Through the Looking Screen-Glass

The Invisible Labor of Commercial Content Moderators (CCM)

—

Sienna Parker

Abstract

On the internet, the scale of user-generated content reaches exuberant numbers of photos, videos, and posts. Social media websites must moderate this content to ensure that it befits their community standards. To do so, these companies hire commercial content moderators (CCMS) who are responsible for viewing flagged content. The job of CCMs requires viewing thousands of hours of video and photos that often contain violent and horrific imagery that can lead to trauma. However, CCMs and the traumatic nature of their work are largely invisible to the public eye.

In this proposal, I have designed an experiential, multimedia exhibit for the public to gain a better understanding of the work of CCMs. This proposal includes a storyboard walkthrough and mock-up video of the exhibit in which viewers partake in a “game” simulating being a CCM. While viewers play the game, video cameras film them and manipulate their faces to age as a demonstration of the emotional toll of being a CCM.

Mark Zuckerberg
3 May 2017 ·

Over the last few weeks, we've seen people hurting themselves and others on Facebook -- either live or in video posted later. It's heartbreaking, and I've been reflecting on how we can do better for our community.

If we're going to build a safe community, we need to respond quickly. We're working to make these videos easier to report so we can take the right action sooner -- whether that's responding quickly when someone needs help or taking a post down.

Over the next year, we'll be adding 3,000 people to our community operations team around the world -- on top of the 4,500 we have today -- to review the millions of reports we get every week, and improve the process for doing it quickly.

These reviewers will also help us get better at removing things we don't allow on Facebook like hate speech and child exploitation. And we'll keep working with local community groups and law enforcement who are in the best position to help someone if they need it -- either because they're about to harm themselves, or because they're in danger from someone else.

In addition to investing in more people, we're also building better tools to keep our community safe. We're going to make it simpler to report problems to us, faster for our reviewers to determine which posts violate our standards and easier for them to contact law enforcement if someone needs help. As these become available they should help make our community safer.

This is important. Just last week, we got a report that someone on Live was considering suicide. We immediately reached out to law enforcement, and they were able to prevent him from hurting himself. In other cases, we weren't so fortunate.

No one should be in this situation in the first place, but if they are, then we should build a safe community that gets them the help they need.

Jenny Lee, Jasmine Teng and 142K others 11K comments 9.3K shares

What is the scale of content posted?

350M photos/day

720K hours video/day

500M tweets/day

https://www.internetlivestats.com/

Who are content moderators?

"The workers act as **digital gatekeepers** for a platform, company, brand, or site, deciding what content will make it to the platform and what content will remain there." (Roberts, 2016)

Statistics:

- $11.8B Industry
- 700-2,000 posts viewed per day (Facebook)
- Must sign NDAs
- Mostly contract workers with low wages
- Facebook = 15,000 workers
- Youtube = 10,000 workers
- Twitter = 1,500 workers

Cognizant

amazon mechanical turk

Who gets hired as a commercial content moderator?

Young College Graduates

"During my interview and before the interview there was a written test thing, and t**here they were very clear that like, you are going to see disturbing content, this is not necessarily the easiest job, emotionally speaking**. So I think they made it very clear before I started, during the written part of the interview, and communication between my offer and start date that it was serious material—not something that you can just kind of bullshit your way through, like, you have to be very prepared for this" (Roberts, 2019; p. 77)

Business Processing Offshoring (BPO) Professionals

"John arrived at commercial content moderation as a sort of consolation prize for having failed the employment examination that Douglas [Staffing] administered to potential workers hoping to get a position on live calls. In some aspect of that process—whether their ability to think on their feet, to speak in colloquial, Americanized English, or some combination of the two—their skills were deemed insufficient, and **they were relegated to the commercial content moderation work that Douglas [Staffing] also undertook**" (Roberts, 2019; p. 178)

Individuals hired as content moderators are not always fully aware of the work that they will be doing

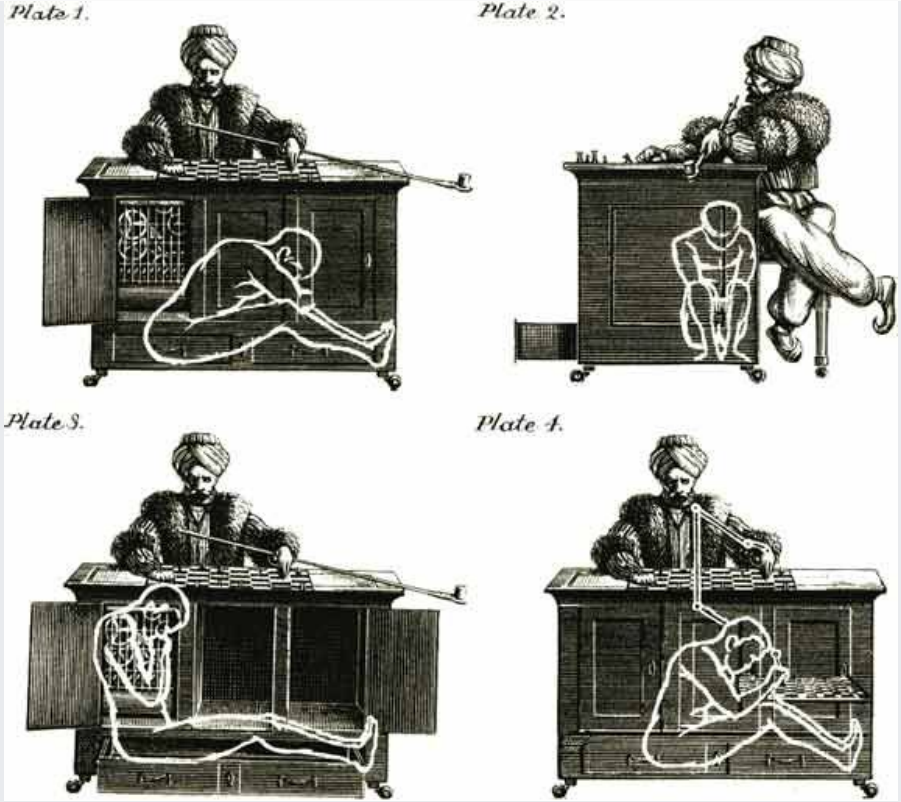
What type of content do moderators view?

