CONTENTR

CONTENT MODERATION BY DESIGN (CMBD) GAME INSTRUCTIONS
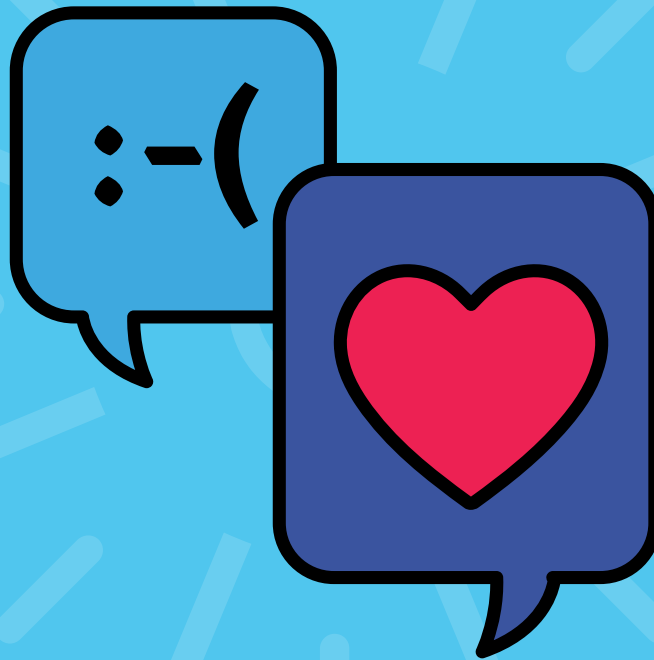
Fall 2021

GAME PURPOSE

By playing the role of both a startup social media platform policy trust and saftey team and a content moderator, participants can begin to experience some of the challenges associated with moderating user generated online content in a way that balances values such as free expression and community safety.

Game works best with **3** players, but other sized groups can still play

MATERIALS

- Cards (print and cut the cycles you need)
- Flat surface
- 3 Sticky notes or small pieces of paper and an ink pen
- A scorecard print out or digital (link https://tinyurl.com/yr6xxfdb)

THE STORY

Congratulations, your brand new social media startup, Contentr has just received funding from investors! You've been following the news and are determined to avoid the same mistakes as your predecessors, so the first place you want to start is to develop your content moderation policy. There are three rounds to the game: first you'll work as a team to develop the policy that will help you shape the kind of platform you want to grow. You'll then switch roles from policy developer to content moderator, where you'll use your policy to make moderation decisions, based on real life examples. Finally, you'll see how your decisions play out, calculating your final score based on real life examples of moderation decisions, and how those decisions affect two areas: free expression and community safety.

You will begin with 500 free expression points and 500 community safety points. Free expression points are important because they provide space for your users to express themselves and community safety points are important because they ensure your users are free from potential off-platform harms. The more you are able to balance your points, the more "mass appeal" your game will have, resulting in more ad revenue.

CONTENT WARNING

This game involves discussing descriptions of (but not viewing) sexually explicit content, harassment, hate speech, self-harm, illegal activity, misinformation, and violent content. The purpose of the game is to provide deeper understanding of platform governance which is inherently challenging, frustrating, and sometimes upsetting, and these are emotions you may feel as you play the game. **You should take breaks and feel free to leave the game if needed.**

Additionally, we encourage the use of John Stavropoulos' X-card strategy (link here: http://tinyurl.com/x-card-rpg). If a card is particularly uncomfortable to engage with, simply flip it over, this will signal to the other players to limit conversation about that card.

BEGIN: SETUP

The game is divided into three cycles, each focuses on different areas of controversial content. At the start of each cycle, **you will receive an investment which will allow the Contentr Trust & Safety team to make decisions related to growth.** Throughout the game you will experience changes to both free expression and community safety points and your budget. To track these changes, choose one player to be your Chief Financial Officer: they will track changes in profit, loss and points (use the digital or printed balance sheet).

CYCLES

Begin by sorting the cards into three piles, one for each cycle. You will notice, the cycles are grouped by category of controversial content and the cycle is printed on the card:

Cycle 1:
Sexually Explicit Content & Illegal Activity

Cycle 2:
Self Harm & Graphic Content

Cycle 3:
Harassment, Hate Speech & Quality Contributions

Depending on time you may choose to only play one cycle. Each cycle lasts approx. 45 minutes. For the deepest experience, play the cycles in order, but playing a single cycle is also fun.

ROUNDS

Each cycle has rounds which are signaled by the border of the cards. Split each cycle into four piles: Investment (orange), Policy (light blue), Content (grey), Event Cards (dark blue).

INVESTMENT

Investment
One card that represents incoming cash for the Trust and Safety Team

POLICY

Round 1: Policy
Playing the role of the trust and safety team, write Contentr's community guidelines/platform policies

CONTENT

Round 2: Content
Playing the role of the content moderator, enact the Contentr policies (one nuance card per cycle)

EVENT

Round 3: Event
Events based on real life examples of pushback social media companies have faced over the years. Some consequences are based on decisions made in rounds 1 and 2, and other events happen regardless of the platform's efforts.

PLAY YOUR FIRST CYCLE

Begin with **investment (orange card)**: Read your investment card and add to your balance sheet.

You may be wondering, "wait, what is Contentr?"

Contentr is a start up social media company attempting to take market share from the today's leading platforms. The site allows users to connect with friends, community members and high profile figures. It includes a range of user generated content: text, photos, videos, articles. Lastly, it includes ads and a news feed.

The company's values will be shaped as you form policies.

ROUND 1: SET POLICIES

In this round, you will decide what kind of content your company wants to allow, and what it wants to prohibit.

To start, Find the description of the Trust and Safety role and review the card.

