Sinders, Caroline; Ahmad, Sana

Article — Accepted Manuscript (Postprint)

The labor behind the tools: using design thinking methods to examine content moderation software

Interactions

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Sinders, Caroline; Ahmad, Sana (2021) : The labor behind the tools: using design thinking methods to examine content moderation software, Interactions, ISSN 1558-3449, Association for Computing Machinery (ACM), New York, NY, Vol. 28, Iss. 4, pp. 6-8, https://doi.org/10.1145/3470492

This Version is available at:
https://hdl.handle.net/10419/266363

The Labor Behind the Tools: using design thinking methods to examine content moderation software

Caroline Sinders, Convocation Design + Research,

Sana Ahmad, Freie Universitat Berlin and Weizenbaum Institute

Content moderation is widely known to be hidden from the public view, often leaving the discourse bereft of operational knowledge on social media platforms. Media and scholarly articles have shed light on the asymmetrical processes of creating content policies for social media and their resulting impact on the rights of communities who are marginalized. Some of these investigations, including documentaries, journalistic investigations, academic research and from individual whistleblowers who work at these companies, have also uncovered the practices of moderating user generated content at social media platforms like YouTube, Facebook, TikTok and others. In doing so, the complex and hidden outsourcing relationships between social media companies and third-party companies, who are often located away in different geographical locations, have been made visible. Most importantly, the identification of these companies and outsourcing practices have brought to public attention the secretive work processes and working conditions of content moderators.

This article aims to illustrate a unique method undertaken to examine the software used in the content moderation labor process. Across December 2020, we held two focus group research

workshops with ten content moderators in total. Our participants included both former and current employees at a third-party IT-enabled services company in Berlin which supplies content moderation service to a social media monopoly based in the US. The participants of our study have a background in immigration, with most of them living in Germany for less than five years. Fewer employment opportunities due to lack of German language skills, motivated all of our participants to apply for this work process. With the focus of this article lying on the methodological part of our research workshops, we will not be able to elaborate further on the recruitment and work process of the participants.

The content moderation labor process is highly confidential, and workers face intimidation from their employers in terms of sharing or disclosing the kinds of work they do, and how they do it. When crafting these workshops, we were inspired by collective memory practices. The workshops allowed us to use design thinking exercises, somewhat reverse engineered, to uncover and gauge the infrastructural design, user interface and user experience design of the systems they work in.While design thinking has a broad meaning and definition, it can be used to problem solve, and create new software or designs, we reverse engineered the process and used exercises designed to "help frame questions" and translation ideas tangible interfaces and architectural layouts. Caroline has observed from her time in industry UX design that product designers often lean towards product design and building a specific or concrete 'thing' be it a new product, a product augmentation or process when utilizing design thinking exercises. Our goal in this workshop was to ground the participants in space of making; thus, by asking people about their day, the hardships they face, and the literal structure of their workday, their work, and how iteratively they approach work, it starts to build a foundation to think through building something to hold that work. These

exercises helped ground the moderators in the logic of how their software functions, and spark memories of the software they use. By working with multiple content moderators in a workshop setting, the workshop could create a space of organic reflection and comparison on the tools and protocols amongst the content moderators, leading to discussions on how the Berlin-based employer and the social media client managed the work. In holding our workshops, we followed necessary research protocols, informed by the research ethics standards at the WZB Berlin Social Science Center.

The workshop was divided into five exercises, with each exercise sequentially iterating off of the previous exercise, starting with writing out a workflow of their day such as what is the first thing they see when they login, what do they do after login, where are the tasks stored, how do they engage in the tasks, amongst others. Activities belonging to the first, second and third exercises were aimed at deciphering the layouts, workflows, and design. Activities four and five had the group discussion format which invited participants to share their experiences on workplace surveillance and propose work-related changes and potential improvements.

Designing the Workshops

Ensuring the privacy of content moderators was integral to this project. Considering the secretive work practices, we were aware of the potential threats to moderators' job security and therefore they were not requested for screenshots of their work. Additionally, the wireframes have been redrawn for this article as an added precautionary measure. The workshops were entirely held using the audio communication function of a web-based and open-source software, with appropriate attention given to anonymizing the participants.

