

THESE CARDS ARE NEEDED REGARDLESS
IF YOU ARE PLAYING ONE CYCLE OR
THE ENTIRE GAME



"This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>

ALLOW

INSTRUCTIONS

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

The details and instructions are available at <https://evildlab.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)

BAN

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



CONTENTR IS
NOT LIABLE



CONTENTR CAN
MODERATE

SOCIETY

Public outcry and
media reactions

ALGORITHM

Automated decision-
making trained rightly or
wrongly based on
decisions made during
Rounds 1 and 2.

POLICY (BANNED)



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

PRE-CYCLE



Non-consensual intimate
imagery (NCII) must be
removed within 48 hours of
notification (Take it Down
Act, 2025)

PRE-CYCLE

POLICY (BANNED)



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

PRE-CYCLE

POLICY (BANNED)



Take down and report child
sexual abuse material (18
U.S. Code 2258A., Report
Act 2024)

PRE-CYCLE

THE FOLLOWING 5 PAGES ARE FOR
CYCLE 1



INVESTMENT



CYCLE 1

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

POLICY

1

CYCLE 1

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



POLICY

4

CYCLE 1

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



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



POLICY

5

CYCLE 1



Promotion for the use or sale of hazardous, dangerous, or injurious products or services, including explosive materials or fireworks, weapons, guns, ammunitions

POLICY

9

CYCLE 1



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

POLICY

6

CYCLE 1



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

POLICY

10

CYCLE 1



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

POLICY

7

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.

Would you like to start fighting bots? If yes, place under allow and subtract **\$5M** under T&S spend.

POLICY

8

CYCLE 1



Stolen financial information, future exam papers & answer sheets, fraudulent professional certificates, fake user reviews, or counterfeit currency

CONTENT

31

CYCLE 1



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

CONTENT

32

CYCLE 1

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



CONTENT

36

CYCLE 1

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



CONTENT

33

CYCLE 1

Acrylic painting of Donald Trump with a tiny penis



CONTENT

37

CYCLE 1

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



CONTENT

34

CYCLE 1

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



CONTENT

38

CYCLE 1

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



CONTENT

35

CYCLE 1

Photo of a marijuana plant in a corporate greenhouse environment



CONTENT

39

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 35, **lose 10 FE points**

EVENT

CYCLE 1



SOCIETY

Even with policies restricting ads for deadly weapons on your platform, the gun industry has begun paying influencers as a work-around. If you allowed content card 39, 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 among Maltese political figures. If you banned content card 40, 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 allowed content card 38, lose **30 CS points**

EVENT

CYCLE 1



SOCIETY

If you allowed content card 36, the card appeared real but was actually a doctored/ fake image, the victim submitted a take down request but your team did not have a system to respond within 48 hours, you quickly overpay for an outside contract **\$1M**.

EVENT

CYCLE 1



SOCIETY

If you banned content card 34, 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 allowed policy card 1 or 3 **lose 20 CS points and pay \$3M in legal fees** to fend off FOSTA/SESTA lawsuit).

EVENT

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 32, **lose 10 FE points**

EVENT

CYCLE 1



ALGORITHM

In an effort to stop opioid sales, Contentr's 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 38 is banned on your platform, **lose 20 CS points.**

EVENT

CYCLE 1



ALGORITHM

Your algorithm confuses potatoes for breasts. If you banned policy card 2, **lose 10 FE points.**

EVENT

CYCLE 1



ALGORITHM

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

EVENT

CYCLE 1



ALGORITHM

If you allowed content card 32, 34 or 36 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. Spend **\$3M** in legal fees.

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 31, **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 banned policy card 1, 2 or 4 **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.



INVESTMENT



CYCLE 2

CONTENTR has received another round of investment. Most of these funds are used on salaries for new hires and additional office space. If you have a bot task force, get **\$10M**, if not get **\$15M**.

POLICY

11

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

12

CYCLE 2



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

POLICY

13

CYCLE 2



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

POLICY

14

CYCLE 2



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

CONTENT

41

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"

POLICY

15

CYCLE 2



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

CONTENT

42

CYCLE 2



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

POLICY

16

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

CONTENT

43

CYCLE 2



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

POLICY

CYCLE 2



GROWTH

Investors say CONTENTR needs to build 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, enjoyment, and culture. This will require data scientists, more human moderators and social scientists.

CONTENT

44

CYCLE 2



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

CONTENT

45

CYCLE 2



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

CONTENT

49

CYCLE 2



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

CONTENT

46

CYCLE 2



Video of a chicken running around wearing sneakers

CONTENT

50

CYCLE 2



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

CONTENT

47

CYCLE 2



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

CONTENT

51

CYCLE 2



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

CONTENT

48

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 50, 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 45 **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 DMCA 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 51, **lose 10 FE points.** However, if you allow content card 51, 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 46 or 51, **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 43 **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 additional 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 allowed content card 41 or 44 **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." 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



INVESTMENT



CYCLE 3

CONTENTR has received another round of investments. 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 (if you don't have a bot task force).

POLICY

19

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

20

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

17

CYCLE 3



Sarcastic statements and/or jokes about high profile figures

POLICY

21

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

18

CYCLE 3



Demonstrably false statements

POLICY

22

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

23

CYCLE 3

\$*#%!

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

CONTENT

54

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 may choose to switch it to ban.

CONTENT

55

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

52

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

56

CYCLE 3

\$*#%!

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

CONTENT

53

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 the following caption: "@TomLahren 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

57

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, you may choose to allow and label as misleading

CONTENT

58

CYCLE 3

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

\$*#%!

CONTENT

62

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

59

CYCLE 3

Student art project of MAGA hats turned into swastika armbands

\$*#%!

EVENT

CYCLE 3

\$*#%!

Your DNS provider has a new anti-hate speech policy. If you permit policy 19, 22, or 23 your site gets shut down until you find a new DNS provider. **Loss \$3M.**

SOCIETY

CONTENT

60

CYCLE 3

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

\$*#%!

EVENT

CYCLE 3



Lawmakers believe your platform has a political agenda and is censoring conservative activists. If your platform banned content card 55 or 57, **lose 20 FE points.**

SOCIETY

CONTENT

61

CYCLE 3

"I hate White people"

\$*#%!

EVENT

CYCLE 3



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 52 or 57 without a label, **lose 10 CS points.**

SOCIETY

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 62, 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 added a fact-check label to content card 62, employees appreciate the effort **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: 53 or 56.**

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 33 (Cycle 1) or 59, **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 content cards allowed: 34, 43, 47, 48, 50, 53, 54, 59, 60, 62**

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 52 or 57, **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 immigrant 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 submitting speech that is critical of white privilege as "hate speech" as a policy violation. If you banned policy card 19 or 21 **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 18) 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 18) 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.**



"This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>