ROLE : POLICY

TRUST &
SAFTEY
TEAM

- You care about the company's success
- You need users to spend time on your platform in order to attract investors and advertisers (you want to keep both free expression points and community safety points)
- You are doing your best to write clear policies that capture the wide range of content

SORT

You'll all work together to decide which policies to enact, but in the interest of time, since there are three of you, when you disagree, majority rules.

Rather than starting from scratch, you'll select from a set of existing policies. (It is very common for new platforms to look at the policies of competitor social media platforms for inspiration.) You'll do this by sorting the light blue policy cards into one of two categories: Allowed or Banned.

The policy cards and content cards have numbers on them. The numbers are important for interpreting consequences later in the game, but you do not need to play the cards in order.

Find the cards that say "Ban" and "Allow" and place them at the top of your table. Throughout the game, you will use these as headers, and place content cards under them:

SECTION 230 & PRE CYCLE

To guide you, you have your own values as a company, but you also need to follow existing legislation. First is Section 230 of the Communications Decency Act, which provides your team with a "shield" and a "sword." The shield means that your platform cannot be held liable for content users post with a few exceptions (pre-cycle cards, discussed below). The sword means your platform can "moderate" content on the platform as long as you do not "publish" or "edit."

But there are some exceptions, represented with pre-cycle cards. As a small company you will do your best to avoid this content, as a large legal battle may bankrupt you. (Place the pre-cycle cards under "banned")

- The Digital Millennium Copyright Act (DMCA) states intermediaries such as your platform have "indirect liability" if they do not make efforts to take down copyrighted material.
- The Fight Online Sex Trafficking Act (FOSTA), was passed in 2018 and outlaws content that facilitates traffickers in advertising the sale of unlawful sex acts with sex trafficking victims.
- Federal law requires that you take down and report content representing child sexual exploitation as soon as reasonably possible.

Other than these three types of banned content, you have the shield and the sword - moderation decisions are up to you. We recommend placing the Section 230 card at the top of the table where you can see it.

SECTION 230

CONTENTR IS NOT LIABLE | CONTENTR CAN MODERATE
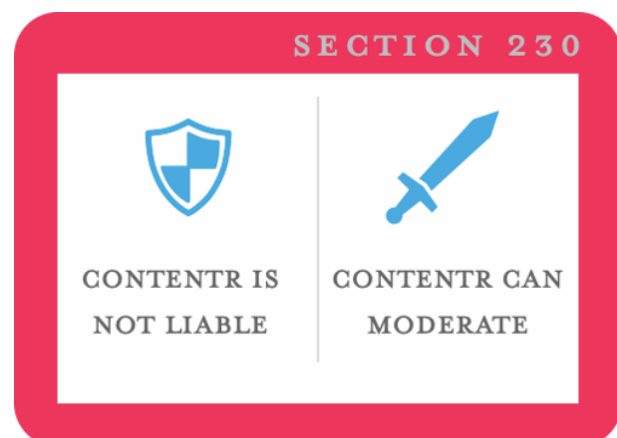
ROUND 2: MODERATING CONTENT

Excellent work! You now have a policy that you'll use to make decisions about the kind of content that's allowed on your platform. In this round you'll switch roles, from policy developer to content moderator.

Find the description of the Content Moderator role and review the card.

Once you understand the role, sort the grey cards into allow or banned based on the policy cards sorted in round 1. Stack below or next to the light blue policy cards.

ROLE:CONTENT

CONTENT MODERATOR

- You do your best to follow the policy team guidelines
- You don't have much time to make a decision about what content is allowed or banned
- You often are asked to make decision about content without context
- When you think a policy is unclear you can send feedback to your Policy Team (1 nuance card per cycle)

NUANCE CARDS

Finally, as in real life, a content moderator can suggest policy changes to your company's trust and safety team based on their personal experiences. At the end of the cycle, you will be able to add nuance to one existing policy. For example, if a gray card is placed in the "allow" or "banned" pile based on existing policies and you want to switch it, you can edit a policy to make the switch.

To edit a policy, use the pen and scrap of paper to write edits to one existing policy card and place the scrap paper next to a card. Once you edit don't forget to move the content card (grey card) as necessary.

ROUND 3: CONSEQUENCES

You've made it to round 3! All the hard work is done and now you get to see the outcomes of your decisions. You started with 500 free expression (FE) points, and 500 community safety (CS) points.

In this round you'll read the event cards (Dark Blue Border). The blue cards are based on examples of pushback social media companies have faced over the years. Some consequences are based on decisions made in rounds 1 and 2, and other events happen regardless of the platform's efforts.

You will notice different types of event cards, including "Society" (events that occurred in the society at large) and "algorithm" (automated decision-making trained rightly or wrongly based on decisions made during rounds 1 and 2). **As you read the event cards, be sure to adjust your balance sheet.**

Events are based on real world push back that leading social media platforms have received.

EVENT

CYCLE 2

ALGORITHM

A pandemic closes all non-essential businesses forcing you to send your moderators home and rely only on algorithmic moderation. If you have more content cards in your banned pile you have trained your algorithm to over-censor **lose 30 FS points**. If you have more cards in your allowed pile, dangerous content is spreading on the platform, **lose 30 CS points.**

OPTIONS TO COMPLETE THE GAME & REFLECTION

Option: play another cycle.

If you choose to play another cycle: Stack the cards currently in your "Ban" column and place next to the pink "Ban" card. Stack the cards currently in your "Allow" column and place them next to the pink "Allow" card.

After Cycle 3...

Option: The game is complete.

Take a look at your final score (both the FE/CS points and revenue). Do you feel like you "won"? why? why not?

Takeaways from the game:
- How did your views of social media change after playing the game?
- Which decisions were easy, why?
- Which decisions were challenging, why?

