

Youchun Zhang

Thesis Studio 1

6 November 2018

Milestone Paper – Prototype

This thesis prototype experiments with the forms to answer the questions from the research phase, including how the algorithmic transparency affects the daily experience of users and employees in an algorithm-controlled system, and what are the current academic and legal reactions responding to this phenomenon. It chooses YouTube, one of the biggest video-sharing websites in the world as the primary research object and focuses on unboxing one specific algorithm-controlled system: the recommendation system.

Impact and Coverage of YouTube

YouTube is the biggest online video platform featuring a wide variety of user-generated contents. It was purchased by Google in 2006. Originally, YouTube was designed for the users to post original content that was either funny or worth being shared for many other reasons. In 2016, the number of YouTube users had reached 1.35 billion. According to the forecast, this number will climb to 1.86 billion in 2021.¹

In the past ten years, the categories of content being generated by its users have been extended to video blogs, gaming videos, instructional videos, and educational videos. As of July 2015, more than 400 hours of video were uploaded to YouTube every minute.² As of September 2018, the number of subscribers of the most popular YouTube channels has reached over 65 million. 81 percent of global internet users have visited YouTube in the recent month, and 31 percent of YouTube users access it more than once per day.³

¹ "YouTube: Statistics & Data". Accessed November 10, 2018.
<https://www-statista-com.libproxy.newschool.edu/topics/2019/youtube/>.

² "Hours of Video Uploaded to YouTube Every Minute as of July 2015". Accessed November 10, 2018.
<https://www-statista-com.libproxy.newschool.edu/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/>.

³ "Number of YouTube Users Worldwide from 2016 to 2021 (in Billions)". Accessed November 10, 2018.
<https://www-statista-com.libproxy.newschool.edu/statistics/805656/number-youtube-viewers-worldwide/>.

YouTube uses Google AdSense, a program closely working with Ad-tech and targeting on making the publisher make money from their online content. In 2017, Google's share of total U.S. digital advertising revenues amounted to 38.6 percent, as compared with 19.9 percent share from Facebook. Some of the top paying AdSense niches are Finance, Internet Marketing, Technology, Web Hosting, Internet & Computers.

Prototypes: Experiments

Three prototypes are created in the experiments. The first prototype creates a workflow that encourages the users to actively request for the disclosure of the mechanics within an algorithm-controlled system. The physical part is consist of a soft-circuit button, a mini-size thermal printer, the Arduino Uno, and several wires. The Arduino Uno has been pre-programmed through its IDE specifying a message sent from "Algorithm Services Department". Once the audience presses the button, a piece of paper will be printed and come out from the thermal printer describing the process and the goal of the YouTube recommendation algorithm. The audience can keep the paper as a personal copy of the record.⁴

The message on the paper is based on a reverse engineering article for the YouTube video publishers. Some of the conclusions drew by the author of this article are, YouTube uses user's Watch History, Search History, and Demographic Information, and the goal of the YouTube's promotional algorithm is to promote videos to lead to longer viewer sessions.



Figure 1. Photo of the first prototype

⁴ Gielen, Matt. "Reverse Engineering The YouTube Algorithm." Tubefilter. February 16, 2017. Accessed November 10, 2018. <https://www.tubefilter.com/2017/02/16/youtube-algorithm-reverse-engineering-part-ii/>.

Based on the feedback collected from the first prototype, the message evolves into a form of receipt. The receipt serves as a proof of the list of services, the personal data and cookies being collected from users, and the list of technologies being used in the YouTube system. The mechanics of requesting for this receipt is the same as in the first one. Moreover, the barcode on the top will take the user to the Google accounts page, as a reference to the barcode on the regular receipt that can help the salesperson quickly scan and retrieve the detail of the original transaction. The design scenario of acquiring the receipt is intended to be when the user first signs up with Google account and when the user accesses the service.

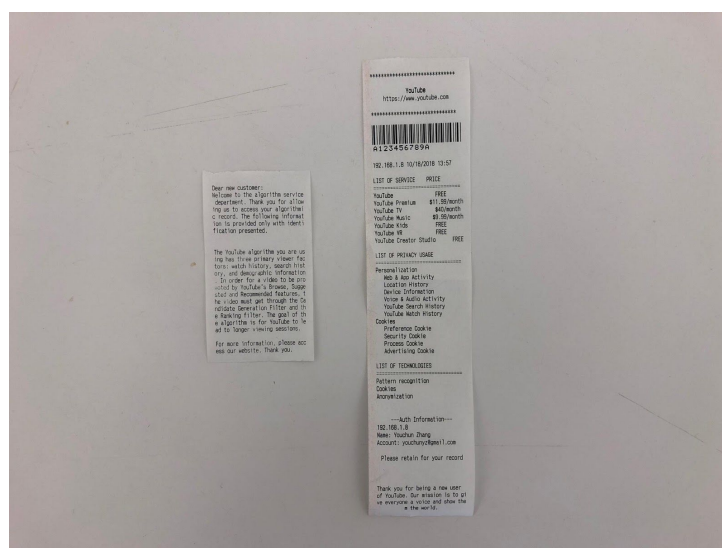


Figure 2. Photo of the printed receipts in the first and the second prototype

In the third prototype, more formats are presented, including one trying to explain the detail of search history, one elaborating the Ad profiling of a real Google account and Google's estimation of the owner's interests, and another one with steps to develop a simple recommender system using IMDB Top 250 movies dataset.⁵ The detail of search history is a list of search queries embedded in the URL as shown in the Google "Download your data" package. Some are written in Chinese, therefore, they are not all readable.

