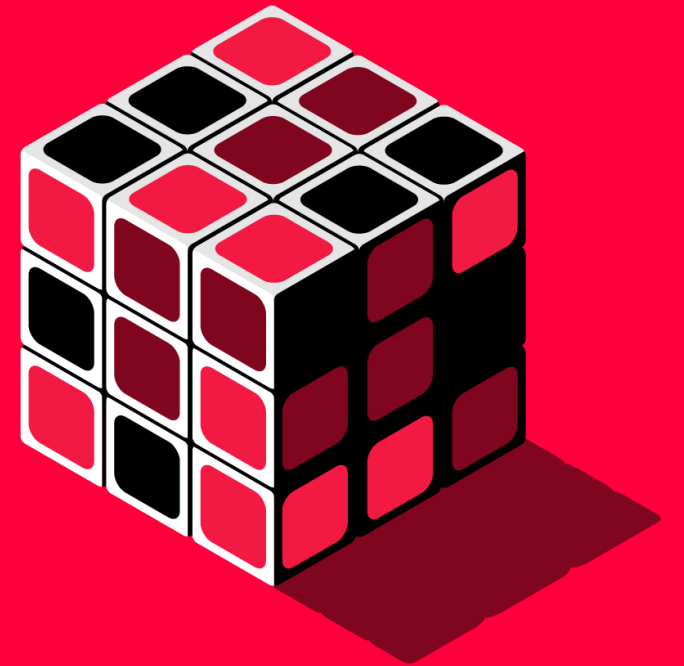WFA World Federation of Advertisers

GARM:
An Overview for ABG

18 Oct 2021

# GARM: The Global Alliance for Responsible Media

GARM is the Global Alliance for Responsible Media, **a cross-industry initiative** that **unites advertisers**, **media agencies**, **media platforms** and **industry groups** to **confront the threat of harmful content online**.

The goal is to **remove harmful content from advertiser-supported media on digital media platforms**, and ensure that the industry works together via uncommon collaboration against this challenge. This new way of working prioritizes harnesses executive collaboration, enabled by media experts and technicians.

GARM was **started by the World Federation of Advertisers** and is **supported by the CMO Growth Council of the Association of National Advertisers** in the US.

WFA

GARM: Recognizing the need for an all-in approach

COORDINATION:
PREVENTING ECOSYSTEM EXPLOITATION
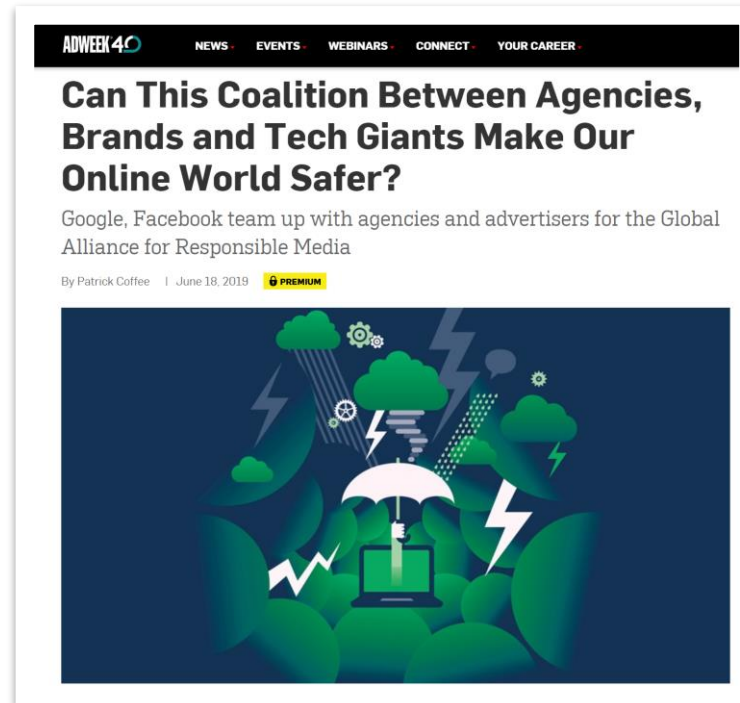
More viewer safety for consumers