Apart from privacy considerations, there were other challenges to consider. In order to examine the infrastructural design, UX and UI of the software, we needed to guide the participants through the self-drawing process of the software they used for moderating content. Design is a specific medium with a language and vernacular unto itself. Content moderators may not know the names or how to describe the elements in the software they use. Therefore, we designed the flow of activities to accommodate for this constraint and guide the participants through an ideation process which could ensure their participation. Our approach was to create exercises that slowly, organically, and iteratively helped the participants sketch out the software they use.

Workshop Structure

Both the workshops commenced with presentations from the two organizers, with Sana introducing the existing studies on the labor process of content moderation and the importance of undertaking this research project and followed by Caroline who explained basic design elements that we assumed were elements they would see in their own software. The main question guiding the workshops were: how to examine the design of the content moderation software through the experiential inputs of the workers while protecting the workers. With the collective knowledge of our organizing team, including academic research by Sana on the labor processes of content moderation in third-party IT BPO companies in India, along with the design background of Caroline with her expertise in designing digital tools and software, we aimed to answer the research question in a multi-disciplinary manner. The practical knowledge of two student assistants at the WZB Berlin Social Science Center, further benefited our workshops.

Using Iterative Activities

As mentioned earlier, the indispensable aspect of the workshops has been recognized as their iterative structure, i.e., knowledge about the content moderation software was gradually developed, using a step-by-step process wherein participants could build out, think, and iterate on the software design they were recalling. This gave the participants space to reflect and redraw for better accuracy. Each activity started with looking at general questions on the content moderation work process, including the generic layout of the software, the moderators' workflows and which increasingly became more specified in the process of adding complexities in each subsequent workshop activity. Much like generating an outline, we could then go more granular and specific-asking about smaller and specific features, and more qualitative questions on the kinds of content they were moderating, how their specific company organized a working day, and how their specific team used the software. Such a sequential technique was further aimed at allowing the participants to gradually ease into the design thinking process.

The first activity was planned to garner an undetailed view of the daily workflow and routines of the content moderators. Accordingly, the participants were asked to list down the daily sequences of their work. Through this, we could firstly determine the type of software used, i.e., whether it is web-based or not. Other elements which could be assessed from the experiences of the workers included their observations while logging on to start the moderation work, the first task they undertook after being logged in, and the functions available in the software, including those related to internal communication such as the ability to send emails to their management and interpersonal messaging option with their team members or other workers in the company. This overview was

important in getting a preliminary grasp on how workers accessed content moderation tasks and if they could exercise flexibility between carrying out their work-related tasks and other assignments. Such an activity was useful in persuading the participants to revisit their daily work routine and recall the essential elements of their work process. It also gave them a basis for remembering the different kinds of screens and windows in their software, which was subsequently carried forward into the second activity.

Building on the first, the succeeding activity allowed the participants to further iterate and lay out the different screens and states of their work software, pulling directly from the initial list that they had made. The list helped the participants to look back at what each ‘state’ or ‘screen’ held and the kind of actions each screen allowed for.

The third exercise was focused more specifically on the software page layout and gaining insights into the granular elements. During this activity phase, the participants started drawing and building out a rough wireframe, with focus placed on elements such as menus, buttons, locating different information in distinct pages and positions in the software. Along with the illustrations developed by the participants, the experiential narratives of the content moderation labor process, including the management control embedded in the software, were also shared with us. Our questions related to these themes roused the responses by the participants.

