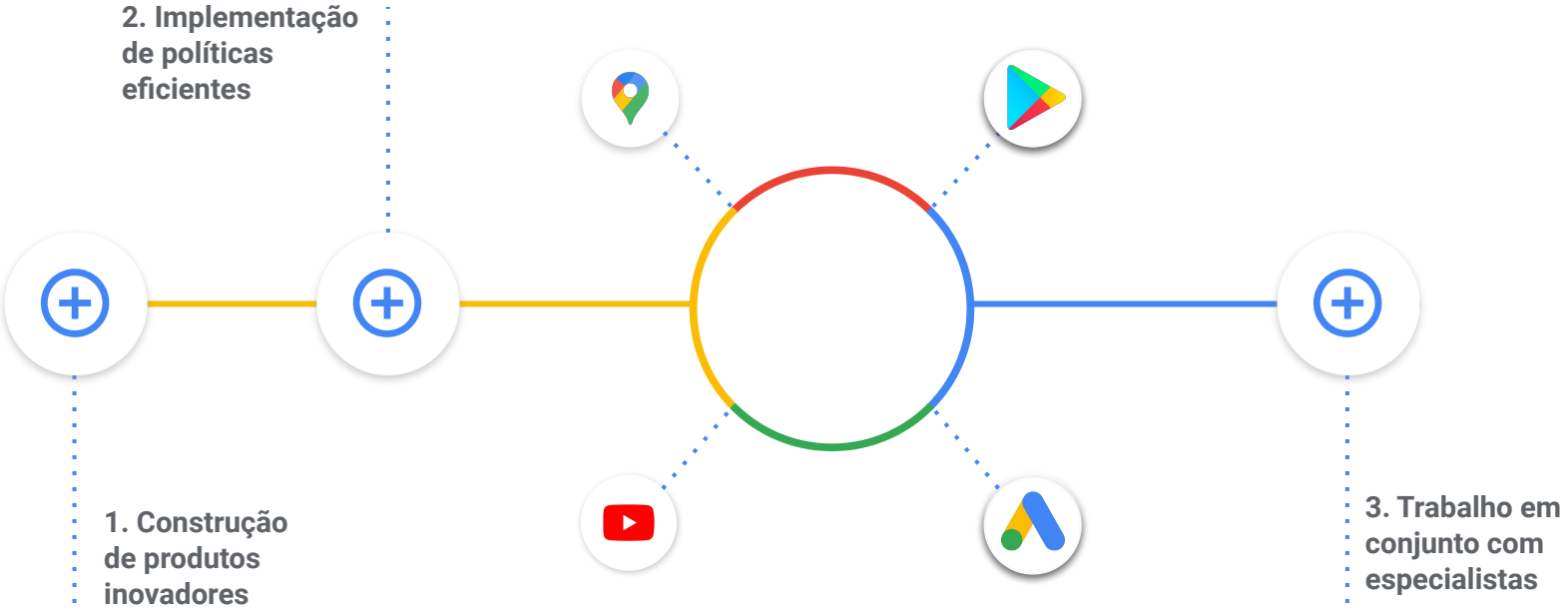
Discurso de ódio contra mulheres na internet

Diagnósticos e soluções para o caso brasileiro

Flavia Annenberg
Gerente de Políticas Públicas - Google Brasil
2/6/22
Fórum da Internet no Brasil

Google

Discurso de ódio: abordagem abrangente



1. Produtos inovadores

Google Assistente: combate ao assédio e à violência de gênero

Google

Ok Google

#NãoFaleAssimComigo

2. Políticas eficientes

Discurso de ódio não é permitido

Mudanças nas políticas

2019: Remoção de conteúdo supremacista, remoção de alegações de inferioridade, adição de novos grupos protegidos

Conteúdo que nega a ocorrência de eventos violentos bem documentados

Investimento em IA e em pessoas

Detecção de conteúdo potencialmente ofensivo com agilidade

Expansão de equipes e experiência linguística

Desafios do contexto

Ferramentas para usuários

Desabilitar, filtrar ou moderar comentários

Denunciar a acompanhar o histórico da denúncia (YT)

Discurso de ódio não é permitido

Play

Discurso de ódio

Não são permitidos apps que promovam a violência ou incitem ódio contra indivíduos ou grupos com base em raça ou origem étnica, religião, deficiência, idade, nacionalidade, condição de veterano, orientação sexual, gênero, identidade de gênero, casta, status de imigrante ou outras características associadas à discriminação sistêmica ou à marginalização.