Less risk for brand advertisers

More credibility for platforms

A sustainable industry

WFA

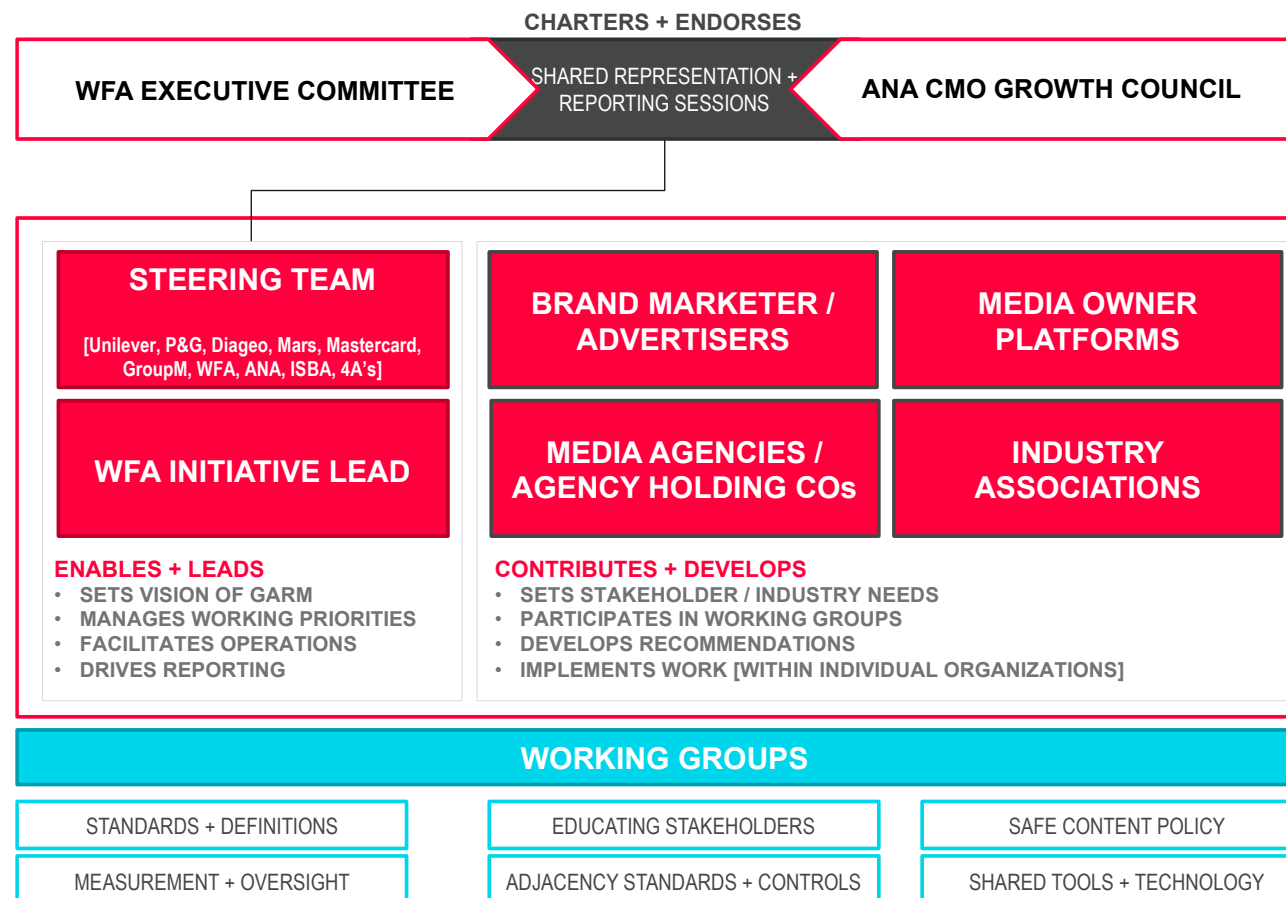# GARM: Governance for prioritized development

EXECUTIVE OVERSIGHT:

- The GARM will be overseen by the WFA Executive Committee + ANA CMO Growth Council
- The GARM will report into both groups via shared representation + reporting sessions
- This will allow for both organisations to endorse and aporove the GARM recommendations
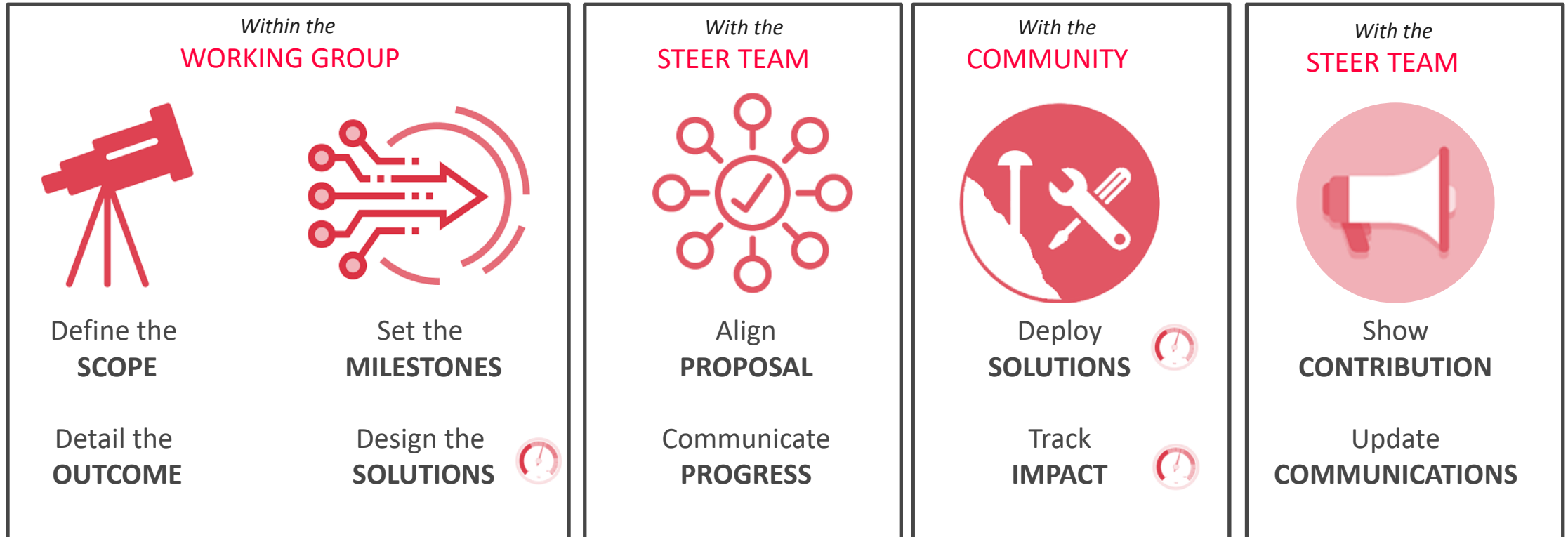
PRIORITY SETTING + FACILITATION:

- The GARM operations: vision, priority setting, facilitation, reporting, review of work wil be done by the Steering Team (of founders) and a WFA consultant
- GARM Steering Team members will represent constituent groups, and will also help drive working group work

DEVELOPMENT + DEPLOYMENT

- The GARM members from all areas will be asked to participate in on research around needs and priorities
- GARM members will have the opportunity to participate in work streams
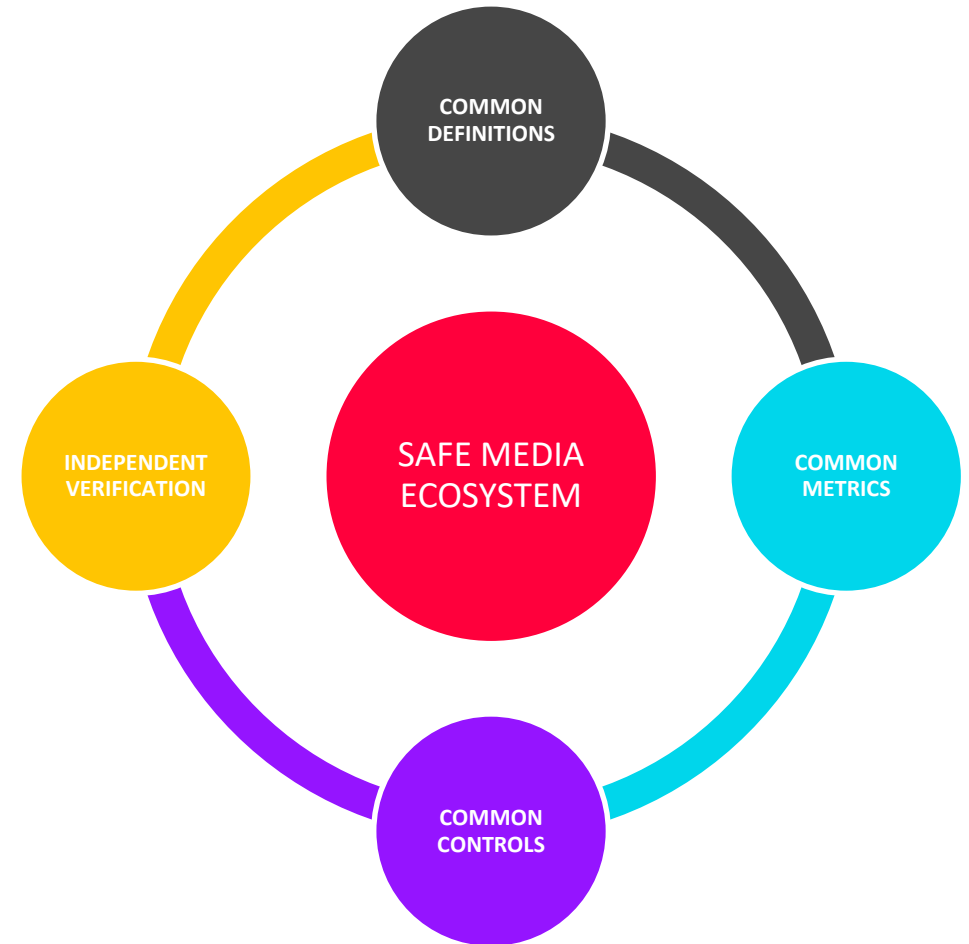- GARM recommendations will be implemented by individual organisations