Subsequently, the workshops proceeded to the final activities which invited the participants to engage in discussions amongst themselves. The fourth activity picked up on the themes of workplace surveillance and monitoring strategies by the management. This included examining

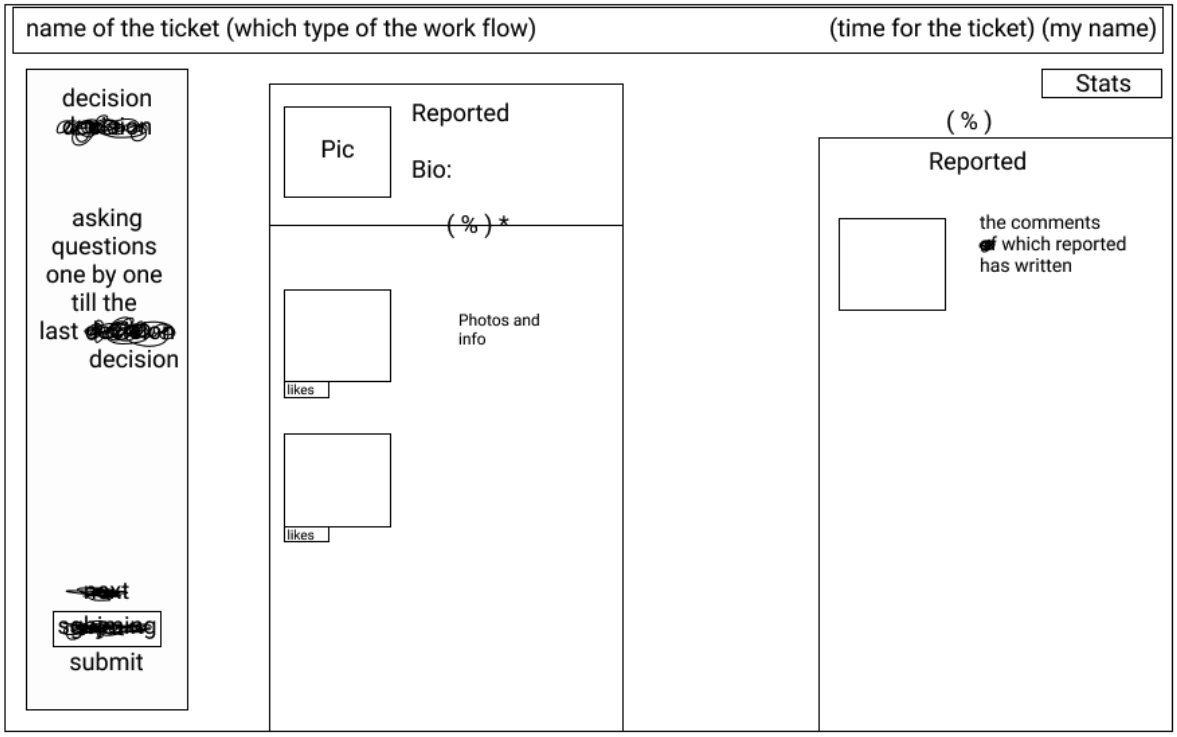
participants' views on the possible embeddedness of surveillance technologies in the content moderation software. The fifth and the final activity was aimed at drawing on the participants' understanding of the content moderation work process and the imaginable ways in which work could be made better for the moderators. In doing so, they were also provoked to think about the probability of machine learning tools in their work software and if it could potentially risk their job security.