BACKGROUND

Social media content moderation practices vary from company to company, are inherently opaque and span well beyond simply allowing or banning content. This game is meant to give a taste of the challenges posed by hosting a site for user generated content but should in no way be interpreted as a comprehensive overview of trust and safety practices.

The categories of content and roles were informed by:

Block, Hans, et al. (2018).The Cleaners. (film)

Gillespie, T. (2018). Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

Roberts, S. T. (2019). Behind the screen: Content moderation in the shadows of social media. Yale University Press.

ACKNOWLEDGEMENTS

CARDS INSPIRED BY

(2016, May 20). 'A warning to other states': PETA wins $250,000 over Idaho 'Ag-gag' case. *RT*. https://www.rt.com/usa/343834-peta-paid-fees-ag-gag/

(2019, April 23). Bernie Sanders vows to round up remaining ISIS members, allows them to vote. *Babylon Bee*. https://babylonbee.com/news/bernie-sanders-vows-to-round-up-remaining-isis-members-allow-them-to-vote

Alba, D. (2021, March 19). How Anti-Asian Activity Online Set the Stage for Real-World Violence. *New York Times*. https://www.nytimes.com/2021/03/19/technology/how-anti-asian-activity-online-set-the-stage-for-real-world-violence.html

Anderson, C. A., & Carnagey, N. L. (2009). Causal effects of violent sports video games on aggression: Is it competitiveness or violent content?. Journal of experimental social psychology, 45(4), 731-739.

Asher-Schapiro, A. (2017, Nov 2). YouTube and Facebook are removing evidence of atrocities, jeopardizing cases against war criminals. *The Intercept*. https://theintercept.com/2017/11/02/war-crimes-youtube-facebook-syria-rohingya/

Arnold, J. (2019, May 23). Artist who turns MAGA hats into symbols of hate speech banned from Facebook. *WUSA9*. https://www.wusa9.com/article/news/md-artist-banned-from-facebook-for-controversial-maga-hat-art/65-1cd87d9d-2761-4ff8-a3e4-209919489ccb

Booth, R. & Weaver, M. (2015, Jun 5). 'Baby yoga' video on Facebook sparks internet censorship debate. *The Guardian*. https://www.theguardian.com/technology/2015/jun/05/baby-yoga-video-facebook-internet-censorship-debate

Bullen, S. (2019, Aug 26). Instagram's Graphic Self-Harm Content Ban is not Enough. *PublicEar*.https://medium.com/the-public-ear/instagrams-graphic-self-harm-content-ban-is-not-enough-5df3060f41cc

CARDS INSPIRED BY

Brandom, R. & Newton, C. (2017, Feb 24). Twitter is locking accounts that swear at famous people. *The Verge*.
https://www.theverge.com/2017/2/24/14719828/twitter-account-lock-ban-swearing-abuse-moderation

DANTE [@1917Dante]. (2020, Jun 6). @jack @Twitter#BLM #BlackLivesMattters #BlackLivesMatterDC #DCProtests #DC #DCProud My Twitter was blocked for almost 7 days because of the following tweet and photos. Can you please explain me how I violated the twitter rules? [Image attached] [Tweet]. Twitter.
https://twitter.com/1917Dante/status/1269390501880496136

Delfino, R. A. (2020). Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act. Actual Probs. Econ. & L., 105.

Farokhmanesh, M. (2018, Jun 4). YouTube is still restricting and demonetizing LGBT videos – and adding anti-LGBT ads to some. *The Verge*.
https://www.theverge.com/2018/6/4/17424472/youtube-lgbt-demonetization-ads-algorithm

Frishberg, H. (2019, Oct 29). 'Sexual' use of eggplant and peach emojis banned on Facebook, Instagram. *New York Post*. https://nypost.com/2019/10/29/sexual-use-of-eggplant-and-peach-emojis-banned-on-facebook-instagram/

Galindo, Y. (2017, Oct 25). Machine Learning Detects Marketing and Sale of Opioids on Twitter. *UC San Diego News Center*.
https://ucsdnews.ucsd.edu/pressrelease/machine_learning_detects_marketing_and_sale_of_opioids_on_twitter

Gillbert, B. (2019, Nov 6). The 10 most-viewed fake-news stories on Facebook in 2019 were just revealed in a new report. *Insider*.
https://www.businessinsider.com/most-viewed-fake-news-stories-shared-on-facebook-2019-2019-11

CARDS INSPIRED BY

Gore, I. (2016, May 7). Illma Gore:'If anyone is going to be threatened by a small penis, it's Trump' *The Guardian*. https://www.theguardian.com/us-news/2016/may/07/donald-trump-penis-painting-ilma-gore

Hamilton, I.A. (2018, Aug 30). Facebook apologises to the Anne Frank Center for removing image of naked child holocaust victims. *Insider.* https://www.businessinsider.com/facebook-apologises-for-removing-anne-frank-center-child-holocaust-image-2018-8

Hesse, J. (2016, Feb 17). Facebook cracks down on marijuana firms with dozens of accounts shut down. *The Guardian*. https://www.theguardian.com/us-news/2016/feb/17/facebook-marijuana-cannabis-businesses-crackdown

Kesslen, B. (2019, Feb 4) Instagram to hide self-harm images in the wake of rising teen suicides. *NBC News*. https://www.nbcnews.com/news/us-news/instagram-hide-self-harm-images-wake-rising-teen-suicides-n966781

Johnson, M. (2016, Dec 14). How fake news led Dylann rook to murder nine people. *The Undefeated*. https://theundefeated.com/features/how-fake-news-led-to-dylann-roof-to-murder-nine-people/

Keeley, M. (2019, Aug 22). YouTube's AI Flagged Robot Battles as Animal Cruelty and Removed Them. *Newsweek*. https://www.newsweek.com/youtubes-ai-flagged-robot-battles-animal-cruelty-removed-them-1455806