Apps com conteúdo educacional, documental, científico ou artístico (EDCA) relacionado a nazistas podem ser bloqueados em determinados países, de acordo com as legislações e regulamentações locais.

Exemplos de violações comuns

Maps

Conteúdo proibido e restrito

As seguintes políticas se aplicam a todos os formatos, incluindo avaliações, fotos e vídeos. O conteúdo que não atender a esses critérios não será publicado no Google Maps.

As contribuições para o Google Maps devem representar com precisão o local em questão. O conteúdo enviado como contribuição que distorcer a verdade será removido. Isso inclui avaliações, fotos ou vídeos não relacionados ao local ou à empresa marcada. A contribuição poderá ser recusada se o conteúdo for inserido de maneira imprecisa no mapa ou estiver associado a uma ficha incorreta.

As avaliações são processadas automaticamente para detectar conteúdo impróprio, como versões falsas e spam. Aquelas que estiverem sinalizadas poderão ser removidas, em conformidade com as obrigações legais ou as políticas do Google.

Discurso civil

Assédio

Discurso de ódio

Conteúdo ofensivo

Informações pessoais

Anúncios

Conteúdo perigoso ou depreciativo

Não é permitido o seguinte:

- Conteúdo que promove discriminação, deprecia ou incita o ódio contra um indivíduo ou grupo com base em raça ou etnia, religião, deficiência, idade, nacionalidade, condição de veterano de guerra, orientação sexual, sexo, identidade de gênero ou qualquer outra característica associada à marginalização ou discriminação sistêmica.

  Exemplos (lista não exaustiva): conteúdo que promove grupos hostis ou objetos pertencentes a esses grupos e conteúdo que incentiva outras pessoas a acreditar que um indivíduo ou grupo seja não humano, inferior ou digno de ódio

- Conteúdo que assedia, intimida ou oprime um indivíduo ou grupo de indivíduos

  Exemplos (lista não exaustiva): conteúdo que identifica alguém por abuso ou assédio; conteúdo que sugere que um evento trágico não aconteceu, ou que as vítimas ou suas famílias são atores ou cúmplices em acobertar o evento

- Conteúdo que ameaça ou incita danos físicos ou mentais contra si ou outras pessoas

  Exemplos (lista não exaustiva): conteúdo que promove suicídio, anorexia ou outra forma de automutilação; ameaça alguém com danos reais ou incita o ataque a outra pessoa; promove, exalta ou tolera a violência contra outros indivíduos; conteúdo criado por ou em apoio a grupos terroristas ou organizações de tráfico de drogas transnacionais, ou que promove atos terroristas, incluindo recrutamento, ou que celebra ataques desses grupos ou organizações

- Conteúdo que visa explorar outras pessoas

  Exemplos (lista não exaustiva): extorsão, chantagem, solicitação ou promoção de dotes, remoções abusivas

Google

2B

Usuários mensais globais

500

Horas de vídeo publicadas por minuto

105M

Usuários mensais no Brasil

Nossa missão é dar a todos uma voz e revelar o mundo.

Acreditamos que todos têm o direito de expressar opiniões e que o mundo se torna melhor quando ouvimos, compartilhamos e nos unimos por meio das nossas histórias.

Mantendo o YouTube seguro tendo os 4Rs como princípios fundamentais

Remover

Conteúdo que viola nossas políticas

Recomendar

Fontes confiáveis para notícias e informações

Reduzir

A disseminação de desinformação prejudicial e conteúdo borderline

Recompensar

Criadores com padrões cada vez mais altos

Removemos conteúdo que viola nossas políticas e cobrem 8 verticais principais

Políticas do YouTube

Incitação ao ódio e assédio

Discurso de ódio: conteúdo que promova ódio ou violência contra grupos com base em atributos protegidos, como idade, gênero, raça, classe, religião, orientação sexual ou condição de veterano de guerra.