CHARTERS + ENDORSES

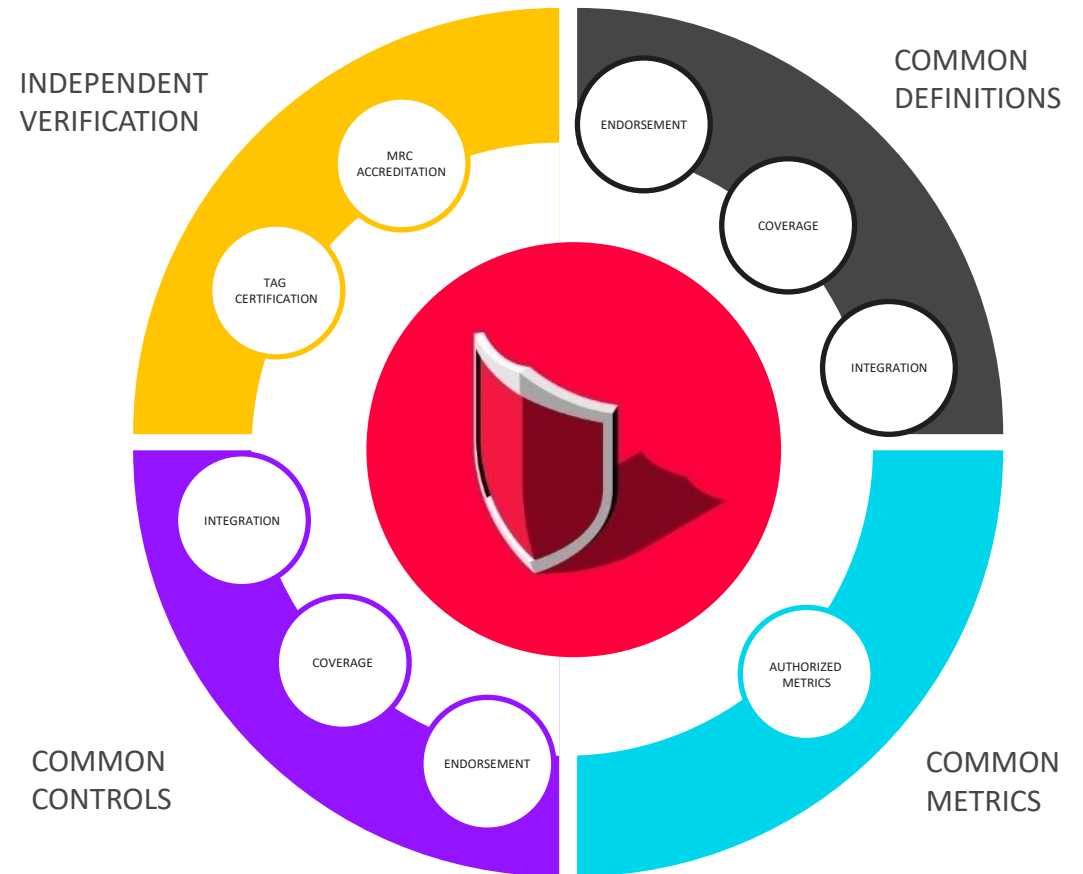| WFA EXECUTIVE COMMITTEE	SHARED REPRESENTATION + REPORTING SESSIONS	ANA CMO GROWTH COUNCIL

STEERING TEAM

[Unilever, P&G, Diageo, Mars, Mastercard, GroupM, WFA, ANA, ISBA, 4A's]

WFA INITIATIVE LEAD

BRAND MARKETER / ADVERTISERS

MEDIA OWNER PLATFORMS

MEDIA AGENCIES / AGENCY HOLDING COs

INDUSTRY ASSOCIATIONS

ENABLES + LEADS
- SETS VISION OF GARM
- MANAGES WORKING PRIORITIES
- FACILITATES OPERATIONS
- DRIVES REPORTING

CONTRIBUTES + DEVELOPS
- SETS STAKEHOLDER / INDUSTRY NEEDS
- PARTICIPATES IN WORKING GROUPS
- DEVELOPS RECOMMENDATIONS
- IMPLEMENTS WORK [WITHIN INDIVIDUAL ORGANIZATIONS]

WORKING GROUPS

STANDARDS + DEFINITIONS	EDUCATING STAKEHOLDERS	SAFE CONTENT POLICY
MEASUREMENT + OVERSIGHT	ADJACENCY STANDARDS + CONTROLS	SHARED TOOLS + TECHNOLOGY

WFA

GARM: How we develop solutions

Within the **WORKING GROUP**	*With the* **STEER TEAM**	*With the* **COMMUNITY**	*With the* **STEER TEAM**
Define the **SCOPE** Set the **MILESTONES**	Align **PROPOSAL**	Deploy **SOLUTIONS**	Show **CONTRIBUTION**
Detail the **OUTCOME** Design the **SOLUTIONS**	Communicate **PROGRESS**	Track **IMPACT**	Update **COMMUNICATIONS**

WFA

GARM:
How our 'Agenda Items' create a safer media ecosystem

COMMON DEFINITIONS

INDEPENDENT VERIFICATION

SAFE MEDIA ECOSYSTEM

COMMON METRICS

COMMON CONTROLS

WFA

GARM:
How our 'Agenda Items' create a safer media ecosystem

8 Steps for markets to take forward on GARM

We've broken a local mobilization of GARM into 4 phases with 8 key steps:

In **PLAN** it's essential to make sure that we have a critical mass of members in the NAC to scale GARM locally. It's advisable to convene your board to obtain approval and also secure participation.

In **PREPARE**, making GARM assets like the Brand Safety Floor + Suitability Framework accessible and relevant by translation and also assessing safety topics against local cultural and regulatory contexts.

In **MOBILIZE** we recommend making sure that you rollout efforts with the right brands, agencies and platforms that reflect local or regional needs. It's also key that NACs start to attend or follow GARM Community Calls.

In **MEASURE** we encourage you to take on member feedback and needs and work with GARM to address those, while also tracking the pull-thru of your efforts.

PLAN

1. Gauge Member Interest
2. Secure+ Coordinate Global / Local Brands to Mobilize

PREPARE

3. Translate Floor + Suitability Framework
4. Assess Local Topics (Cultural / Regulatory) For Attention

MOBILIZE

5. Plan a Representative Launch (Brands / Agencies / Platforms)
6. Attend GARM Community Calls

MEASURE

7. Feedback on Reception / Future Needs
8. Track Uptake / Use

WFA

Developing common definitions for content safety and risk

GARM: Brand Safety Floor + Suitability Framework

SOLUTION OVERVIEW

The Brand Safety Floor + Suitability Framework is a consistent set of definitions that create a safety and risk standard for monetized content across the industry. This is now included in MRC and IAB TechLab Standards.