"This can include a wide range of material, but often focuses on content that is **highly sexual** or **pornographic**, depicts the **abuse** of adults, the abuse of children (physical and/or sexual), the abuse and **torture** of animals, content coming from **war zones** and other areas besieged by **violent conflict**, and any material that is designed to be **shocking, prurient or offensive by nature**. This is the material that CCM workers seek out, view and adjudicate, second by second of their entire working day, only to do so again the next." (Roberts, 2016)

Torture	Abuse	Hate Speech
Violence	War	Pornography

Humans vs AI: Why isn't content moderated only by algorithms?

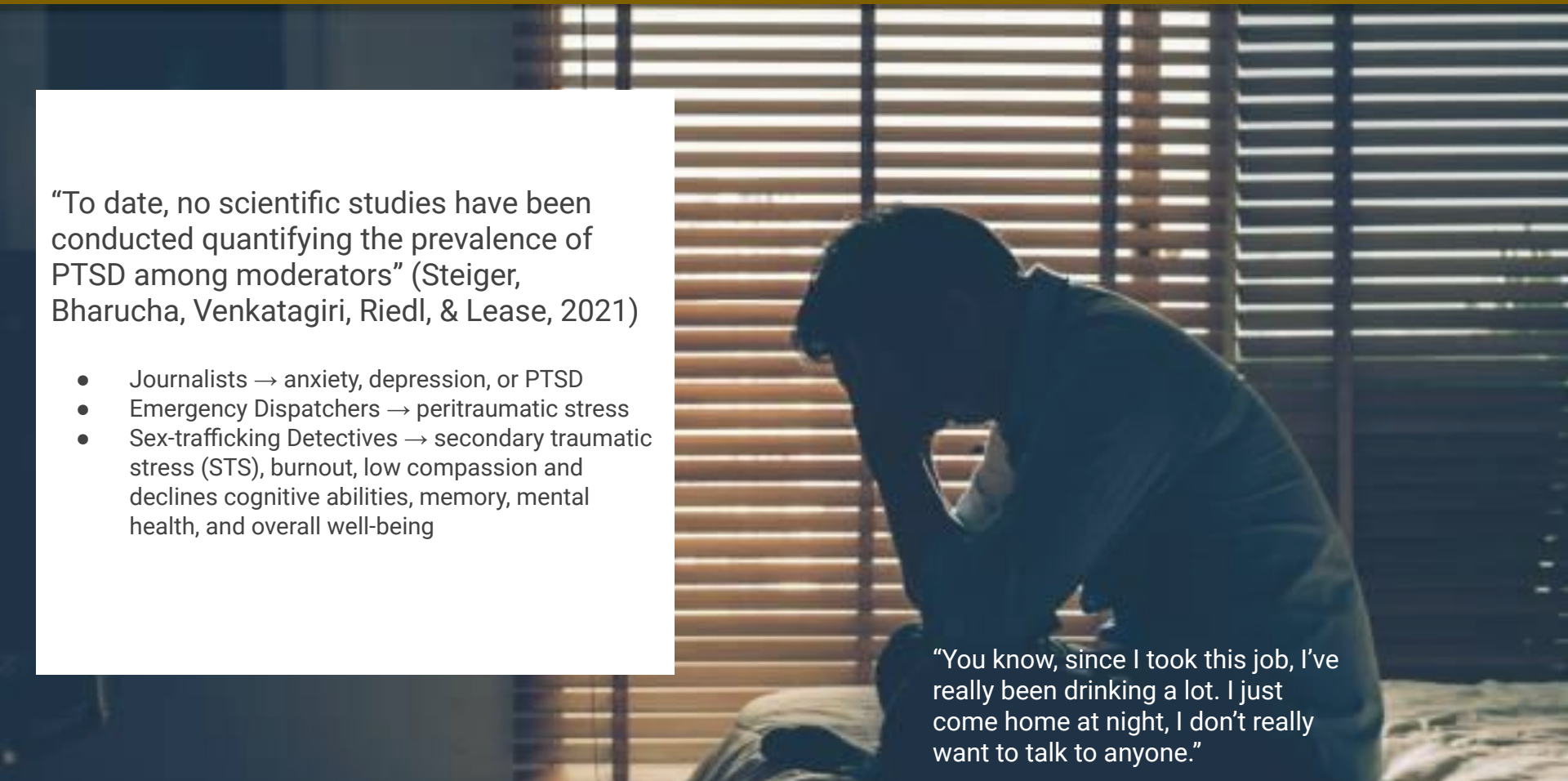
"Ideally, algorithms would automatically detect and filter out such content, and machine learning approaches toward this end are certainly being pursued. **Unfortunately, algorithmic performance remains today unequal to the task in large part due to the subjectivity and ambiguity of the moderation task**, thus making it necessary to fall back on human labor (Roberts 2018a; Roberts 2018b). While social platforms could ask their own users to help police such content, such exposure is typically considered untenable since **these platforms typically want to guarantee their users a protected Internet experience, safe from such exposure, within the confines of their curated platforms**" (Dang, Riedl, & Lease, 2020)



What are the effects of this work?