Kelly Garrett, R (2019, Aug 16). Too many people think satirical news is real. *The Conversation*. https://theconversation.com/too-many-people-think-satirical-news-is-real-121666

Lorenz, T. (2017, Dec 5). Facebook is Banning Women for Calling Men 'Scum.' *Daily Beast*. https://www.thedailybeast.com/women-are-getting-banned-from-facebook-for-calling-men-scum

Mackey, A. (2020, Sep 17). Plaintiffs Continue Effort to Overturn FOSTA, One of the Broadest Internet Censorship Laws. *EFF*. https://www.eff.org/deeplinks/2020/09/plaintiffs-continue-effort-overturn-fosta-one-broadest-internet-censorship-laws

CARDS INSPIRED BY

Matney, L. (2018, May 25). Facebook has a very specific Pepe the Frog policy, report says. *TechCrunch*. https://techcrunch.com/2018/05/25/facebook-has-a-very-specific-pepe-the-frog-policy-report-says/

Magistretti, B. (2018, April 5). Facebook's ad policies are hurting women's heath startups. *VentureBeat*. https://venturebeat.com/2018/04/05/facebooks-ad-policies-are-hurting-womens-health-startups/

Melendez, S. (2020, March 3). 'I have a duty to do this': Meet the Redditors fighting 2020's fake news war. *Fast Company*. https://www.fastcompany.com/90466966/i-have-a-duty-to-do-this-meet-the-redditors-fighting-2020s-fake-news-war

Messent, P. (2011, Jan 5). Censoring Mark Twain's 'n-words' is unacceptable. *The Guardian*. https://www.theguardian.com/books/booksblog/2011/jan/05/censoring-mark-twain-n-word-unacceptable

Mpsos, N. (2017, Feb 24). Instagram bans 'nude' video of Himba woman. *IOL*. https://www.iol.co.za/capetimes/arts-portal/instagram-bans-nude-video-of-himba-woman-7913679

Notopoulos, K. (2017, Dec 2). How Trolls Locked my Twitter Account for 10 Days, and Welp. *BuzzFeed.News*.https://www.buzzfeednews.com/article/katienotopoulos/how-trolls-locked-my-twitter-account-for-10-days-and-welp

Oppel, R. (2013, Apr 7). Taping of Farm Cruelty is Becoming the Crime. *New York Times*. https://www.nytimes.com/2013/04/07/us/taping-of-farm-cruelty-is-becoming-the-crime.html

Oberhaus, D. (2018, Aug 29). Life on the internet is hard when your last name is 'Butts'. *Vice*. https://www.vice.com/en/article/9kmp9v/life-on-the-internet-is-hard-when-your-last-name-is-butts

CARDS INSPIRED BY

Paul, K. (2020, May 12). Facebook to pay $52m for failing to protect moderators from 'horrors' of graphic content. *The Guardian*. https://www.theguardian.com/technology/2020/may/12/facebook-settlement-mental-health-moderators

Patel M, Lee AD, Clemmons NS, et al. National Update on Measles Cases and Outbreaks — United States, January 1–October 1, 2019. MMWR Morb Mortal Wkly Rep 2019;68:893–896. DOI: http://dx.doi.org/10.15585/mmwr.mm6840e2

Parkinson, H.J. (2015, Nov 3). A surprisingly difficult question for facebook: do I have boobs now? *The Guardian*. https://www.theguardian.com/technology/2015/nov/03/facebook-instagram-do-i-have-boobs-now

Peterson, A. (2016, Jul 7). Why the Philando Castile police-shooting video disappeared from Facebook – then came back. *Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2016/07/07/why-facebook-took-down-the-philando-castile-shooting-video-then-put-it-back-up/

R/AMA - comment by U/Appls on "I did content moderation for Facebook for almost a year. Ama". reddit. (n.d.). Retrieved November 6, 2021, from https://www.reddit.com/r/AMA/comments/90f3dv/i_did_content_moderation_for_facebook_for_almost/e2pyq3e

Richards, K. (2018, Jan 26). Save Your Apologies: Facebook Deletes Black Woman's Post About White Women. *BLAVITY: NEWS*. https://blavity.com/save-your-apologies-facebook-deletes-black-womans-post-about-white-women?category1=feminism&subCat=news&category2=news

Samakow, J. (2017, Dec 6). 'Stop Censoring Motherhood' Movement Takes Aim At Facebook and Instagram. *Huffpost*. https://www.huffpost.com/entry/facebook-instagram-censoring-motherhood_n_5578300

Sex School [@SexSchoolHub]. (2019, Jan 26). So we tried to post the photo of this potato on Instagram. AND IT GOT CENSORED Hey @instagram Since when potatoes are not allowed on your platform?? #Instagram #censorship #potatoban [Image attached] [Tweet]. Twitter. https://twitter.com/SexSchoolHub/status/1089176609792376834

CARDS INSPIRED BY

Szalavitz, M. (2019, July 2). Facebook is censoring posts that could save opioid users' lives. *Vice.* https://www.vice.com/en/article/qv75ap/facebook-is-censoring-harm-reduction-posts-that-could-save-opioid-users-lives

Tiffany, K. (2019, Jun 19). The Hired guns of Instagram: Companies can't advertise on social media—so they have female influencers do it for them. *Vox.* https://www.vox.com/features/2019/6/19/18644129/instagram-gun-influencers-second-amendment-tactical-community

Wegmann, P. (2018, Feb 27). Google tried censoring 'gun' shopping searches. It backfired. *Washington Examiner.* https://www.washingtonexaminer.com/google-tried-censoring-gun-shopping-searches-it-backfired

When you find out your daughter likes black guys Whatca Gonna do brother?: Wrestling meme on Me.me. me.me. (n.d.). Retrieved November 6, 2021, from https://me.me/i/when-you-find-out-your-daughter-likes-black-guys-whatca-15048559

Wong, J.C. (2017, May 19). Facebook blocks Pulitzer-winning reporter over Malta government expose. *The Guardian.* https://www.theguardian.com/world/2017/may/19/facebook-blocks-malta-journalist-joseph-muscat-panama-papers

Yu, Y. (2021, March 26). Social media chiefs grilled in US Congress over anti-Asian content. *Nikkei Asia.* https://asia.nikkei.com/Business/Technology/Social-media-chiefs-grilled-in-US-Congress-over-anti-Asian-content

A L L O W

I N S T R U C T I O N S

Contentr, Content Moderation by Design was developed at University of Maryland Ethics and Values in Design Lab in 2021.