- *For Advertisers + Agencies* this is an essential foundation for targeting to be used in media planning & buying, and use of contextual targeting tools
- *For Media Platforms + Ad Tech Partners* this is an essential framework for establishing buying modes, tools, and reporting.

SOLUTION STATUS

- Deployed – September 2020

- Forthcoming update in 4Q 2021 to incorporate Misinformation as a 12th category

BENEFITS

1. Creates a safety standard for the advertising industry to ensure that harmful content is not monetized

2. Establishes a system based on risk to improve transparency and control for advertiser sellers and buyers

3. Drives a framework that is common across the entire industry

LIMITATIONS

1. Requires marketer + agency activation (ideally thru multidisciplinary workshop)

2. Monetization standard and moderation standards are mapped (v a straight 1:1 relationship)

3. Roll-out is continuous and requires investment in time and resource

WFA

GARM Agenda Item 1: Standards + Definitions
Developing common definitions for content safety and risk

GARM: Brand Safety Floor + Suitability Framework — GARM Global Alliance for Responsible Media (Brand Safety Floor + Suitability Framework table)

Question	Critical Step	YouTube	Facebook	Instagram	Twitter	TikTok	Pinterest	Snapchat
Has the platform endorsed the GARM Brand Safety Floor + Framework?	Executive Endorsement	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Has the platform linked content moderation and/ or monetization policies to GARM standards?	Mapping Exercise	Yes	Yes	Yes	Yes	Yes	Yes	Yes
To what extent has the platform embedded GARM standards into first-party buying tools?	Coverage in all ad inventory	Yes	No	No	No	No	Yes	Yes
	Inclusion in or linked to first-party tools	Yes	Yes Select inventory	No	Yes	No	No	No
	Availability of third-party integrations	Yes	Yes Select inventory	No	Underway	Yes Select inventory	No	No

GARM: Implementation Tracker – Sept 2021

WFA

GARM Agenda Item 1: Standards + Definitions
Developing common definitions for content safety and risk

GARM: Brand Safety Floor + Suitability Framework — Global Alliance for Responsible Media

Brand Suitability Framework: Sensitive content appropriate for advertising supported by enhanced advertiser controls

CURRENT UPTAKE: POLICY MAPPING

GARM Content Category	Relevant Platform Policy						
	YouTube	Facebook	Instagram	Twitter	TikTok	Pinterest	Snap
Adult & Explicit Sexual Content	• Nudity & Sexual Content • Child Safety	• Adult Nudity and Sexual Activity	• Adult Nudity and Sexual Activity	• Non- Consensual Nudity • Sensitive Media • Child Sexual Exploitation	• Minor safety – sexual exploitation of minors • Adult nudity and sexual activities	• Adult Sexual Services • Adult Content	• Sexually Explicit Content
Arms & Ammunition	• Firearms	• Regulated Goods: Firearms	• Regulated Goods: Firearms	• Illegal or certain regulated good or services	• Illegal activities and regulated goods - weapons	• Dangerous Goods and Activities	• Regulated Goods
Crime & Harmful acts to individuals and Society, Human Right Violations	• Harmful or Dangerous Content • Hate Speech • Harassment or cyberbullying	• Violent and Graphic Content • Bullying and Harassment • Child Nudity and Sexual Exploitation • Suicide and Self-Injury	• Violent and Graphic Content • Bullying and Harassment • Child Nudity and Sexual Exploitation • Suicide and Self-Injury	• Violence • Abuse and harassment	• Illegal activities and regulated goods -criminal activities	• Child Sexual Exploitation • Self-Harm • Harassment & Criticism	• Threatening / Violence / Harm:
Death, Injury or Military Conflict	• Violent or Graphic Content • Harmful or Dangerous Content • Suicide & Self-Injury	• Violent and Graphic content	• Violent and Graphic content	• Promoting Self-harm	• Violent and Graphic Content	• Graphic Violence and Threats	• Threatening / Violence / Harm
Online piracy	• Fake Engagement • Impersonation • Sale of illegal or regulated goods or services • YouTube Terms of Service	• Intellectual Property • Copyright • Intellectual Property Counterfeit • Intellectual Property Trademark	• Intellectual Property • Copyright • Intellectual Property Counterfeit • Intellectual Property Trademark	• Copyright • Trademark	• Integrity and authenticity – intellectual property violations	• Copyright • Trademark	• Spam
Hate speech & acts of aggression	• Hate Speech	• Hate speech • Bullying and Harassment	• Hate speech • Bullying and Harassment	• Hateful Conduct	• Hate Speech • Hateful Behavior	• Hateful Activities	• Threatening / Violence / Harm
Obscenity and Profanity, including language, gestures, and explicitly gory, graphic or repulsive content intended to shock and disgust	• Violent or Graphic Content • Age Restriction	• Hate Speech • Bullying and Harassment	• Hate Speech • Bullying and Harassment	• Sensitive Media	• Hateful Behavior – Slurs • Harassment & Bullying	• Harassment & Criticism	
Illegal drugs, tobacco, e - cigarettes, vaping	• Sale of Illegal or Regulated Goods or Services • Harmful or dangerous content	• Regulate Goods: Drugs	• Regulate Goods: Drugs	• Illegal or certain regulated goods or services	• Illegal activities and regulated goods – drugs, controlled substances, alcohol and tobacco	• Dangerous Goods and Activities	• Regulated Goods
Spam & Malware	• Spam, Deceptive Practices & Scams	• Spam	• Spam	• Private information • Impersonation • Platform manipulation	• Integrity and authenticity – spam and fake engagement	• Spam	• Spam
Terrorism	• Violent criminal organizations	• Dangerous Organizations: Terrorism • Dangerous Organizations: Organized Hate	• Dangerous Organizations: Terrorism • Dangerous Organizations: Organized Hate	• Terrorism or Violent Extremism	• Violent Extremism • Dangerous individuals and organizations - Terrorists and terrorist organizations	• Violent Actors	• Terrorism
Debated Sensitive Social Issues		• Hate Speech • Bullying and Harassment	• Hate Speech • Bullying and Harassment		• Hateful Behavior	• Civic Misinformation • Conspiracy Theories • Medical Misinformation	
Other	• COVID Misinformation Policy	• COVID-19 and Vaccine Policy and Protections	• COVID-19 and Vaccine Policy and Protections	• Covid Integrity • Covid-19 Misleading Information			