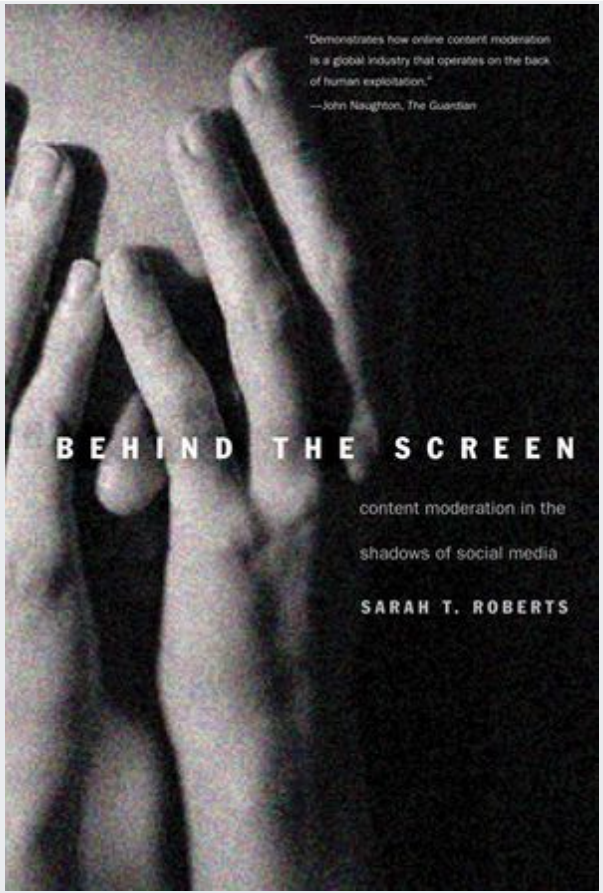
"To date, no scientific studies have been conducted quantifying the prevalence of PTSD among moderators" (Steiger, Bharucha, Venkatagiri, Riedl, & Lease, 2021)

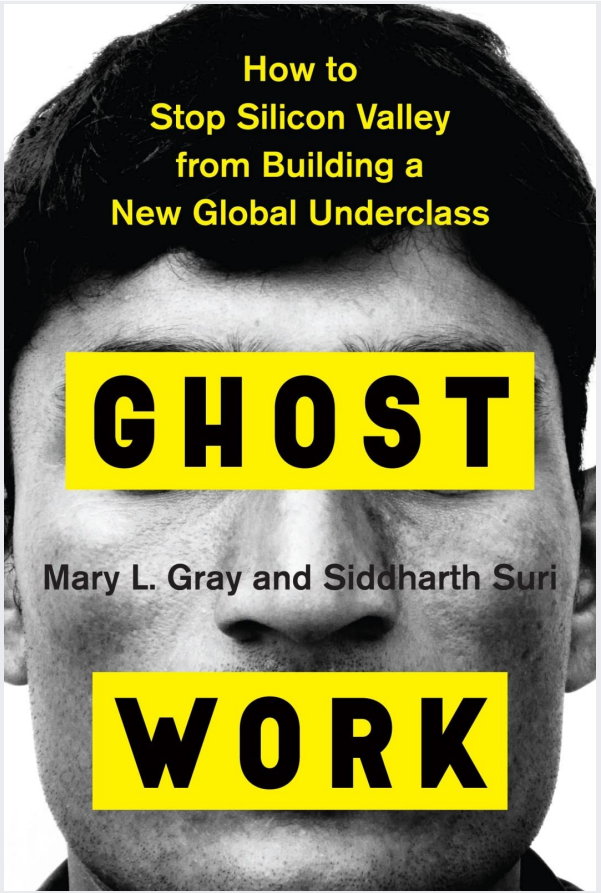
- Journalists → anxiety, depression, or PTSD
- Emergency Dispatchers → peritraumatic stress
- Sex-trafficking Detectives → secondary traumatic stress (STS), burnout, low compassion and declines cognitive abilities, memory, mental health, and overall well-being



"You know, since I took this job, I've really been drinking a lot. I just come home at night, I don't really want to talk to anyone."