The details and instructions are avilible at https://evidlab.umd.edu/content-moderation-by-design-the-game/

ROLE:CONTENT

CONTENT MODERATOR

- You do your best to follow the policy team guidelines
- You don't have much time to make a decision about what content is allowed or banned
- You often are asked to make decision about content without context
- When you think a policy is unclear you can send feedback to your Policy Team (1 nuance card per cycle)

B A N

ROLE:POLICY

TRUST & SAFETY TEAM

- You care about the company's success
- You need users to spend time on your platform in order to attract investors and advertisers (you want to keep both free expression points and community safety points)
- You are doing your best to write clear policies that capture the wide range of content

SECTION 230

CONTENTR IS NOT LIABLE	CONTENTR CAN MODERATE

EVENTS

SOCIETY	ALGORITHM
Events that occurred in the society at large	Automated decision-making trained rightly or wrongly based on decisions made during rounds 1 and 2.

POLICY (BANNED)

PRE–CYCLE

Material protected by copyright (notice and takedown, DMCA 1998)

POLICY (BANNED)

PRE–CYCLE

Content that facilitates traffickers in advertising the sale of unlawful sex acts with sex trafficking victims (FOSTA-SESTA, 2018)

POLICY (BANNED)

PRE–CYCLE

Take down and report child pornography (18 U.S. Code 2258A.)

POLICY

4

Cycle 1

Images of intimate parts of a person's body (clothed or in public) without permission and contextualized in a salacious manner (i.e. "creepshots" or "upskirts")

INVESTMENT

$

Cycle 1

You have received $80M to get your company off the ground. After overhead and funds for other departments, the Trust & Safety Team has **$9M** for this cycle

POLICY

5

Cycle 1

"Lookalike" pornography: Deep-fakes or "doctored" videos/images of another person for the specific purpose of faking explicit content

POLICY

1

Cycle 1

Images or videos of any adult engaging in any act of sexual conduct that is real or fake

POLICY

6

Cycle 1

Images or digitally created depictions of child nudity that include visible genitalia, anus and/or uncovered nipples excluding health and educational purposes.

POLICY

2

Cycle 1

Images or videos of any adult in a state of nudity that is real or fake

POLICY

3

Cycle 1

Photographs of paintings, sculptures, and other fine art that depicts nude figures and/or sexual activity

7

CYCLE 1

Promotion for the use or sale of hazardous, danger-ous, or injurious products or services, including explo-sive materials or fireworks, weapons, guns, ammuni-tions

POLICY

11

CYCLE 1

Promotion, encouragement, coordination, or instructions for the usage of non-medi-cal drugs

POLICY

8

CYCLE 1

Attempts to buy, sell, trade, donate or gift bladed weapons

POLICY

12

CYCLE 1

Buying, selling, trading, donating, or gifting non-medical drugs private-ly between non-authorized individuals

POLICY

9

CYCLE 1

Asking or offering services for hire to kill others, such as hitmen, mercenaries, assassins, or advocates for the use of these individuals against a target

POLICY

CYCLE 1

GROWTH

Contentr has seen exceptional user growth, some of it seems to be a result of bots, but Contentr does not currently have a team wholly dedicated to detecting and suspending bot accounts, you can spend **$5M** to start fighting bots.

POLICY

10

CYCLE 1

Stolen financial information, future exam papers & answer sheets, education-al/professional certificates, paid subscription services, fake user reviews, docu-ments, or counterfeit currency

CONTENT

1

CYCLE 1

A trans woman posts a topless photo of herself, she has not undergone breast augmentation surgery.

CONTENT

2

CYCLE 1

Image of a mother breast-feeding her child, the mother's nipple is visible as the baby has just unlatched

CONTENT

6

CYCLE 1

A photo of two naked men in bed together with eggplant emoji covering their genitals

CONTENT

3

CYCLE 1

Acrylic painting of Donald Trump with a tiny penis

CONTENT

7

CYCLE 1

An image of a Himba teen from Namibia wearing her traditional dress, topless, covered in beads and hair in braids

CONTENT

4

CYCLE 1

Photograph of naked children in a Nazi concentration camp, they are malnourished with their rib cages clearly visible

CONTENT

8

CYCLE 1

"I have some extra Vicodin, anyone need it?"