GARM: Implementation Tracker – Sept 2021

WFA

GARM Agenda Item 1: Standards + Definitions
Developing common definitions for content safety and risk

GARM: Brand Safety Floor + Suitability Framework — Global Alliance for Responsible Media

Brand Suitability Framework: Sensitive content appropriate for advertising supported by enhanced advertiser controls

(Content category framework table with High Risk / Medium Risk / Low Risk columns covering: Adult & Explicit Sexual Content; Arms & Ammunition; Crime & Harmful acts to individuals and Society, Human Right Violations; Death, Injury or Military Conflict; Online piracy; Hate speech & acts of aggression; Obscenity and Profanity; Illegal Drugs/Tobacco/e-cigarettes/Vaping/Alcohol; Spam or Harmful Content)

CURRENT UPTAKE: INTEGRATIONS

Question	Critical Step	ZEFR	CHANNEL FACTORY	PIXABILITY	PEER 39	DOUBLE VERIFY	INTEGRAL AD SCIENCE	OPENSLATE	MOAT
To what extent has the provider embedded GARM standards into first-party buying tools?	Content Targeting / Avoidance	Yes	Yes	Yes	Yes	Partial	Partial	Partial	In Progress
	Post-campaign Reporting	Yes	Yes	Yes	In Progress	Partial	In Progress	Partial	In Progress
Where are the service providers active with GARM standards?	YouTube	Y	Y	Y	N	Y reporting only	N	Y	N
	Facebook	Y video only	N	N	N	Y	N	Y video only	N
	Instagram	N	N	N	N	N	N	N	N
	Twitter	N	N	N	N	In Progress	In Progress	N	N
	TikTok	Beta	N	N	N	N	Beta	Y	N
	Pinterest	N	N	N	N	N	N	N	N
	Snap	N	N	N	N	N	N	N	N
	Other				Open web	Open web	Open web		Open web

GARM: Implementation Tracker – Sept 2021

WFA

GARM Agenda Item 2: Measurement + Oversight
Establishing common metrics to track progress

GARM Global Alliance for Responsible Media

GARM Aggregated Measurement Report

CORE QUESTION	AUTHORIZED METRIC	DEFINITION + OVERVIEW	RATIONALE
How safe is the platform for consumers?	Prevalence of violating content or Violative View Rate	The percentage of views that contain content that is deemed as violative	Establishes a ratio based on typical user content consumption. Prevalence or Violative View Rate examines views of unsafe/violating content as a proportion of all views.
How safe is the platform for advertisers?	Prevalence of violating content or Advertising Safety Error Rate	The percentage of views that contain content that is deemed as violative The percentage of views of monetized content that contain violative content	Monetization prevalence examines unsafe content viewed as a proportion of monetized content viewed
How effective is the platform in policy enforcement?	Removals of Violating Content + Removal of Violating Accounts Removals of Violating Content expressed by how many times it has been viewed	Pieces of violating content removed Accounts removed due to repeat policy violation Pieces of violating content removed categorized by how many times they were viewed by users	Platform teams spend a considerable amount of time removing violating content and bad actors from their platforms – the magnitude of the efforts should be reported to marketers. It is also important to marketers to understand how many times harmful content has been removed.
How does the platform perform at correcting mistakes?	Appeals Reinstatements	Number of pieces of violating content removed that are appealed Number of pieces of violating content removed that are appealed and then reinstated	Platform should be responsive to their users and policy should be consistent with a policy of free and safe speech. For this reason we look at appeals and reinstatement of content removed.

SOLUTION OVERVIEW

GARM created a consistent and consolidated process to allow the media industry to focus on the questions that matter, and more importantly highlight methodological best practices in answering core questions: how safe is the platform for consumers, how safe is the platform for advertisers, how effective is the platform in enforcing its policies, and how responsive is it in correcting mistakes.