Scholarship and Research

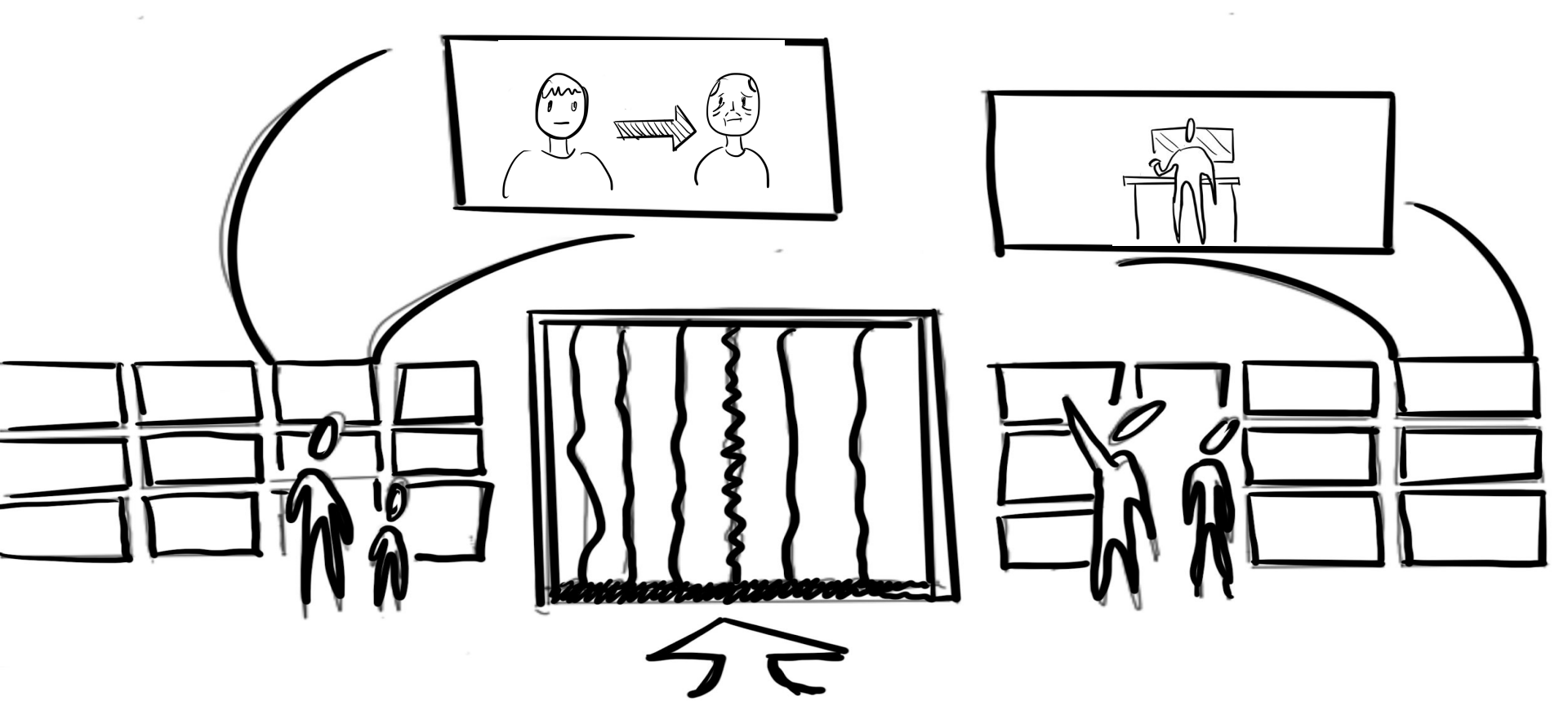




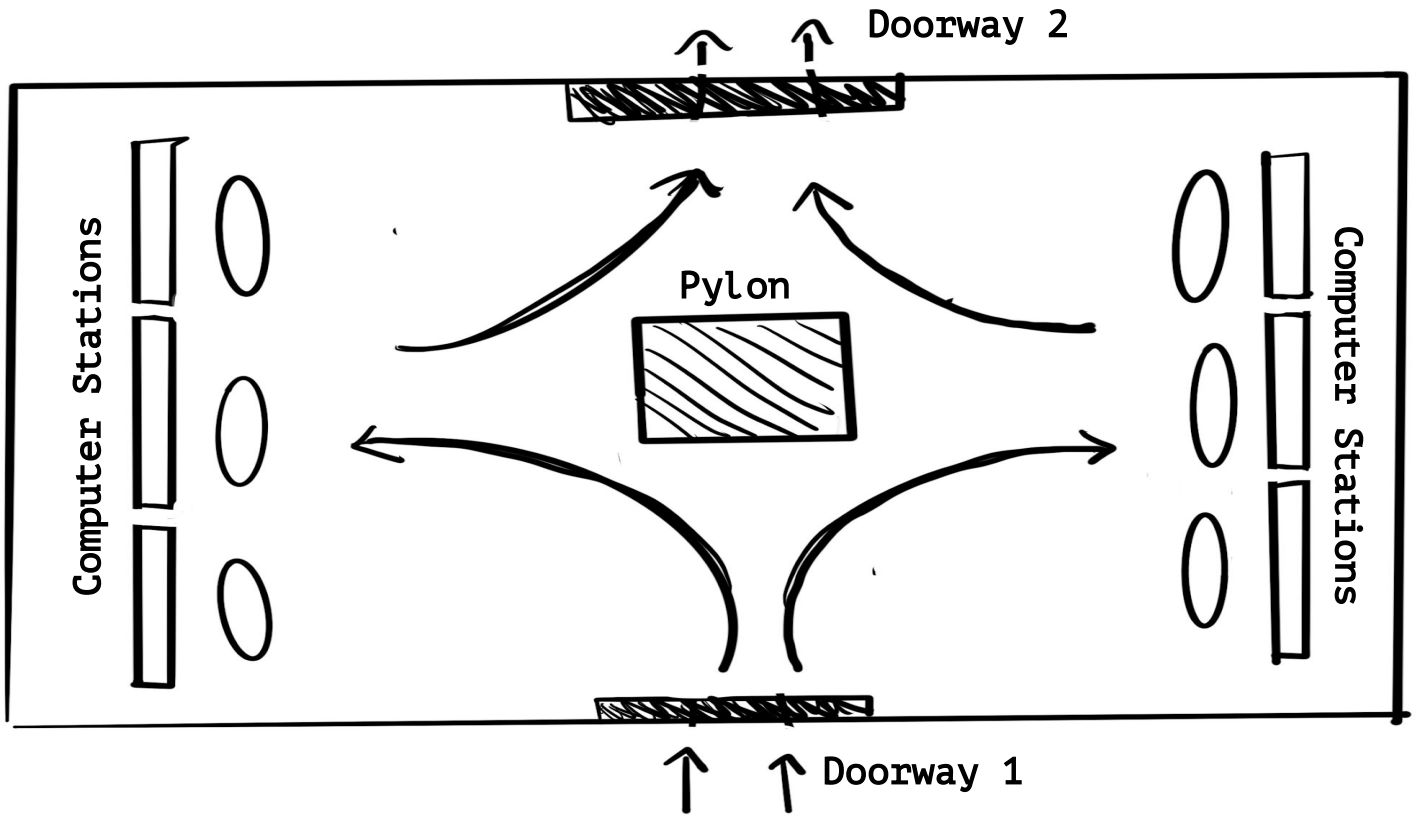
Documentary Film: “The Cleaners” (2018)