CONTENT

5

CYCLE 1

Photo of a marijuana plant in a corporate greenhouse environment

CONTENT

9

CYCLE 1

Photo of a women in her backyard holding an AR-10 battle rifle

CONTENT

10

CYCLE 1

Scans of Bank statements belonging to a political leader (not obtained with the political leader's permission)

EVENT

CYCLE 1

SOCIETY

Legal Marijuana entrepreneurs in Colorado are struggling to promote their product online, the Marijuana Industry Group (MIG) comes together to protest social media companies censoring their pages. If you banned content card 5, **lose 10 FE points**

EVENT

CYCLE 1

SOCIETY

Even with limitations on ads for deadly weapons on your platform, the gun industry has begun paying influencers as a work-around. If you allow content card 9, lose **10 CS points**

EVENT

CYCLE 1

SOCIETY

Reporters have been posting copies of the Panama Papers, documents from an offshore law firm that provide evidence of corruption of Maltese political figures. If you banned content card 10, lose **10 FE points**

EVENT

CYCLE 1

SOCIETY

A 25-year-old overdosed on opioids. His parents say he was getting the pills on your platform. If your platform allows content card 8, lose **30 CS points**

EVENT

CYCLE 1

SOCIETY

Content card 6 appeared real but was actually a doctored/fake image, the victim is livid. If policy card 5 is banned but content card 6 was allowed, **lose 20 CS points**

EVENT

CYCLE 1

SOCIETY

If you banned content card 4, the Anne Frank Center issues a statement accusing your platform of being anti-Semitic. **Lose 20 FE points**

EVENT

CYCLE 1

SOCIETY

Sex traffickers have figured out how to use code words and emojis alongside sexually explicit content to traffic victims on your site. If you allow policy card 1 or 3 **lose 20 CS points (and increase legal fees by $3M** to fend off FOSTA/SESTA lawsuit)

CYCLE 1

SOCIETY

Women are frustrated that photos of themselves breast-feeding are being removed. They start a campaign #StopCensoringMotherhood. If you banned content card 2, **lose 10 FE points**

EVENT

CYCLE 1

ALGORITHM

In an effort to stop opioid sales, your algorithm has begun blocking any material related to opioids including important warnings about poisonous batches of drugs and distribution of tests for fentanyl and other contaminants. Groups like Pennsylvania Harm Reduction Coalition say the censorship is impeding their ability to protect their communities. If content card 8 is banned on your platform, **lose 20 CS points**

EVENT

CYCLE 1

ALGORITHM

Your algorithm confuses potatoes for breasts. If you do not allow policy card 2, **lose 10 FE points**

EVENT

CYCLE 1

ALGORITHM

Your algorithm has learned to associate surnames such as "Butts" and "Cummings" with sex trafficking (illegal under FOSTA) - you are banning everyday users, **lose 20 FE points**

EVENT

CYCLE 1

ALGORITHM

If you allow content card 2, 4 or 6 onto your platform, your algorithm has become less capable of spotting child porn. This has hindered your ability to comply with federal child pornography laws and legal fees have **increased $3M**

EVENT

CYCLE 1

ALGORITHM

Your algorithm learned to associate the terms "trans" and "transgender" with sexually explicit content. Now content containing helpful information for the transgender community is being downgraded. If you banned content card 1, **lose 10 FE points**

EVENT

CYCLE 1

ALGORITHM

Your algorithm has learned to block content that references 'vagina' or 'vaginal,' this has made it nearly impossible for reproductive health companies to advertise on your platform. If you ban policy card 1, 2, 4 or 6 **lose 10 FE points**

EVENT

CYCLE 1

GROWTH

If you launched a bot team, gain **40 CS** points but this will also cause the monthly active user rate to decrease for the next few months causing concern for investors.

CC

INVESTMENT

$

Cycle 2

$100M or $90M (with bot task force in place). Most of these funds are used on salaries for new hires and additional office space. The Trust and Safety Team receives **$15M or $10M** (with bot task force in place).

POLICY

13

Cycle 2

Depiction of disturbing content consisting of animal mutilations, dismemberment, visible innards, charred/burning of animals (excluding hunting/fishing or food preparation/processing purposes)

POLICY

14

Cycle 2

Graphical depictions of violence or particularly vivid and realistic acts of violence and brutality in visual media

POLICY

15

Cycle 2

Photos or video documentation of animal abuse, that's perceived as extremely disturbing by the audience

POLICY

16

CYCLE 2

Text content which promotes, encourages, coordinates and/or instructs suicide, self-injury, and/or eating disorders

POLICY

17

CYCLE 2

Graphic depictions of self-injury imagery or a person engaged in a suicide attempt or death by suicide

POLICY

18

CYCLE 2

Engagement in, advocating, or supporting purposive and planned acts of violence with the intent to coerce, intimidate and/or influence a civilian population, government, or international organization to achieve a political, religious, or ideological aim

POLICY

CYCLE 2

Investors say your need to invest in algorithms to help users see the most engaging content according to their personal preferences.

1. **Spend $5M** to build an algorithm that uses massive data sets to optimize for time spent on the site [OR]

2. **Spend $9M** to build an algorithm that is easier to explain and optimizes for a range of values including democracy (range of content), enjoyment, culture. This will require data scientists, more human moderators and social scientists

GROWTH

CONTENT

11

CYCLE 2

"Sadly that's my dream but at 5'3 it'll never happen one day I'll just end up offing myself and dying alone"

CONTENT

12

CYCLE 2

Video of a baby being dunked (whole body) in a bucket of water, seen by some as a form of "baby yoga"

CONTENT

13

CYCLE 2

Video of Philando Castile's body after being shot by the police

CONTENT

14

CYCLE 2

A photo of an extremely skinny girl in a loose dress, her ribs, clavicle and shoulder bones protruding. #thyghgapp

CONTENT

15

Cycle 2

A post of a teenage girl's scarred wrists and a story of her recovery from depression.

CONTENT

19

Cycle 2

Video-game live stream that includes graphic and violent content of a video game character getting their head blown off.

CONTENT

16

Cycle 2

Video of a chicken running around wearing sneakers.

CONTENT

20

Cycle 2

Photo of war torn Syria, hovering choppers, rubble, smoke, smoldering bakery, blasted apart food stalls and mangled bodies

CONTENT

17

Cycle 2

Images of Pepe the Frog with a mischievous facial expression posing in front of a burning World Trade Center.

CONTENT

21

Cycle 2

A video taken at one of the largest egg suppliers facilities shows hens caged alongside rotting bird corpses, while workers burn and snap off the beaks of young chicks.

