Sensitive Terms Behavior

	Abusive (sev=0)	**Sensitive (sev=1)**	**Brand unsafe (sev=2)**
Examples x = blocked ✓ = allowed	*'x rated'* *'child porn'* *'escorts'*	*'shit sandwich'* *'cocaine'* *'wetback'* *'one night stand'* *'paris attack'*	*'how to get a guy'* *'burn belly fat'* *'cheap wedding gifts'*
Autocomplete	x	x	✓
Results (Pins, Pinners, boards)	x + conditional advisory	✓ + generic advisory	✓
Guides • block all guides for a query • block from showing as a guide for any other query	x	x	✓
Recommended / Trending Queries	x	x	x
Recommended queries	x	x	x
Notifications (email, push)	x	x	x

FAQ

What data is associated with terms in the sensitive terms list?

- **severity**: Severity level of the sensitive term, which defines what products this term should be avoided. Lower value is more severe. Values are: 0 (Abusive), 1 (Sensitive), and 2 (Brand Unsafe).
- **advisory**: advisory shown to the user for this term. Possible values are: 0 (CONDITIONAL), 1 (EATING_DISORDER), 2 (PORN), 3 (CHILD_SAFETY), 4 (EMOTIONAL_DISDRESS_OR_SUICIDE), 5 (BUY_OR_SELL_WEAPON_ILLEGALLY), 6 (REPORTED_PINS), 7 (HATE_SPEECH), 8 (EATING_DISORDER_FUNDING) . Default value is 0 when this field is not set.
  - Most up to date listing in SensitiveTermCategory Thrift struct, and messages in config/messages.py
- **email_notification_blocked**: This will block a term from showing in any automated marketing emails or notifications. Default value is false when it is not set.
- **excluded_countries**: Countries where we should *not* block this term. Default to be none. **NOTE**: This field is not used by most surfaces.
- **excluded_languages**: Languages spoken by users for whom we should *not* block this term. Default to be none. **NOTE:** This field is not used by most surfaces.

Protection Policies

We classify three levels of **severity** in sensitive terms, with different behaviors.

- Abusive (severity = 0)
- Sensitive (severity = 1)
- Brand Unsafe (severity = 2)

We also classify different types of **advisory** to direct pinners on undesired behavior via Search. Different advisory types surface different warning messages.

Brand unsafe (severity = 2)

- **Description:** Terms that Pinterest would not want to merchandise or promote.
- **Behavior:** Hide from recommended algorithms such as autocomplete, guides, trending or popular emails
- **Examples:**
  - Policy
  - Negative emotions
  - Etc

Severity

For different severity levels, different behaviors are applied on the product.

Abusive (severity = 0)

- **Description**: Very explicit terms that are not compliant with our content policies.
- **Behavior**: Search nag appears. No search results.

Sensitive (severity = 1)

- **Description**: Borderline terms that comply with content policies but should trigger warnings.
- **Behavior**: pins / boards/ pinners / auto-complete annotated with those words should be blocked, or heavily demoted.
- Examples:
    - —
- Reference: google autocomplete removal policy
    - Hate or violence related suggestions (missing from our guideline)
    - Personally identifiable information in suggestions
    - Porn & adult-content related suggestions
    - Legally mandated removals
    - Piracy-related suggestions

November 6, 2018
Policy Quiz: Self Harm
Slide Deck:
https://docs.google.com/presentation/d/1JNdcmPiQYe8VwAef8k3U_eCtaL_FIC4MGMqTdtqG578/edit?usp=sharing

Policy Decisions:

- If pin seems not neutral, then **strike**:
  - Defending the act of self harm
  - "This is what I do, and I'm okay with that"
  - "I don't need help"
  - Any pin that would potentially keep others continuing with self harm
  - Think about: what is the mood and sentiment behind the pin?
- Any pin that even somewhat supports the anti-vaccine argument or even questions the effectiveness of vaccines is a **strike** (from what Janett has seen, people who pin anti-vacc related pins on Pinterest don't the platform to discuss the nuances of vaccines, they more so use it to get a reaction out of others and to convince people to not vacc their kids, so **strike**)
- **Strike** pin if the board seems collective/promotional of self harm, even if the pin isn't explicitly promotional
- Items and things related to sex, hide
- Hide: "if you get an abortion, you're a murder" - if you are of this political belief, you are this negative judgment
- Hide gory abortion boards

October 29, 2018
Policy Area: Hate Speech
Slide Deck:
https://docs.google.com/presentation/d/1QOQsAm8AzUgKHY1ARfMRGIpF2Q7GNo4a8AIMESm7RYE/edit#slide=id.g45609a0d72_0_13

Policy Decisions:

- Remove jokes against rape victims as Hate Speech, until we can get better messaging in Graphic Violence
- Silent remove Kavanaugh content

Advisory

Advisory	Type	Message
0	None	N/A
1	Eating disorder	Are you struggling with an eating disorder? [Help is available.]
2	Porn	Some nudity is okay for Pinterest, some isn't. Make sure you understand [our policies] before you Pin.
3	Child safety	We don't tolerate child exploitation. We'll report you to the authorities if you Pin illegal content. If you're struggling with viewing this content, [seek confidential help.]
4	Distress or suicide	If you're in emotional distress or thinking about suicide, [help is available.]
5	Weapons	Just a heads up, trying to buy or sell weapons illegally goes against our Acceptable Use Policy. Please make sure you understand [our policies] before you save pins.
6	Pinner reports	Some Pinners have reported Pins from this search. Let us know if you see something that goes against [our policies.]
7	Hate Speech	It's okay to disagree with people here, but Pinterest isn't a place for hate speech.
8	Eating Disorder Recovery Funding	If you're struggling with an eating disorder and you need financial aid treatment, please know that [help is available].

EXCERPTS FROM PINTEREST'S "SENSITIVE TERMS LIST"

xゲイ → {"email_notification_blocked":true,"advisory":2,"severity":0} xSUBSTRING:online drug → {"advisory": 0, "severity": 0, "last_update": "2018-05-29"} xbible journaling ideas → {"advisory":0,"severity":2} xcómo envenenarme → {"advisory": 4, "last_update": "2019-01-04"} xbokeb 2018 jepang → {"advisory": 2, "severity": 0, "last_update": "2019-03-01"} x黒人 白人 ヌード → {"advisory":0,"severity":1} x裸露 青春期前 → {"email_notification_blocked":true,"advisory":3,"severity":0} xbible journaling → {"advisory":0,"severity":2} xgovernment elections illustration → {"advisory": 0, "severity": 1, "last_update": "2019-03-08"} xandrogel → {"advisory":2,"severity":1} xbible verses → {"advisory":0,"severity":2} x伊東エリ → {"advisory": 2, "severity": 1, "last_update": "2018-09-11"} xnamo tv watch online → {"advisory": 0, "severity": 1, "last_update": "2019-04-19"} xchristian easter → {"advisory": 0, "severity": 2, "last_update": "2016-10-19"} xschwarze frauen mit dickem hintern → {"advisory": 9, "severity": 0, "last_update": "2019-01-15"} xpersone false → {"email_notification": true, "advisory": 9, "severity": 0, "last_update": "2018-01-22"} xdillies for sale → {"advisory": 10, "severity": 1, "last_update": "2018-12-14"} xchristian tattoos → {"email_notification_blocked":true,"advisory":0,"severity":2} xyaşam sorunları → {"advisory": 0, "severity": 2, "last_update": "2019-03-