See full trailer here: https://www.youtube.com/watch?v=iGCGhD8i-o4

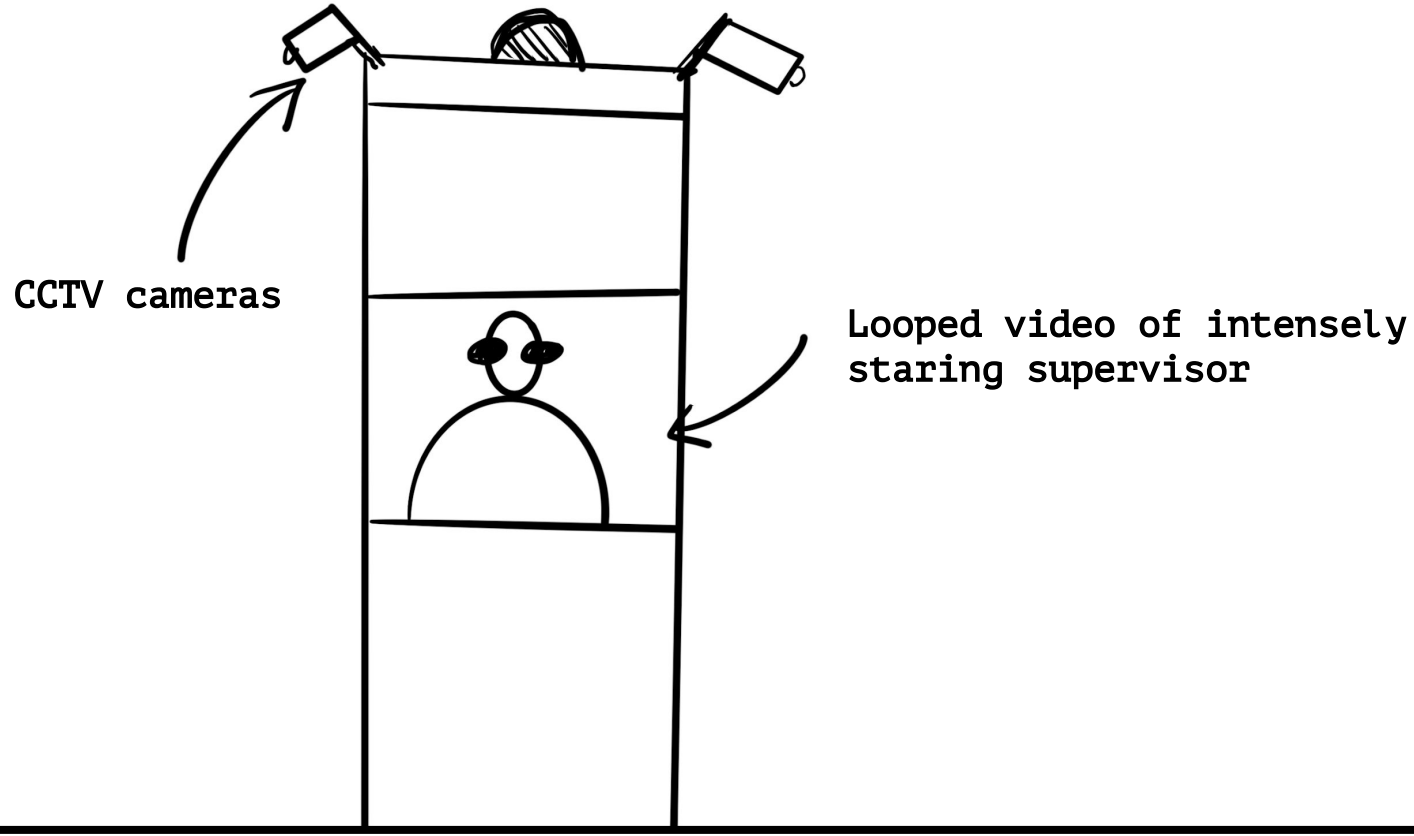
How can we make the emotional toll of commercial content moderation more visible?



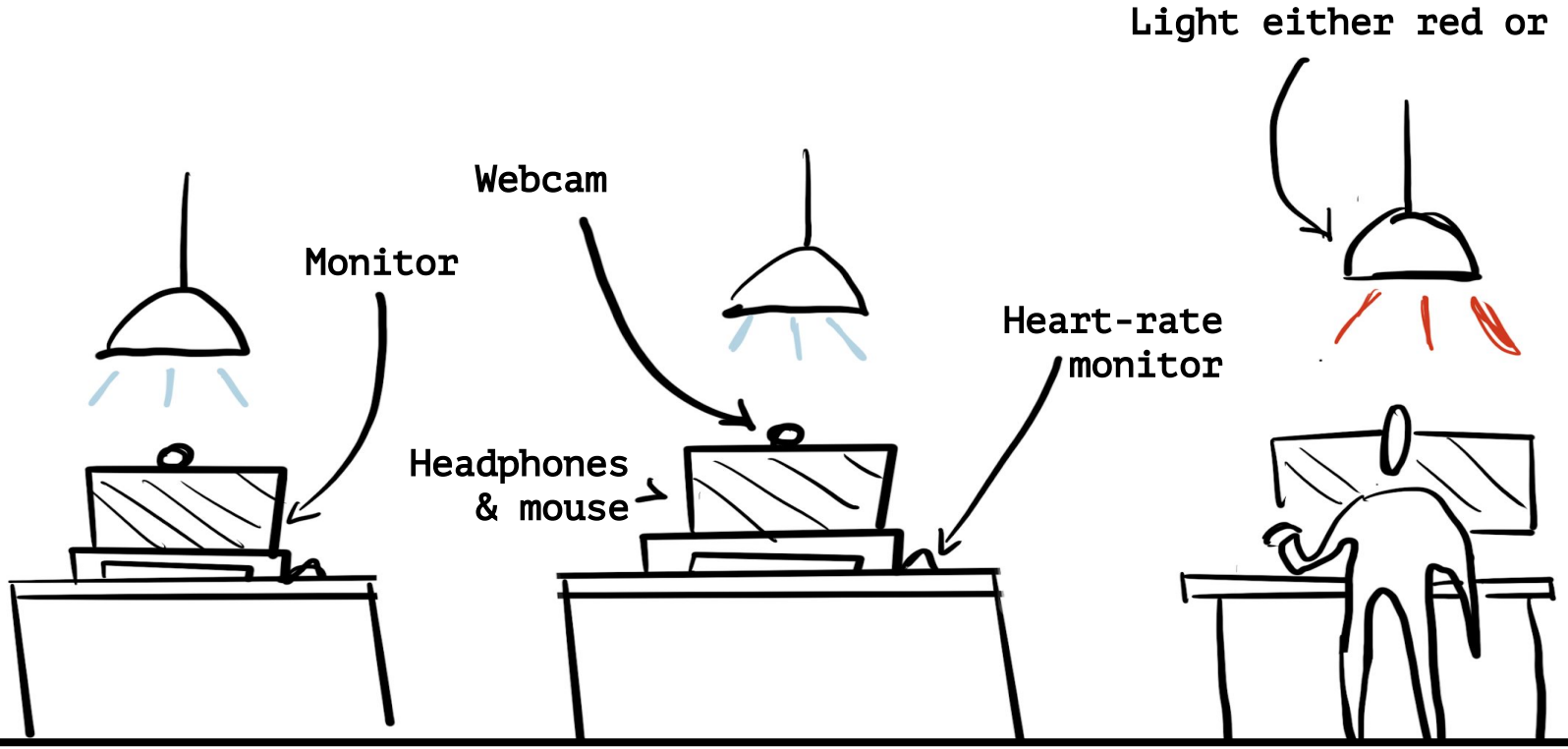
The exhibit begins with a curtain-covered doorway and wall mural made of screens. The screens show videos of aging faces and of CCTV footage of people working on computers. Viewers are unaware of what this footage means but are told they can learn more by entering through the doorway into the next room.



