

# Content Policy at Meta

Emilar Gandhi - Content Policy Stakeholder Engagement

# Content Policy at Meta

# Billions

Of items of  
content shared  
everyday

**3.65  
Billion**

People use our  
family of  
applications every  
month

**2.88  
Billion**

People use our  
family of  
applications every  
day

**87%**

Of users are  
outside USA and  
Canada

# Meta's Content Policy Team



## **CONTENT POLICY**

Develops our  
Community  
Standards

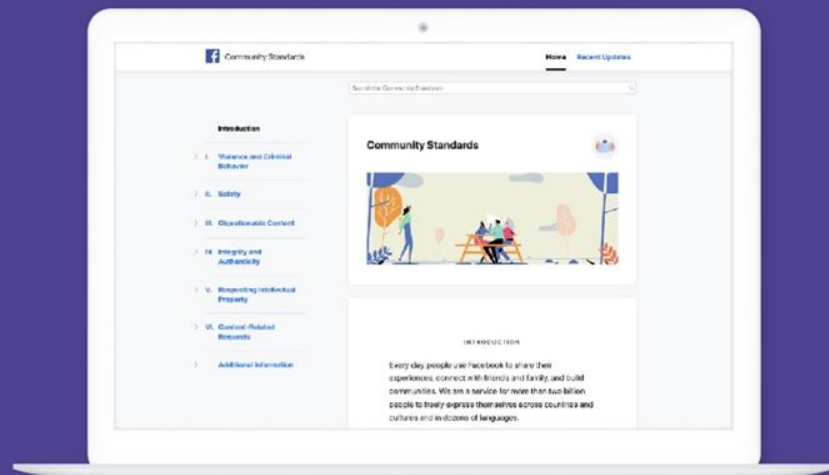
## **COMMUNITY INTEGRITY**

Builds the  
technology that  
enforces our  
Community  
Standards at scale

## **GLOBAL OPERATIONS**

Enforces our  
Community  
Standards through  
human review

# Community Standards



# Overview

We have policies for areas including:

**Suicide & Self-Injury**

**Fraud & Deception**

**Dangerous  
Individuals &  
Organizations**

**Hate Speech**

**Cruel & Insensitive**

**Violence & Incitement**

**Bullying /  
Harassment**

**Human Exploitation**

**Inauthentic  
Behavior**

**Privacy Violations &  
Image Privacy Rights**

**Spam**

**Hacked Accounts**

**Memorialization**

**Coordinating Harm &  
Publicizing Crime**

**Adult Nudity &  
Sexual Activity**

**Violent & Graphic Content**

**Regulated  
Goods**

# VOICE

Create a place  
for expression and  
give people voice.

## PRIVACY

We are committed  
to protecting  
personal privacy  
and information.

## SAFETY

Expression that has  
the potential to  
intimidate, exclude  
or silence others  
and isn't allowed on  
Facebook.

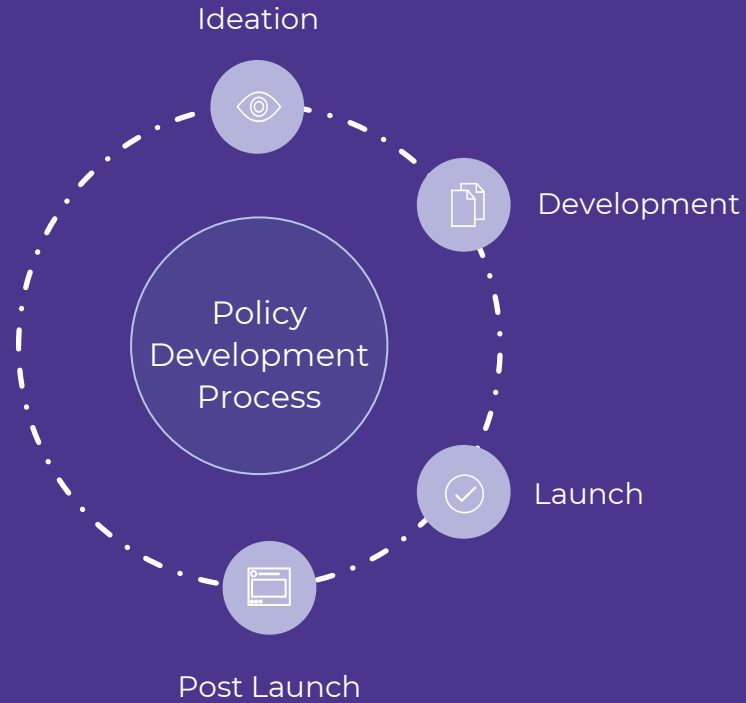
## AUTHENTICITY

Authenticity creates  
a better  
environment for  
sharing.