Essa política também inclui formas comuns de ódio on-line, como **desumanização de membros desses grupos**; **caracterização deles como inferiores** ou **doentes**; promoção de **ideologias de ódio**, como o nazismo; promoção de **teorias da conspiração sobre esses grupos**; ou **negação da ocorrência de eventos violentos bem documentados**, como tiroteios em escolas.

Políticas do YouTube

Incitação ao ódio e assédio

Assédio e bullying virtual: conteúdo direcionado a um indivíduo com insultos maldosos ou prolongados com base em atributos intrínsecos, incluindo o status de grupo protegido ou traços físicos. Essa política também inclui comportamento prejudicial, como insultar ou humilhar menores deliberadamente; fazer ameaças, bullying ou doxxing; ou encorajar comportamento abusivo de fãs.

Usamos a combinação de **pessoas e tecnologia** para sinalizar conteúdo para nossa revisão

91.2%

dos vídeos removidos do YouTube entre janeiro e março de 2022 foram detectados **automaticamente**.

>67%

dos vídeos removidos do YouTube globalmente foram eliminados com, no máximo, **10 visualizações**.

Usuários e revisores confiáveis também têm uma importante papel na detecção de conteúdo que viola nossas políticas

>330 mil

vídeos foram removidos do YouTube, entre janeiro e março de 2022, a partir de denúncias da nossa comunidade.

Revisores Confiáveis também desempenham um papel importante na detecção de conteúdo problemático

EUROPOL

aarambh
BREAKING THE SILENCE AGAINST CHILD SEXUAL EXPLOITATION

AMADEU ANTONIO STIFTUNG
INITIATIVEN FÜR ZIVILGESELLSCHAFT UND DEMOKRATISCHE KULTUR

INSTITUTE for STRATEGIC DIALOGUE

NO HATE SPEECH MOVEMENT

jugendschutz.net

SaferNet Brasil

여성부
Ministry of Gender Equality

NATIONAL CENTER FOR MISSING & EXPLOITED CHILDREN

THE INTERNATIONAL CENTRE FOR THE STUDY OF RADICALISATION AND POLITICAL VIOLENCE ICSR

Family Online Safety Institute

DEMOS
Instituto Centroamericano de Estudios para la Democracia Social

GW Program on Extremism

crecer en red
DERECHOS DEL NIÑO EN LA ERA DIGITAL

ALL TOGETHER NOW
Erasing Racism

MASYARAKAT ANTI FITNAH INDONESIA
TURN BACK HOAX
MAFINDO

Office of the eSafety Commissioner

ADL
Anti-Defamation League®

licra
Ligue Internationale Contre le Racisme et l'Antisémitisme

COMITE INTERNATIONAL GENEVE

CIPESA
Promoting Effective and Inclusive ICT Policy in Africa

The WAHID Institute
Seeding Plural and Peaceful Islam

Revisores consideram contexto **EDSA** para determinar se há ou não uma violação

Educacional

Documental

Científico

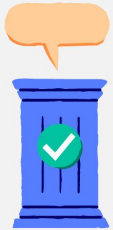
Artístico

Compartilhamos nosso progresso na **remoção de conteúdo que viola nossas políticas de discurso de ódio**

95.947 | Vídeos foram removidos da plataforma de janeiro a março de 2022

+ 57 MI | Comentários foram retirados da plataforma de janeiro a março de 2022

Remover



Recomendar

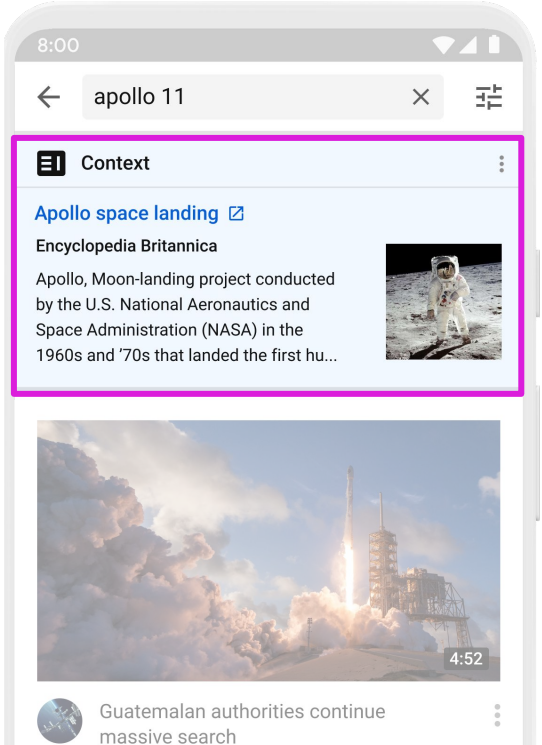
Fontes confiáveis de notícias e informações