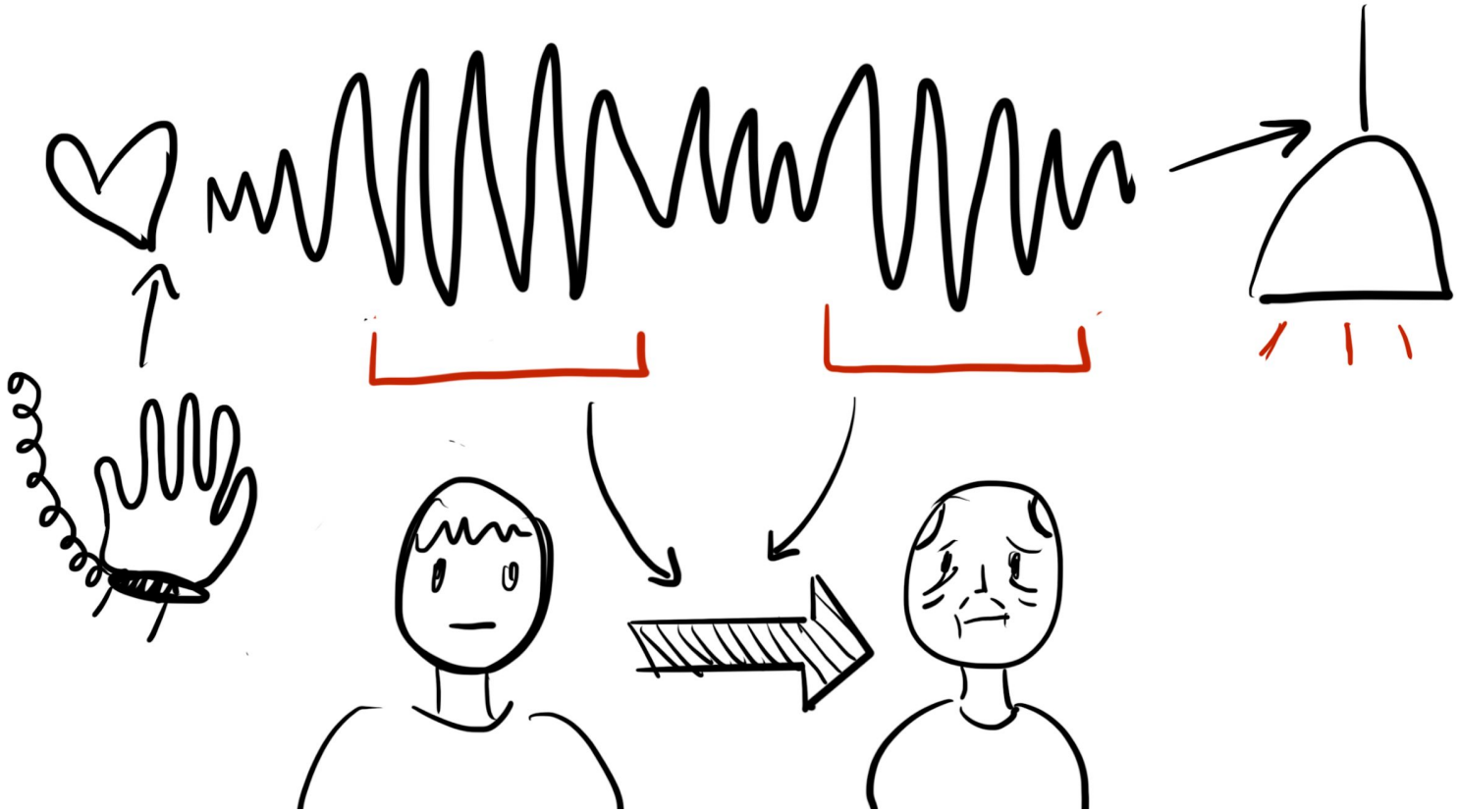
The room is composed of six computer stations and a pylon in the center. Viewers go to an open station and once the finish they exit through a second doorway on the opposite side of the room from which they entered.



The center pylon shows a looped video of a person in business clothes (the supervisor) intensely staring at the computer stations. The top of the pylon also has CCTV cameras that connect to the screens outside the room to live display the people inside.

