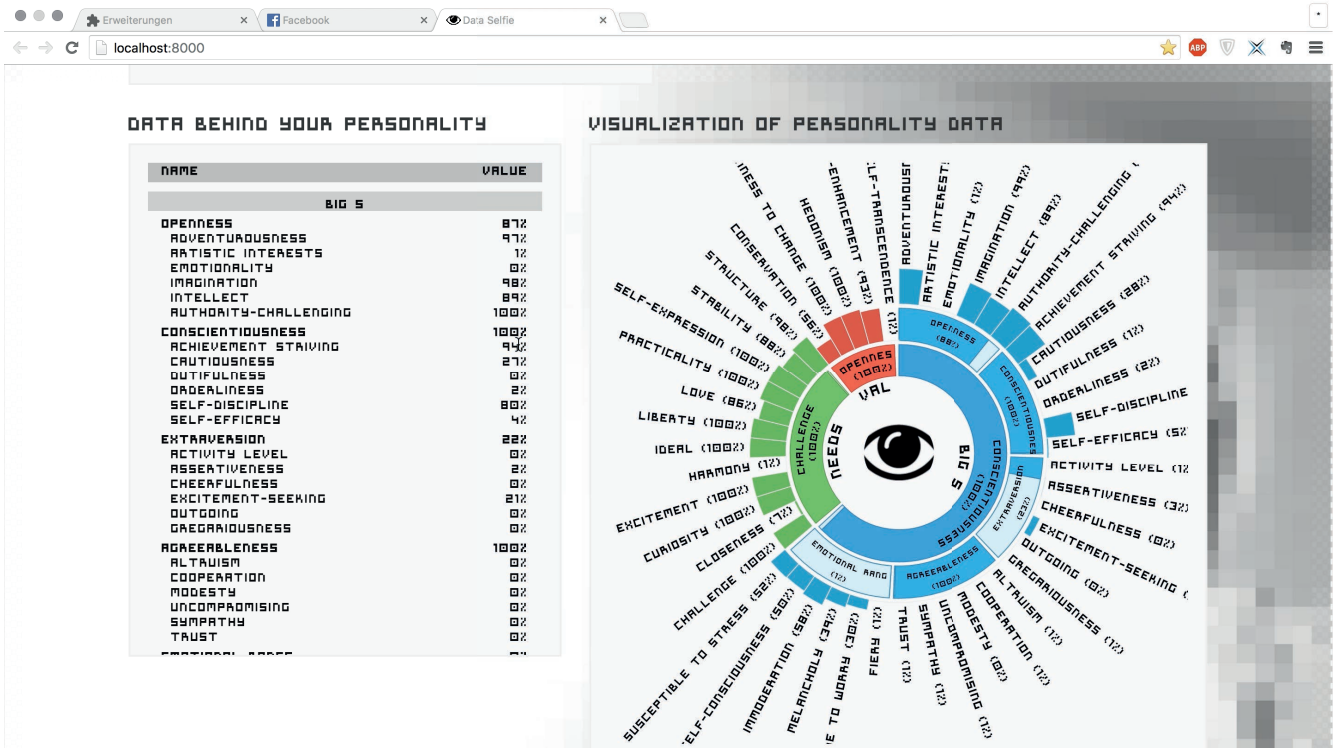
The user then sees all typed words and the top 20 authors of the posts he or she stared at, in a computer terminal like log and can then decide to analyze it. The aesthetic, especially the font, of this should mimic a “hacker view” experience, which is a way of enforcing the idea that this is something one usually doesn’t have access to.