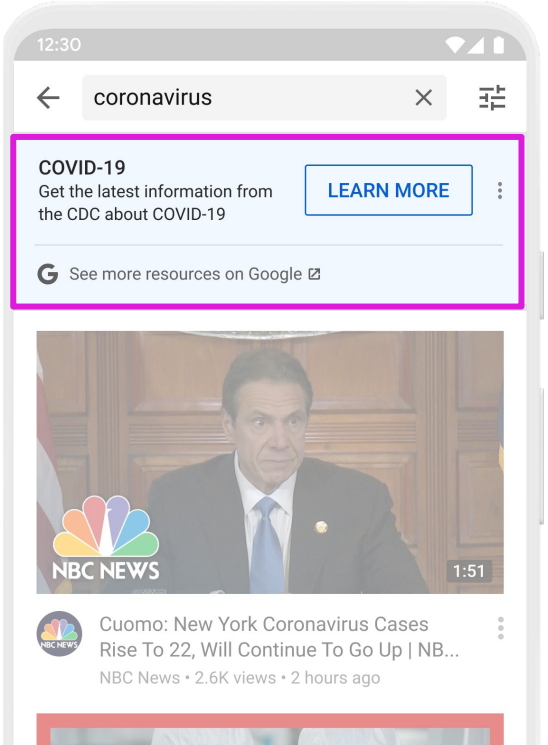
Reduzir

Recompensar

Painel de informações para tópicos sujeitos a informações falsas

Para assuntos sujeitos a informações falsas, o conteúdo de fontes de terceiros oferece mais contexto



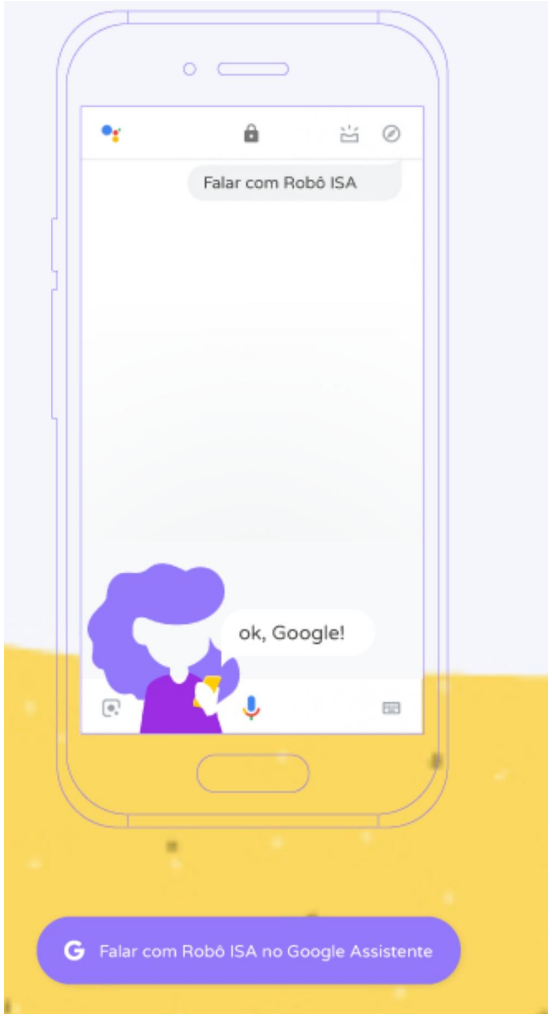


3. Trabalho em conjunto

Parcerias e apoios

Parcerias e apoios

Parcerias e apoios



ISA bot

A mais nova aliada das mulheres durante a quarentena por coronavírus. Uma robô programada para informar e acolher em casos de violência doméstica ou online.

Realização:

think Olga+ Mapa do acolhimento

Obrigada

Google