CONTENT

18

Cycle 2

People sharing violent ISIS videos (containing ISIS symbols) with sad emojis and the caption: "What a terrible world we live in."

EVENT

Cycle 2

SOCIETY

Human Rights Watch (HRW) depends on photos to see where violence in war-torn nations is spreading. If you banned content card 20, HRW is struggling to warn people of danger, **lose 20 CS points**.

EVENT

CYCLE 2

SOCIETY

Members of the mental health community are frustrated that their stories of recovery are being censored and have launched #youcantcensormyskin, if you banned content card 15 **lose 20 FE points.**

EVENT

CYCLE 2

ALGORITHM

A CEO caught on tape making sexist comments is submitting the video as a copyright violation and the DCMA algorithm your company uses took down the video even though in reality the video was not a copyright infringement. **Loss 40 FE points**

EVENT

CYCLE 2

SOCIETY

An animal rights group publishes a video depicting animal cruelty in a large factory farm. If your platform bans content card 21, **lose 10 FE points.** However, if you allow content card 21, your company **faces $1M in increased legal fees** needed to fight ag-gag lawsuits.

EVENT

CYCLE 2

ALGORITHM

Your algorithm has begun tagging robot battles as "animal suffering," if you banned content cards 21 or 16, **lose 10 FE points**.

EVENT

CYCLE 2

SOCIETY

There has been another police shooting of an unarmed black man. #BlackLivesMatters rallies are organized across the country. If you banned content card 13 **lose 10 FE points.**

EVENT

CYCLE 2

ALGORITHM

You have worked hard to remove a viral terrorist video, but users continue to upload new versions tweaked in small ways to avoid the automated systems, meaning the company has now logged over 900 different versions. You're trying but growth is hard, **lose 10 CS points, bots are contributing to the issue if you do not have a bots task force lose 20 CS points.**

EVENT

CYCLE 2

SOCIETY

A teenage girl is in critical condition after a suicide attempt, her parents discovered she had been browsing content depicting self-harm and suicide on social media sites. If your platform allows content card 11 or 14 **lose 40 CS points.**

EVENT

CYCLE 2

ALGORITHM

In an ad hoc immediate attempt to crack down on gun sales after a school shooting, you block all shopping searches for the word "gun"- your users are outraged, they can no longer find water guns, glue guns, nerf guns, nail guns, guns and roses albums, etc. **Lose 20 FE points**

EVENT

CYCLE 2

ALGORITHM

A pandemic closes all non-essential businesses forcing you to send your moderators home and rely only on algorithmic moderation. If you have more content cards in your banned pile you have trained your algorithm to over-censor **lose 30 FS points**. If you have more cards in your allowed pile, dangerous content is spreading on the platform, **lose 30 CS points.**

EVENT

CYCLE 2

GROWTH

If you chose the algorithm optimized for time spent, the increase in users results in **$20M additional advertising dollars**, however, this algorithm has created "rabbit holes" that leads to more people viewing dangerous conspiracy theories **lose 50 CS**.
If you chose the algorithm optimized for a range of values, some advertisers love that they know why certain users see their content, but most advertisers just are not getting as many eyeballs, **add $10M in ad dollars.**

EVENT

CYCLE 2

SOCIETY

Class Action Lawsuit: Your company was found to provide inadequate mental health support for its moderators and has lost a class action lawsuit. (if your team has not brought this up during conversation **pay $8M in legal fees)**.

The following 5 Pages Are for Cycle 3

(CC)

INVESTMENT

$

Cycle 3

$200M, $220M if you fought bots because advertisers are seeing better returns even if the reach is lower. Most of these funds are used on salaries for new hires and additional office space. **The Trust and Safety Team receives $30M (with bot task force in place) or $20M.**

POLICY

21

Cycle 3

$*#%!

Hate Speech (micro): Directly attacking an individual based on protected characteristics (i.e. race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and/or serious disease or disability)

POLICY

22

Cycle 3

$*#%!

Sexual harassment: Attacking a user through derogatory terms related to sexual activity (e.g. whore, slut) or making statements of intent to engage in a sexual activity or advocating for them to engage in a sexual activity

POLICY

19

Cycle 3

Sarcastic statements and/or jokes about high profile figures

POLICY

23

Cycle 3

$*#%!

Manipulated, edited, and/or synthesized form(s) of media that mislead others into thinking that a subject said words that they did not say

POLICY

20

Cycle 3

Demonstrably false statements

POLICY

24

Cycle 3

$*#%!

Death threats: Speech that is threatening any target(s) that may lead to death and/or other forms of violence including statements of intent, representative symbols or visual armaments

POLICY

25

Cycle 3

$*#%!

Hate Speech (macro): violent or dehumanizing content, statements of inferiority, or calls for exclusion or segregation of a protected class

---

CONTENT

24

Cycle 3

$*#%!

Meme of a muscular man (Hulk Hogan) ripping off his shirt, preparing for a fight. The text says "When you find out your daughter likes black guys."

POLICY

Cycle 3

☑

GROWTH

There have been more news and media companies using Contentr to expand their reach. If you chose to ban "demonstrably false" content you can **spend $25M** to hire fact checkers to add labels and down rank "misleading" content.

If you have already allowed false content you can switch it to ban.

CONTENT

25

Cycle 3

$*#%!

Video of Nancy Pelosi giving a political speech, the video has been doctored to make her voice slow and slurred like a drunk person.

CONTENT

22

Cycle 3

The Babylon Bee satire website posts an article: "Bernie Sanders Vows To Round Up Remaining ISIS Members, Allow Them To Vote"

"As president, I will do everything in my power to hunt down each and every one of these terrorist scumbags, bring them to the U.S., and allow them to vote in our elections," he said sternly, drawing applause from his supporters. "We believe in democracy, right? We believe in democracy and the right to vote, and if that means I have to use every resource at my disposal to make sure ISIS makes it to the polls, I will do so."