Conclusion

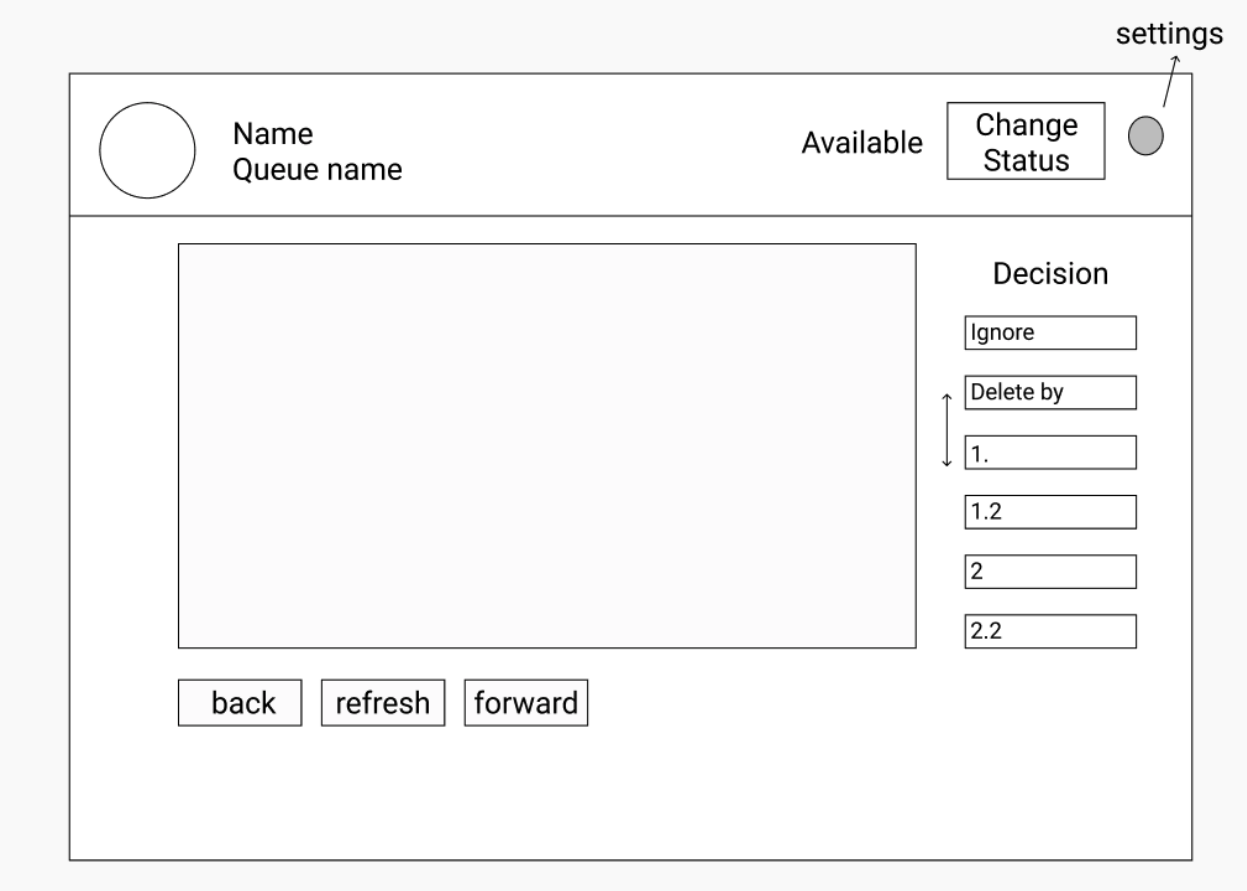
In terms of the research method, we see its merits, especially in being able to draw out human-machine interaction through the collective memories of our participants. Our workshops have yielded design layouts of the software which are unique to the limited information available on the labor processes of content moderation on social media. At the same time, conducting such a workshop or focus group interviews can be more fruitful when combined with one-on-one interviews. Considering the precarious background of our participants and their limited possibilities for exercising collective struggles against the management, we have managed to create a space wherein current and former content moderators have been able to share with us and with each other about their work-related experiences and the management interactions through the focus on the content moderation software. Future research on the use of technical control and novel ways of labor resistance using technology can enrich the existing research on content moderation.