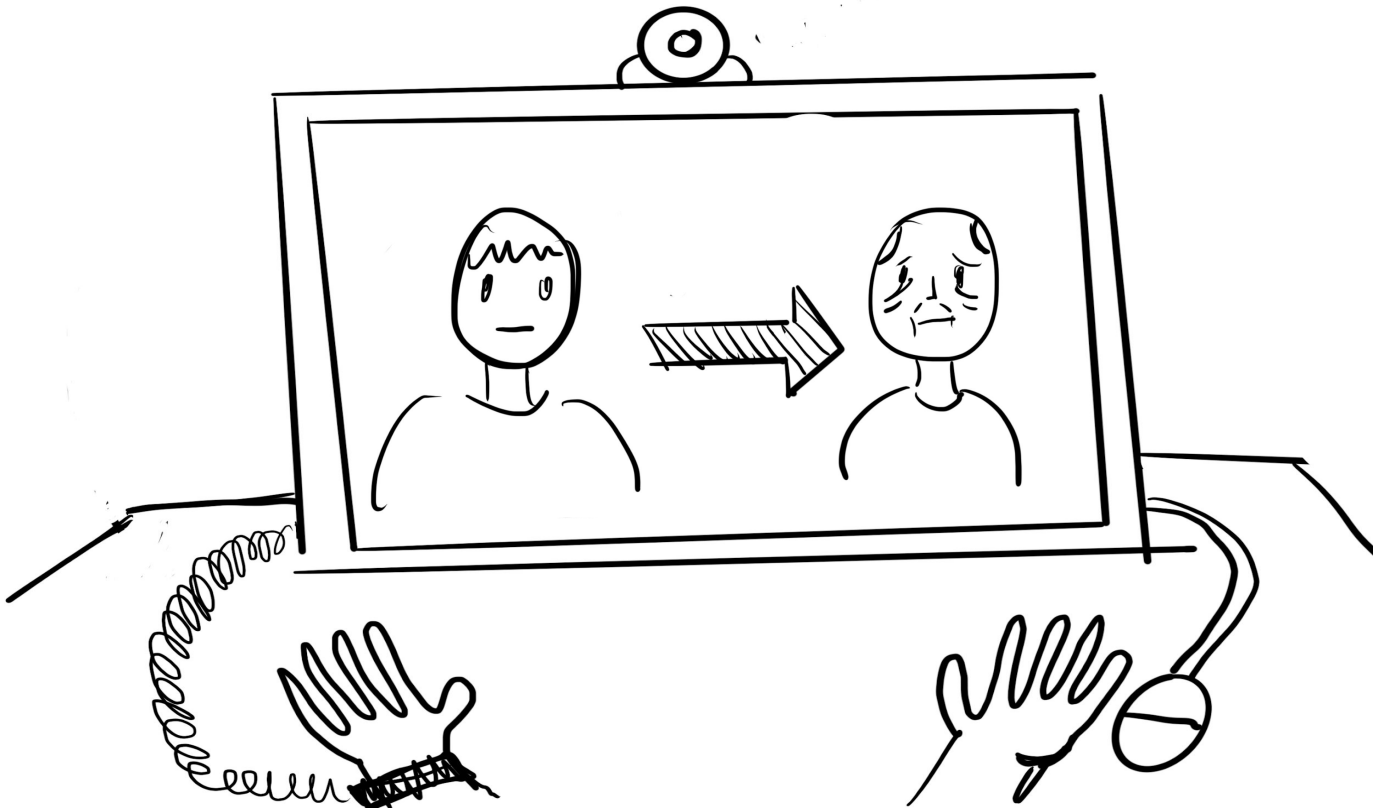


Each computer station has a monitor, mouse, headphones, web camera, and overhead light that either has a cool blue color or bright red color. Only one person can be at a station at a time. Once at the station, the player sees a prompt explaining that the game is categorizing violent and non-violent videos for a social media platform.



The heart monitor captures whenever a player's heart rate goes into a "stress zone" (measured by BPM). This triggers two actions. First, it changes the light at the computer station to red. Second, it begins to gradually add an age filter to the video recording of the player's face.

When a player incorrectly categorizes one of the videos in the game, the game ends. They are then prompted to see a playback of their results.



The player watches the footage of themselves captured by the webcam. This footage displays the face age filter that had been applied throughout the game whenever they were in the stressed heart-rate zone. Thus, the longer they played and the more instances of stress, the older the player will become in the playback footage.

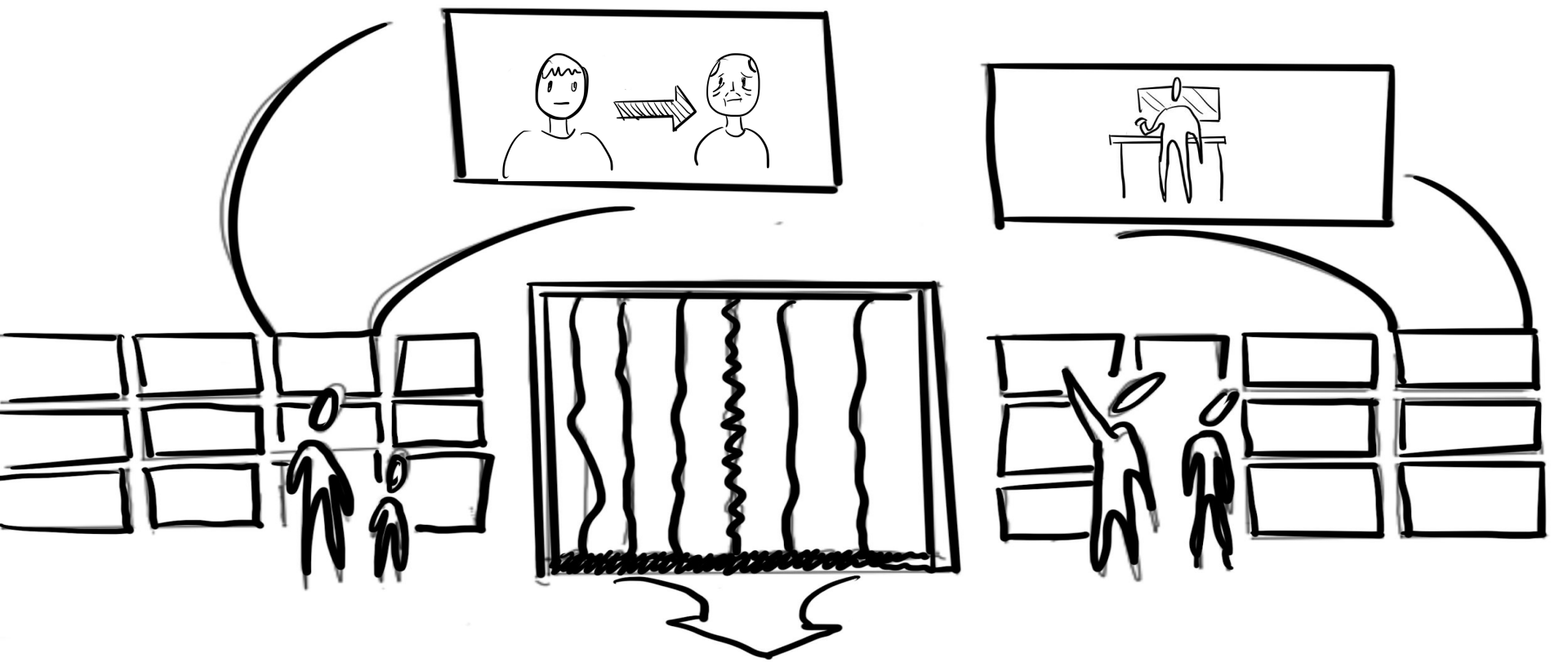
The safety of our platform is our highest priority. We are looking to you to protect this unique and vibrant community. Please remove any videos that contain the following:

1. Extremely dangerous challenges
2. Dangerous or threatening pranks
3. Events promoting or glorifying violent tragedies

We allow videos that depict dangerous acts *if* they're meant to be educational, documentary, scientific, or artistic. Do not remove these videos

Start >

See full video here: https://vimeo.com/717282869



The player then exits the room to a display identical to the entrance. However, now they can see that their aging video has been added to one of the screens in the mural. Having done the game, they now understand the beginning of the exhibit. They can also read a debrief about content moderators and the traumatic impacts of the job.