⁵ "Recommender Systems in Python Tutorial." DataCamp Community. Accessed November 10, 2018. <https://www.datacamp.com/community/tutorials/recommender-systems-python>.

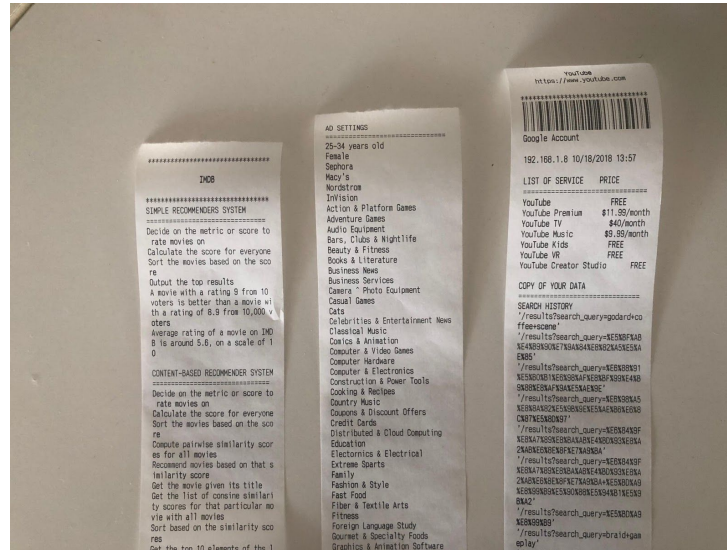


Figure 3. Photo of the printed receipts in the third prototype

In addition, two possible forms of physical setup are discussed at this stage. The experience of this receipt machine can be either private and personal or social and public. Imagine that the machine is a small portable device that the audience can purchase and bring it everywhere. Whenever the audience needs the information, he/she can connect it with an IoT device like a laptop. While in the other scenario, the audience may be watching the machine work in a group. A story is told by combining the information printed by the multiple machines.

Prototypes: Reflection

A 10-minute user interview and testing session is frequently used to collect the feedback for each of the prototypes. The session is started with asking the user their regular usage of YouTube to get a basic idea of how familiar they are with the system and the features mentioned in the prototype. Then the user will need to answer questions like which scenario he/she thinks is related to the most, and what information he/she consider is more important.

The layer of information is a key concern of the prototypes. There are three processes involved in these prototypes - the process of the audience using the YouTube as a context, the process of the device printing the algorithmic information, and the process of the audience understanding the information. The information can be overwhelming if not designed properly.

Particularly, for the audience with limited knowledge of keywords in the computer science field, how to communicate the information but not to lose its accuracy needs to be further investigated.

Receipt as a physical proof of transactions is proved to be effective. Using an analog form with a digital content makes the unboxing process more tangible. The medium quickly sets up the connection between the service and the acknowledge of payment. It also brings up the idea of the warranty, returns, exchange, special offers, and advertisements. It provides an explicit way for the users to indexing information from a bigger repository. Some audience expresses that the machine provides a funny tone.

Users in general find that the goal of the YouTube recommendation algorithm is to engage the viewer with a longer viewer session is surprising. However, when asking about whether people know about the information is accessible on Google's website, most users agree on the fact that they do not care enough. One of the feedback is about the feeling of failure as a result of exchanging privacy for convenience.

During the user test, an interview with one of the users – a part-time fashion blogger also reveals some interesting findings. She usually spends about 40 hours on producing the content. The length of the final version is about 8 minutes. However, when she starts, it is a 15-minute one, but everyone suggests her to cut it shorter so that people will watch till the end. The reason that YouTube becomes the biggest content-sharing platform is related to that YouTube has relatively fair and transparent filtering. This means YouTube's algorithm is more technical and less about human's interference, unlike Instagram and Weibo in China. Because Weibo cannot publish all contents from all the creators one follows, creators need to pay to get their post through the blocking algorithm and recommendations.

Prototypes: Design Value

The goal of designing these prototypes is to enable the process of seeing, checking and monitoring of the logic hidden behind the products we frequently access in everyday life. It tries to use the physical space and the tangible experience to visualize the dynamics of systems behind the screen. It is not a solution-based product; instead it may be only an opening of a series of products signifying the importance of making these information more available and

understandable. When collecting the feedback, it tends to surprise people by the unknowing fact and model people's responses and how they act differently. The impact can be evaluated quantitatively and qualitatively. A feedback collecting system can be designed to follow up with the users about to what degree their interactions with the system can be potentially modified. The dynamics during the exhibition can also be recorded as evidence.