## DIGNITY

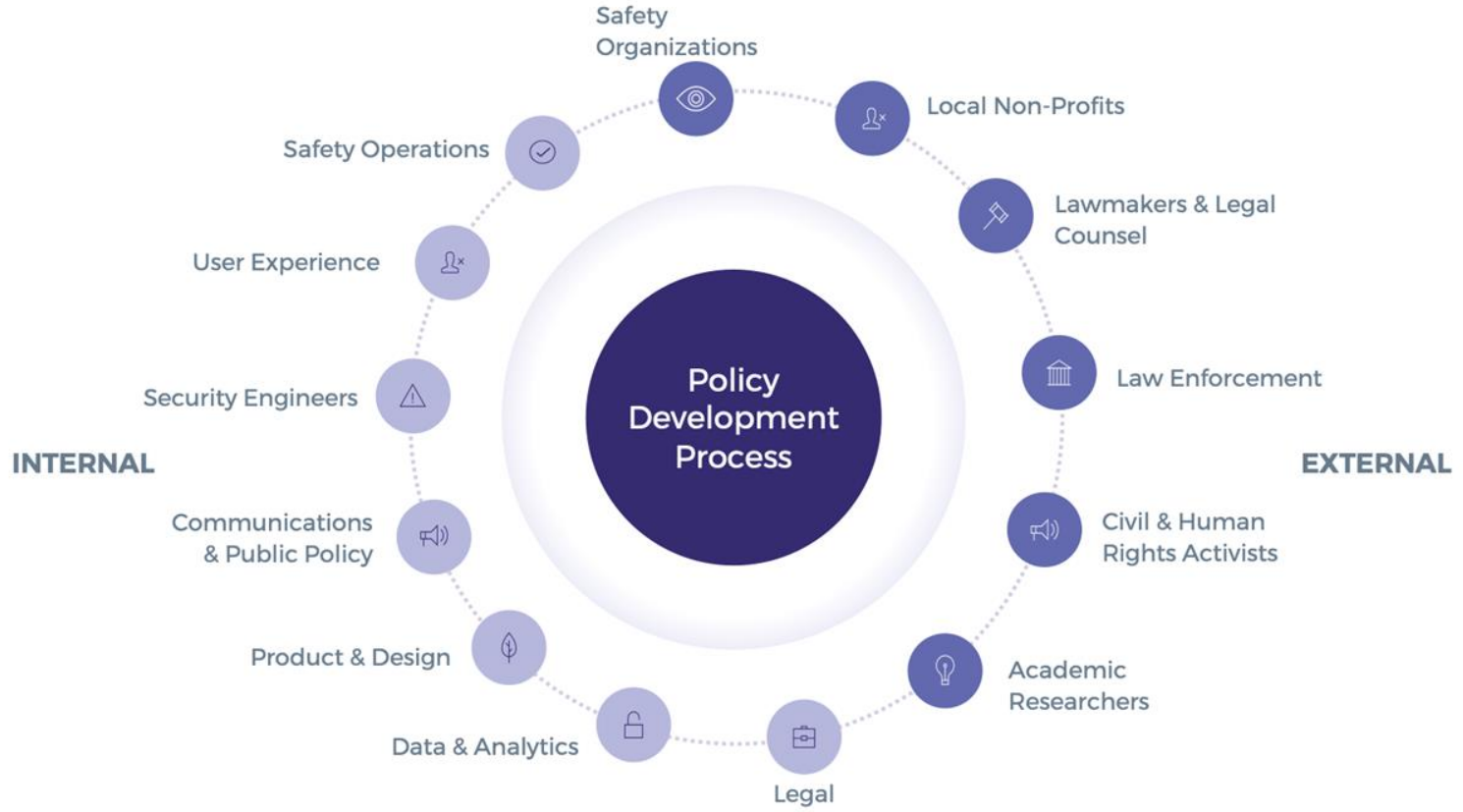
All people are  
equal in dignity  
and rights.