- *For Advertisers + Agencies* this this creates a common source of information in assessing progress and trends
- *For Media Platforms* this is an essential resource for rolling forward best practices in transparency

SOLUTION STATUS

- Deployed – April 2021

- Forthcoming Volume 2 in October 2021

- Ongoing discussions on Local Market Incident Reporting

BENEFITS

1. Creates a single access point for multiple transparency reports

2. Streamlines data and focuses on core questions relevant to strategic choices (investment, independent vendor controls)

3. Creates a methodological baseline in Accepted Metrics

LIMITATIONS

1. Metrics are a reflection of platform policies, which have been mapped to GARM standards, not GARM policies

2. Data is global and may not expose local issues and incidents

3. Data cannot be summed up for industry totals

WFA

GARM Agenda Item 2: Measurement + Oversight
Establishing common metrics to track progress

GARM: Implementation Tracker – Sept 2021

CURRENT UPTAKE: AUTHORIZED METRICS

Question	Authorized Metric	YouTube	Facebook	Instagram	Twitter	TikTok	Pinterest	Snapchat
How safe is the platform for consumers?	Prevalence / Violative View Rate	Authorized Metric	Authorized Metric	Authorized Metric	Next Best Measure	Next Best Measure	Not Submitted	Next Best Measure
How safe is the platform for advertisers?	Advertiser Safety Error Rate or Prevalence	Authorized Metric	Authorized Metric	Authorized Metric	Next Best Measure	Next Best Measure	Not Submitted	Not Submitted
How effective is the platform at enforcing its safety policies?	Removals of violating content	Authorized Metric	Authorized Metric	Authorized Metric	Authorized Metric	Next Best Measure	Authorized Metric	Authorized Metric
	Removal of violating accounts by views	Authorized Metric	Not Submitted	Not Submitted	Next Best Measure	Authorized Metric	Not Submitted	Authorized Metric
	Removal of violating accounts	Authorized Metric	Authorized Metric	Not Submitted	Authorized Metric	Not Submitted	Authorized Metric	Authorized Metric
How responsive is the platform in correcting mistakes?	Appeals (pieces of content)	Authorized Metric	Authorized Metric	Authorized Metric	Not Submitted	Not Submitted	Not Submitted	Not Submitted
	Reinstatements (pieces of content)	Authorized Metric	Authorized Metric	Authorized Metric	Not Submitted	Not Submitted	Not Submitted	Not Submitted

WFA

GARM Agenda Item 3: Adjacency Standards.+ Controls

Improving controls for ad placement in user content

SOLUTION OVERVIEW

In order to improve transparency and control for advertising suitability, we nee to develop common standards and tools that allow for advertisers, agencies, and platforms to more consistently manage advertising placement. This solution drives standards across platforms to help the industry do that.

BENEFITS

1. Opportunity to have platforms allow marketers and agencies to drive contextual targeting and avoidance with common standards

2. Allows for common standards to exist based on formats, across platforms

3. Drives a system of checks and balances via first-and-third-party tools

SOLUTION STATUS

• Conducted Immersions to understand current practice and Community Study to identify needs/expectations

• Gained agreements to drive common standards for individual formats

• Standards to be shared 4Q 2021

LIMITATIONS

1. Still under development towards an agreement

2. May take some platforms longer to deliver to market

WFA

GARM Agenda Item 3: Adjacency Standards.+ Controls
Improving controls for ad placement in user content

CURRENT PARTICIPANTS

YouTube FACEBOOK Twitter TikTok Snapchat Pinterest twitch Spotify

[NB – *still in development*]

WFA

GARM Agenda Item 4: Independent Verification

Improving industry transparency via independent audits

SOLUTION OVERVIEW

Brand safety operations require a rigorous set up and effective operation for the sake of user safety and advertiser trust. GARM believes in driving independent verification on process and execution delivered via independent audits linked to an industry standard. These are delivered by both TAG Brand Safety Control (ex-DTSG) and MRC Brand Safety Controls + Integrations Audits.

BENEFITS

1. Creates independent certification on brand safety structure resource, and validates operational effectiveness

2. Creates a common, consistent, comparable standard for brand safety effectiveness

SOLUTION STATUS

- YouTube successfully passes MRC audit (2020) and is in process of annual recertification

- Facebook commences MRC audit

- Twitter in discovery of MRC audit

LIMITATIONS

1. Audits can be difficult for smaller, newer platforms as they are resource intensive (time, money)

2. Auditing bodies for the advertising industry will have to contend with potentially competitive regulatory audit requests