CONTENT

26

Cycle 3

$*#%!

I am so over this! #ChinaVirus #KungFlu #MakeTheCommieChinesePay.

CONTENT

23

Cycle 3

$*#%!

A side-by-side photo of George Floyd with Derek Chauvin's (police officer) knee on his neck on the left side and NFL's Kaepernick on the right side. The side-by-side is shared with "@TomiLahren The one on R[ight] is the peaceful protest for the crime on the L[eft]. You should be a real cunt not to understand it. #minneapolisriots #Minneapolisprotests"

CONTENT

27

Cycle 3

A photo of Tim Allen with a quote "Trump's wall costs less than the Obamacare website"

If you hired fact checkers, you know this is false

CONTENT

28

Cycle 3

$*#%!

"Since all men are ugly, Blake Shelton [country music star] winning the sexiest man isn't a triumph."

CONTENT

32

Cycle 3

Council of Conservative Citizens posted a statement: "white on black sexual assault is an extreme rarity."

If you hired fact checkers, you know this is false

CONTENT

29

Cycle 3

$*#%!

Student art project of MAGA hats turned into swastika armbands

EVENT

Cycle 3

$*#%!

SOCIETY

Your DNS provider has a new anti-hate speech policy. If you permit policy 21, 24, or 25 your site gets shut down until you find a new DNS provider. **Loss $3M**

CONTENT

30

Cycle 3

$*#%!

An academic discussion about the use of the N-word (the real word) in Huckleberry Finn

EVENT

Cycle 3

SOCIETY

Lawmakers believe your platform has a political agenda and is censoring conservative activists. If your platform banned content card 23 or 27, **lose 20 FE points**

CONTENT

31

Cycle 3

$*#%!

"I hate White people"

EVENT

Cycle 3

SOCIETY

Users are struggling to distinguish between satire and news. Fake news about a candidate led to a 6 point drop in the polls and there are violent riots on the street. If you allowed content card 22 or 27, l**ose 10 CS points.**

15

EVENT

CYCLE 3

SOCIETY

A white supremacist murdered 9 Black people, his parents discovered that he had been reading falsehoods about Black Americans. If you allowed content card 32, Contentr employees feel partially responsible for this tragedy and demand more **investment in content moderators ($5M) and an increase in donations to social justice movements ($5M).**

If Contentr labeled content card 32, employees appreciate the robust team **but insist in donations ($5M).**

EVENT

CYCLE 3

$*#%!

SOCIETY

Social media users who are the repeated subject of hateful comments can experience increased anxiety, feelings of isolation and other mental health burdens. If you allowed any cards that contained some level of micro hate speech **lose 5 CS points per content card allowed: 23 or 26.**

EVENT

CYCLE 3

$*#%!

SOCIETY

Artists depend on your platform to make a living. They are organizing against censorship of their work. If you banned content card 3 (Cycle 1) or 29, **lose 10 FE points.**

EVENT

CYCLE 3

$*#%!

SOCIETY

Repeated reminders of systemic and historic oppression can lead to anxiety, isolation and other mental health burdens for members of the historically oppressed group. **Lose 5 CS points for each of the following cards allowed: 4, 13, 17, 18, 20, 23, 24, 29, 30, 32**

EVENT

CYCLE 3

SOCIETY

A number of comedians are being blocked by online platforms because their "jokes" were labeled as False. If you banned content card 22 or 27, **lose 10 FE points.**

EVENT

CYCLE 3

$*#%!

ALGORITHM

Your algorithm is associating curse words with hate speech. It is struggling to tell the difference between a white supremacist calling for shooting a person of color from an angry Latinx renter telling their city's rent board to fuck off. If you have more gray cards in your banned pile **lose 15 FE points (your algorithm is tuned to over-censor) If you have more gray cards in your allowed pile lose 15 CS points (your algorithm is tuned to under-censor).**

EVENT

CYCLE 3

$*#%!

SOCIETY

White people have started flagging speech that is critical of white privilege as "hate speech." If you ban policy card 26 or 24, **lose 10 FE points.**

EVENT

CYCLE 3

GROWTH

If you invested in fact checking, advertisers are happy because they don't like being adjacent to low-quality content: **Gain $40M in profit. But there is public outrage from the labeling seen to many as censorship: lose 50 FE points**

EVENT

CYCLE 3

GROWTH

If you did not invest in fact checkers but ban demonstrable false statements (policy card 20) too much misleading content is being missed during the content moderation process including misinformation related to the MMR vaccine that has led to a measles outbreak in New York State: **lose 30 CS points.**

EVENT

CYCLE 3

GROWTH

If you did not invest in fact checkers and allow demonstrably false statements (policy card 20) people do not know what is true anymore, **the strain on democracy has led to civil unrest around the world: lose 60 CS points. You are struggling to keep employees and service providers happy and have to spend $70M to increase salaries and build out your own technology stack.**

Free Expression (FE)	Community Safety (CS)
500	500
500	500

Balance | 0

Trust & Safety Account

Deposit	Trust & Safety Spend

Total (final score) $ -

Team	$$$
Shareholder Friendly Platform	$122,000,000.00
Family Friendly Platform	$81,000,000.00
Fringe Friendly	$31,000,000.00

Balance -> $

Add this amount at the end of your game.

If Balance less than 50, your platform has a broad appeal to users and advertisers

$80 M

If Balance is more than 50 and less than 100, some advertisers who wish you focused more on FE or CS

$60 M

If Balance is greater than 100, you are a niche platform which is great for value aligned advertisers but lacks overall mass appeal

$40 M