MAT 255 Final Project Bibliography

Popular Media About Content Moderation

- Baio, Andy. "The Faces of Mechanical Turk." *Waxy.Org*, 20 Nov. 2008, https://waxy.org/2008/11/the_faces_of_mechanical_turk/.
- Chotiner, Isaac. "The Underworld of Online Content Moderation." *The New Yorker*, 5 July 2019. *www.newyorker.com*, https://www.newyorker.com/news/q-and-a/the-underworld-of-online-content-moderation.
- "Digital Content Moderation Expected to Reach $13.60B by 2027." *Brand Minds*, https://brandminds.com/digital-content-moderation-industry-expected-to-reach-13-60b-by-2027/. Accessed 20 May 2022.
- Ghoshal, Abhimanyu. "Microsoft Sued by Its Content Moderators Who Developed PTSD." *TNW | Microsoft*, 12 Jan. 2017, https://thenextweb.com/news/microsoft-sued-by-employees-who-developed-ptsd-after-reviewing-disturbing-content.
- Zuckerberg, Mark. "Adding to Community Operations Team." *Facebook*, 3 May 2017, https://www.facebook.com/zuck/posts/10103695315624661.
- Mattu, Julia Angwin, Jeff Larson,Lauren Kirchner,Surya. "Machine Bias." *ProPublica*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=gl4jHLt-6ZxkcB55q8h_B25ydpK2Tm56. Accessed 24 May 2022.
- Newton, Casey. "The Secret Lives of Facebook Moderators in America." *The Verge*, 25 Feb. 2019, https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona.
- Riesewieck, Mortiz, and Hans Block. *The Cleaners - Official Trailer*. 2018. *YouTube*, https://www.youtube.com/watch?v=iGCGhD8i-o4.
- *The Humans Working Behind the AI Curtain*. https://hbr.org/2017/01/the-humans-working-behind-the-ai-curtain. Accessed 20 May 2022.

Academic Writing about Content Moderation

- Dang, Brandon, et al. *But Who Protects the Moderators? The Case of Crowdsourced Image Moderation*. arXiv:1804.10999, arXiv, 4 Jan. 2020. *arXiv.org*, http://arxiv.org/abs/1804.10999.
- Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Illustrated edition, Yale University Press, 2018.
- Gray, Mary L., and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Illustrated edition, Harper Business, 2019.
- Karunakaran, Sowmya, and Rashmi Ramakrishan. "Testing Stylistic Interventions to Reduce Emotional Impact of Content Moderation Workers." *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, vol. 7, Oct. 2019, pp. 50–58.
- Riedl, Martin J., et al. "The Downsides of Digital Labor: Exploring the Toll Incivility Takes on Online Comment Moderators." *Computers in Human Behavior*, vol. 107, June 2020, p. 106262. *ScienceDirect*, https://doi.org/10.1016/j.chb.2020.106262.
- Roberts, Sarah. "Commercial Content Moderation: Digital Laborers' Dirty Work." *Media Studies Publications*, Jan. 2016, https://ir.lib.uwo.ca/commpub/12.

- Roberts, Sarah T. *Content Moderation*. 2017. *escholarship.org*, https://escholarship.org/uc/item/7371c1hf.
- Ross, Joel, et al. "Who Are the Crowdworkers? Shifting Demographics in Mechanical Turk." *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, Association for Computing Machinery, 2010, pp. 2863–72. *ACM Digital Library*, https://doi.org/10.1145/1753846.1753873.
- Steiger, Miriah, et al. "The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support." *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, 2021, pp. 1–14. *ACM Digital Library*, https://doi.org/10.1145/3411764.3445092.
- Waseem, Zeerak. "Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter." *Proceedings of the First Workshop on NLP and Computational Social Science*, Association for Computational Linguistics, 2016, pp. 138–42. *ACLWeb*, https://doi.org/10.18653/v1/W16-5618.

Exhibition Design Inspirations

- Jaar, Alfredo. *The Sound of Silence, 2006*. Mixed Media, 2006, https://www.wikiart.org/en/alfredo-jaar/the-sound-of-silence.
- Klimov, Elem. *Come and See*. 1985. *YouTube*, https://www.youtube.com/watch?v=UHaSQU-4wss.
- "Milgram Experiment." *Wikipedia*, 19 May 2022. *Wikipedia*, https://en.wikipedia.org/w/index.php?title=Milgram_experiment&oldid=1088648826.
- "Panopticon." *Wikipedia*, 24 May 2022. *Wikipedia*, https://en.wikipedia.org/w/index.php?title=Panopticon&oldid=1089647072.
- Thiel, Tamiko. *Lend Me Your Face!* Mixed Media, 2020, https://www.tamikothiel.com/lendmeyourface/.

Mock-Up Game Resources

Link to Mock-Up Final Video: https://vimeo.com/717282869

- Adobe Inc. *Adobe After Effects*. https://www.adobe.com/products/aftereffects.html
- Adobe Inc. *Adobe Premiere Pro*. https://www.adobe.com/products/premiere.html
- Elesawy, Mohamed, et al. *Real Life Violence Situations Dataset*. 2019, https://www.kaggle.com/mohamedmustafa/real-life-violence-situations-dataset.
- Goncharov, Yaroslav. *FaceApp*. https://www.faceapp.com/.
- Sean Dove. *EbSynth Tutorial - Digital Ageing Visual Effect Using A.I*. 2019. *YouTube*, https://www.youtube.com/watch?v=mOhAJ18V4nA.
- Samkov, Ivan. *Woman Crying with Tear*. https://www.pexels.com/video/woman-crying-with-tear-6689304/. Accessed 1 June 2022.
- Secret Weapons. *EbSynth*. https://ebsynth.com/.