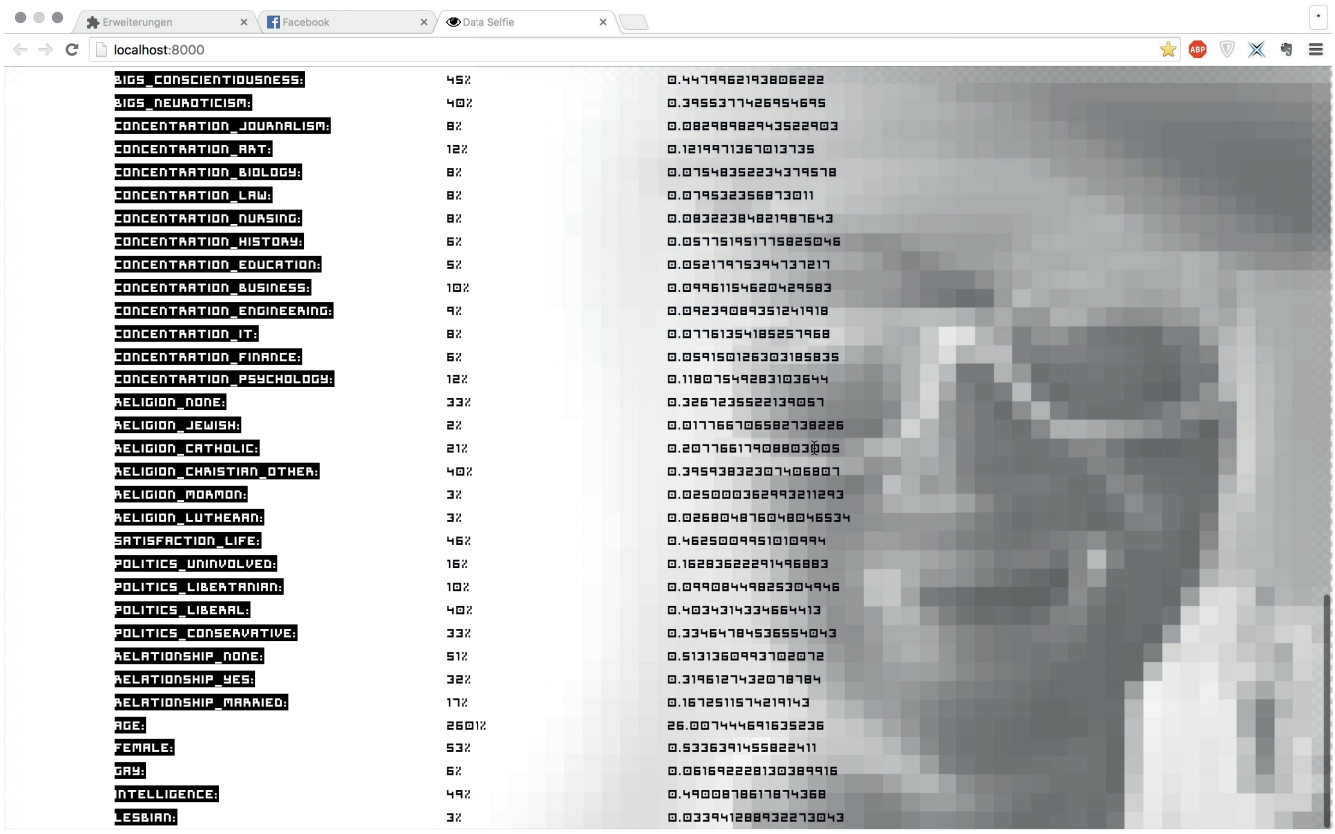
In the initial prototype I analyzed the words I typed with IBM Watson's Personality insights API.



Here is a visualization of the analysis and the prediction regarding my values, needs and the big 5 openness, conscientiousness, extraversion, agreeableness and emotional range.



I iterated my prototype to incorporate what can be predicted based on the top Facebook pages one interacts with, in this case it means the authors of the posts one looks at for a longer period. For this I used Apply Magic Sauce, Prediction API by the University of Cambridge. The analysis through this API include the big 5, concentration, religion, politics, relationship, age, gender, intelligence and sexuality.



This prototype proves the tracking side and analytics side of my concept. Apart from the curiosity from classmates and friends for potentially using this, I have not formally tested this current prototype. The main reason is that I only have limited access to the API and I would need to pay for more usage. At this stage I would not gain a lot of insights from user testing. This makes more sense when I will have implemented more methods of tracking various activities. I am confident I can manage the technology, since the progress from this prototype was achieved in less than 2 weeks and I have 14 more weeks to build a stable version to track an array of a person's Facebook and online activities – possibly mouse positions, eye tracking and cookies – and analyze that through an algorithm. This algorithm can be a mix of IBM Watson API, the Apply Magic Sauce Prediction API and my own set of rules that might involve machine learning. More explorations through making will determine the direction.