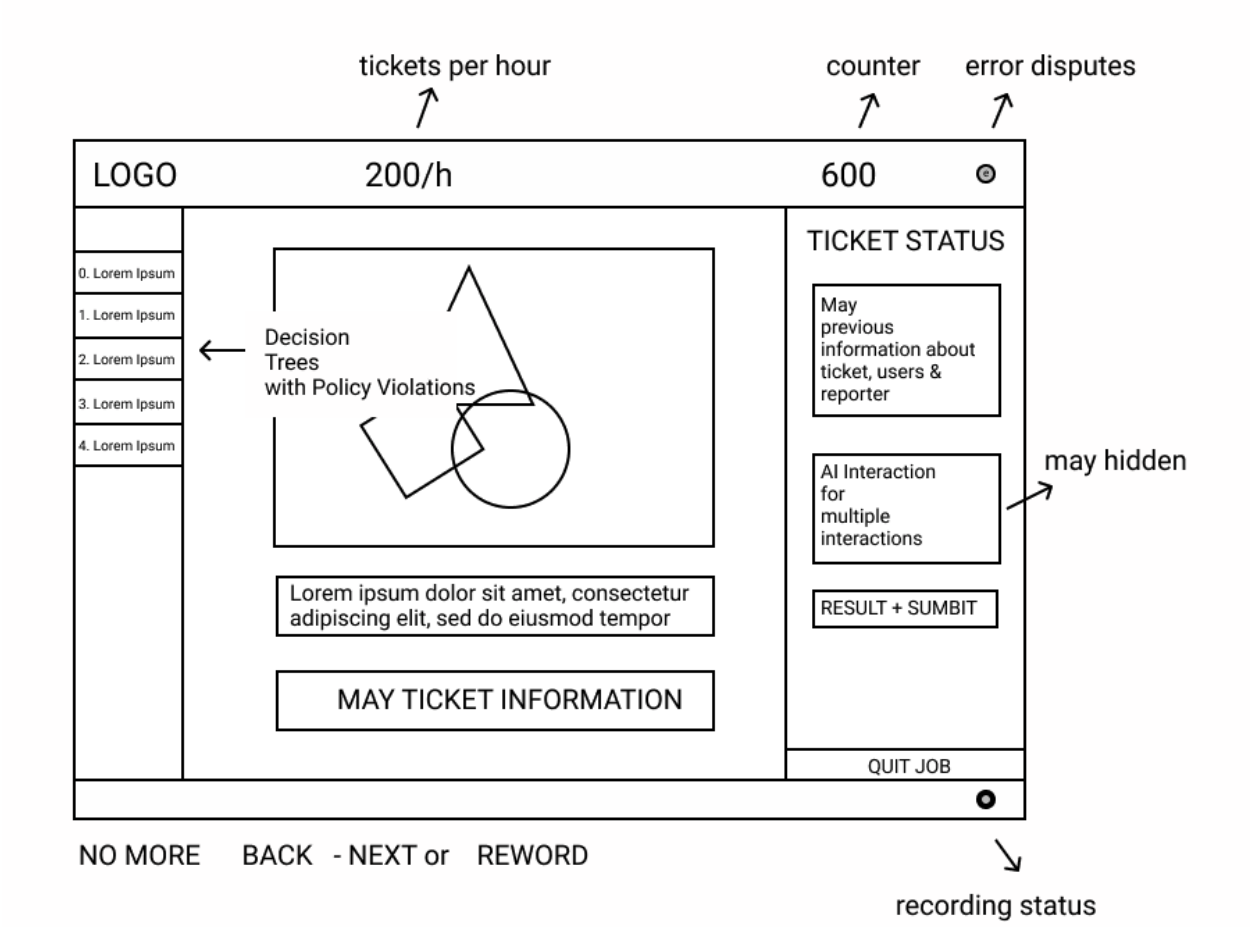
IMAGES HERE

All three figures below are from second workshop's activity 3, detailing the software page layout.



* if the photos or bio was violationg we should choose it and around show a percent or violating content. and the same in the comments that the reported has written. The decision we should choose depend upon the prercent with more than 50%. the page would be disabled.


settings
Name
Queue name
Available
Change Status
Decision
Ignore
Delete by
1.
1.2
2
2.2
back
refresh
forward


tickets per hour
counter
error disputes
LOGO
200/h
600
0. Lorem Ipsum
1. Lorem Ipsum
2. Lorem Ipsum
3. Lorem Ipsum
4. Lorem Ipsum
Decision
Trees
with Policy Violations
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor
MAY TICKET INFORMATION
TICKET STATUS
May previous information about ticket, users & reporter
AI Interaction for multiple interactions
may hidden
RESULT + SUMBIT
QUIT JOB
NO MORE
BACK - NEXT or REWORD
recording status