WFA

Improving industry transparency via independent audits

CURRENT UPTAKE: INDEPENDENT VERIFICATION

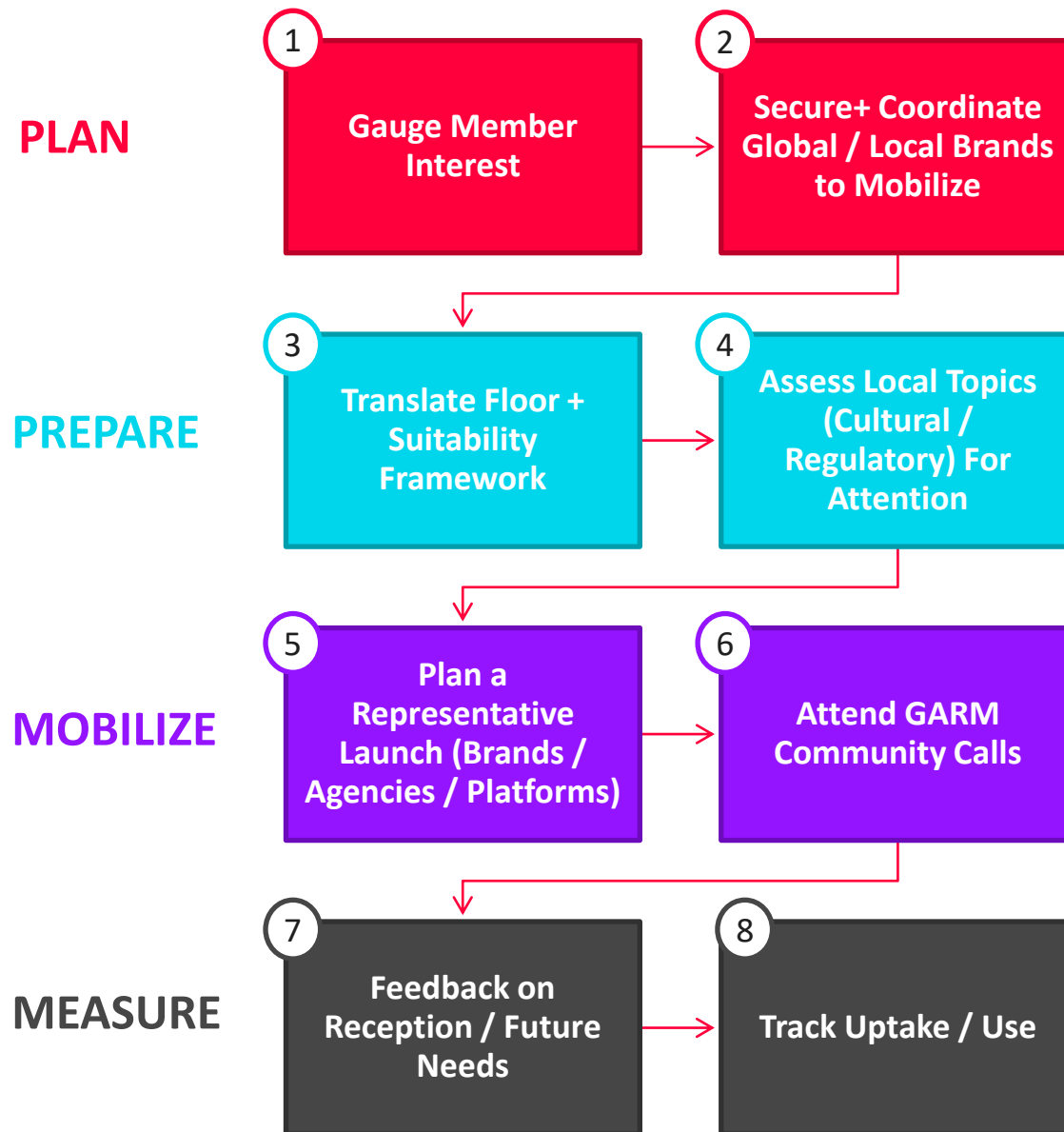| Question	Critical Step	YouTube	Facebook	Instagram	Twitter	TikTok	Pinterest	Snapchat
Which audit standards has the platform secured?	MRC Content-level Brand Safety Controls Audit	Passed	In Progress	In Progress	In Progress	No	No	No
	TAG Brand Safety Certified	Certified	Certified	Certified	Certified	Certified	No	No

WFA

8 Steps for markets to take forward on GARM

We've broken a local mobilization of GARM into 4 phases with 8 key steps:

In **PLAN** it's essential to make sure that we have a critical mass of members in the NAC to scale GARM locally. It's advisable to convene your board to obtain approval and also secure participation.

In **PREPARE**, making GARM assets like the Brand Safety Floor + Suitability Framework accessible and relevant by translation and also assessing safety topics against local cultural and regulatory contexts.

In **MOBILIZE** we recommend making sure that you rollout efforts with the right brands, agencies and platforms that reflect local or regional needs. It's also key that NACs start to attend or follow GARM Community Calls.

In **MEASURE** we encourage you to take on member feedback and needs and work with GARM to address those, while also tracking the pull-thru of your efforts.

PLAN

1. Gauge Member Interest
2. Secure+ Coordinate Global / Local Brands to Mobilize

PREPARE

3. Translate Floor + Suitability Framework
4. Assess Local Topics (Cultural / Regulatory) For Attention

MOBILIZE

5. Plan a Representative Launch (Brands / Agencies / Platforms)
6. Attend GARM Community Calls

MEASURE

7. Feedback on Reception / Future Needs
8. Track Uptake / Use

WFA

GARM: Why NAC membership matters, and successful activation would look like

WHY PARTICIPATION MATTERS

INPUT ON PRIORITIES

SCALE IMPLEMENTATION LOCALLY

FULFILLING ADVERTISER ACCOUNTABILITIES

DRIVING RELEVANCE / IMPACT

WHAT SUCCESS LOOKS LIKE

- ❑ Solutions translated to the local language
- ❑ Selective additions based on cultural and regulatory factors in mind
- ❑ Regular feature in NAC / local platform connects
- ❑ Regular feedback on positive implementation / challenges

Include why we're inviting you early and GARM gaps

WFA

GARM:
An Overview for ABG

18 Oct 2021