# Policy Development Process

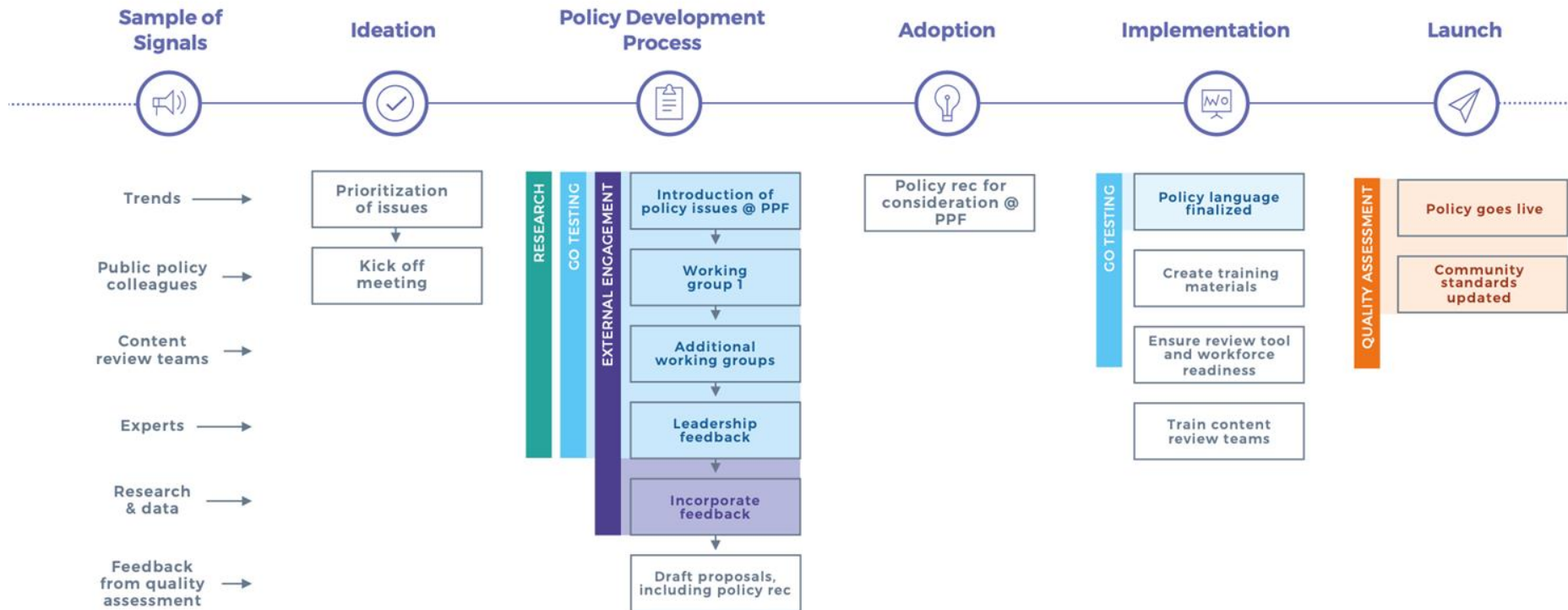




# Who We Work With



# Community Standards: Our Policy Development Process



# Stakeholder Engagement on Content Policy

## Areas of Expertise

Stakeholder Engagement's role is to strengthen our content policies by bringing global knowledge and feedback into the policy development process.

**We cover the following policy areas:**

- Community Standards
- Business Policies
- Misinformation, Content Ranking, New Integrity, and Security Policies
- Product Policies
- Trusted Partner Channel (TP)

## Stakeholder Engagement





# Hate Speech

We do not allow hate speech on Facebook because it creates an environment of intimidation and exclusion, and in some cases may promote real-world violence.

Violent or dehumanizing speech  
Harmful stereotypes  
Statements of inferiority  
Expressions of contempt  
Disgust or dismissal  
Cursing  
Calls for exclusion or segregation

# Defining hate speech

Person or group of people, not  
entities, ideologies or  
institutions

A direct **attack** on **people** based on their  
**protected characteristics**

- Race
- Ethnicity
- National origin
- Disability
- Religious affiliation
- Caste
- Sexual orientation
- Sex
- Gender identity
- Serious disease



- Age (when referenced with another protected characteristic)
- Occupation (when referenced with another protected characteristic)
- Refugees, migrants, immigrants and asylum seekers (protected from the most severe attacks)



We allow for speech that is critical of institutions and practices.



Some people share content containing hate speech to raise awareness.



This policy has evolved over time to include greater protections.



# Hate Speech

All groups receive the same protections under our Hate Speech policy.

Some people share content containing hate speech to raise awareness. Our policies are designed to allow room for this type of speech.

We allow for speech that is critical of institutions and practices provided it isn't likely to contribute to physical harm.

One global policy, which is kept up to date.

We rely on context.

# **Bullying and Harassment**

## Bullying and Harassment

We distinguish between public figures and private individuals because we want to allow discussion, which often includes critical commentary of people who have a large public audience.

For public figures, we remove attacks that are severe as well as certain attacks where the public figure is directly tagged in the post or comment. For private individuals, our protection goes further.



## Bullying and Harassment

Tier 1	Tier 2	Tier 3	Tier 4
<p>Attacks based on a sexual assault, sexual exploitation, or domestic abuse experience</p> <p>Calls for self-injury or suicide</p> <p>Severe sexualized commentary</p> <p>Derogatory sexualized photoshopped imagery</p>	<p>Calls for death or serious disease, epidemic disease or disability</p> <p>Female-gendered cursing terms when used in a derogatory way</p> <p>Claims about sexual activity or sexually transmitted diseases</p> <p>Negative physical descriptions</p>	<p>Comparisons to animals or insects</p> <p>Sexualizing another adult</p> <p>Negative photoshop</p> <p>Content ranking individuals based on physical appearance</p>	<p>Targeted cursing</p> <p>Claims about romantic involvement, sexual orientation or gender identity</p> <p>Negative character or ability claims</p> <p>Expressions of contempt or disgust</p>

Private individuals

(adults must self-report)

Public figures

(purposeful exposure required)

Limited scope public figures

**Thank you.**