Technical Production Plan

I am confident that I will have access to these mentioned parts. The Watson API will probably be available through either IBM employees that Sven Travis is putting me in contact with or through The New School as David Carroll already talked to Senior Vice President and Chief Information Officer Anand Padmanabhan about the university becoming an IBM partner school. Since Apply Magic Sauce is an academic project I hope to collaborate or at least receive support from their team. This is feasible, because they only provide example code for Java developers, but promptly helped me when I had questions about accessing their API through server side JavaScript. Tracking will be the easier part; I will track what *I can* track, assuming that whatever that might be is probably already being tracked. I have been interested in and explored eye tracking since the first semester and as technology advances quickly new methods of eye tracking emerge. Now, there is a way to track eyes just with the webcam, this will be something I will look into as research through experiments.

January I	Reach out to experts. Add tracking methods. Research more about the APIs and how they work in order to possibly manipulate or process their results.
January II	User tests for understanding of concept for a period of 1 week. Manipulate or process API results, i.e. apply own rule set.
February I	Continue: Manipulate or process results (in my own way).
February II	User tests for usability.
March I	Use it myself for 2 weeks. Find bugs optimize structure and performance.
March II	Marketing (web site, teaser film) to situate project and create tone for project's message.
April I	Release extension as tool and experience to public.

Design Statement

I am creating a browser extension for anybody with a Facebook account to track personal activities on the platform, to experience what might be predicted by algorithms based on this collected data and to reflect on their consumption and personality. Simultaneously, this tool should evoke thoughts about the role and increasing importance that is given to artificial intelligence by big internet companies in our everyday life.

Design Questions

What is the reaction to the result (analysis) of one's online activities? How might I create a shocking reaction or one that evokes thinking about the current state (social media, privacy, algorithms)?

Can I, and if yes how, provide a different perspective on online tracking and artificial intelligence and their implications?

How objective and accurate is this kind of analysis or prediction? Does it even matter if big data is seen as truth?

Evaluation

Generally speaking, in my writing and my making, I am trying to answer the questions of why my project is important, unique and a contribution to the debate of data mining, predictive analytics and privacy issues in social media or even the internet as a whole. In my prototypes I am attempting to implement interesting tracking methods that reveal users' activities which they might not always and consciously think about (time spent, time looking at posts, typing). Then there is the question of how might I present this collected data that conveys the notion of a "secret" or "hacker" view, a view that is never accessible to you, although *you* are the subject and provider of that information.

Providing this unique perspective on your data and also analyzing it the way a machine in Silicon Valley might process it is the strength of my project. I feel also very strongly about making this experience only available on your machine, the data will only reside in the local storage of the user's

browser and will not be stored in any database in the cloud. I am not only addressing the fact that data mining, surveillance and big data predictive analytics exist and are almost omnipresent and potentially dangerous, discriminating or “bad”, but I am also demonstrating how the result of all these practices might look like and in turn, how this might be used and can affect you as a person online and offline.

One weakness of my project so far is that I haven’t conducted very extensive research about what experts think is going on behind the scenes. I know for a fact that Facebook is taking the duration of a user looking at a story in their newsfeed into account, because they publicly announced that. The social network also published news that they track your cookies, however I suspect there are more interesting data points that are of interest to the company. I also need to closer investigate artificial intelligence APIs and in general structures and architecture of algorithms, so I can better understand how I might provide results of the data that is processed by me, another person with biases just like the architects of other algorithms. Furthermore it will be important for me to grasp what the users’ reaction generally is, but also specifically if they trust the analysis they will see.

I am at an interesting stage in my process; my proof of concept prototype is a good point to start asking a lot of (new) questions. What if, based on the analysis, I would manipulate the user’s Facebook feed? What if I can give a “report” of why they see the ads they are seeing? There a lot of different ways this could go, possibly after thesis. I could also give scenarios, for example “Based on your interactions on Facebook, you would get this loan or this insurance”. Inspired by Evan Roth’s work I am also thinking about creating an art installation of the data I will be collecting of myself and its analysis.

Production will not be entirely separated from research. I will investigate what can be tracked and what is tracked of us online as I am developing and improving my browser extension. These investigations will include reaching out to experts in the field of data science, artificial intelligence and machine learning and they will inform my further prototypes. I see my future experiments also affecting my research. I strongly believe this tool is interesting, not only to me, but also to other people who are starting to ask the same questions. To conclude, I see myself as an artist, someone that is highlighting processes and concepts that exist and are developing in the world we live in now.