Instagram Youth Well-being Assurance Audit - Phase 2 Audit Report

Company confidential - for intended recipients only

Martin Schneider
Anna Tchernina
Kalyan [redacted]
Yanan [redacted]
Pratima [redacted]

September, 2023

Assurance Audit provides an objective and independent evaluation of the design and operating effectiveness of internal controls.

Meta

PX-01080, Page 1 of 28

Tennessee v. Meta
Case No. 23-1364-IV
PX-01080

TABLE OF CONTENTS

a. Executive Summary

b. Detailed Results

c. Appendix
    1. Management Action Plans - Observations (Phase 1) - Summary
    2. Communication Channels
    3. Audit Report Rating Guidelines
    4. Observation Remediation Guidelines
    5. Management Action Plans - Closed Observations (Phase 1)

PX-01080, Page 2 of 28

Company confidential - for intended recipients only

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504911
METATNAG-011-00622942

EXECUTIVE SUMMARY

Objective:

Instagram (IG) leadership has publicly committed[1] to raising the standards for protecting teens online. As part of this work, in 2021 and 2022, IG announced stricter privacy and safety default settings for teens, and added messaging restrictions on IG Direct (IGD) to protect teens against unwanted contact from adults.
In Q4 2022, Internal Audit (IA) performed the IG Youth Well-being Phase 1 Assurance Audit, which focused on content integrity for teens and interaction restriction controls between teens and potential violators. In Phase 2, we assessed the design and operating effectiveness of the key controls related to IG teen default restrictive settings, and IGD private messaging restrictions between adults and teens. In addition, IA performed remediation validations on the open observations[2] that were identified in the Phase 1 audit.

Scope:

IA independently assessed the design and operating effectiveness of the key controls implemented for the following 3 scope areas:

- **IG Teen Default Restrictive Settings:** IG pre-selects the most restrictive and private settings for teens (e.g., private account by default, contact information not requested by default during conversion to business account).
- **IGD Private Messaging Restrictions:** Unconnected adults[3] are limited from initiating 1:1 private messages with teens, or adding them to private group messages. Teens are limited from initiating 1:1 messages with High Risk Adults[4] that they do not follow.
- **Phase 1 Audit Observations:** 15 open observations from Phase 1 audit, as summarized in Appendix I.

Results Summary: Significant Improvements Needed

Based on the testing completed as of 7/31/2023, the teen default settings for Private by Default (PBD), Sensitive Content Control (SCC), Tag, Guide, Reels Remix and Mention (TGRM) were operating effectively and management has remediated all the observations from the Phase 1 audit. For other default settings and IGD restrictions, IA identified 8 High risk and 1 Low risk observations. Based on the overall effectiveness of controls for in-scope areas, an "Orange" rating is assigned to this report. The results and observation ratings in this report are based on the design and operating effectiveness of internal controls only.
Key observation themes are summarized below:

- Inconsistent suppression on data collection of contact information for teens during the business account conversion; Lack of suppression on data collection for users with unknown age
- Inaccurate and insufficient implementation of restrictive privacy default setting on Website Embeds[5] for teens
- Inconsistent 1:1 and group messaging restrictions between teens and unconnected adults, including High Risk Adults

1. IG Blog Posts on 3/17/2021 and 12/7/2021. Teen (often used interchangeably with "youth") is generally defined as a user over 13 years old and under 18 years old. The definition may vary based on product features and demographics.
2. Observations and Management Action Plans closed before March 15, 2023 were independently validated and deemed complete in IA's standard issue tracking process; therefore were excluded from this audit.
3. Unconnected means the recipient does not follow the sender, and they have not messaged each other previously. Also, young adults (18-20) are able to initiate 1:1 messaging with unconnected teens in close age proximity (i.e., the age gap is equal to or less than 3 years).
4. High Risk Adults are adults that have suspicious behavioral patterns (e.g., Negative user feedback such as reports and blocks).
5. The Website Embeds is an IG product feature that allows a content (e.g., post, reel) or profile of a public IG account to be shared by third-party website outside of IG (e.g., blogs and articles) using an embed code.

PX-01080, Page 3 of 28

CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504912
METATNAG-011-00622943

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
1	**IG Teen Default Restrictive Settings:** Business Account Conversion	**Inconsistent Suppression on Data Collection for Teen Users** The IG Small and Medium Business (SMB) Growth team does not always suppress the prompt for teens to enter contact information during the conversion to a business account. As a result, teens' contact information may be collected and displayed as public information, which may not fully meet Meta's external commitment*. For example, during the period from 6/8/2023 to 7/9/2023, IA observed that out of 995K business account conversion for teens, 13K (1.3%) were prompted to enter contact information during the conversion flow. * As a regulatory response in 2021, Meta made an external commitment that teen users who convert to a business account should not have the option to input contact information within the conversion flow.	High	The Instagram SMB Growth team will: 1. Investigate the leakage to identify the root cause of discrepancy for teen accounts at the time of conversion to a business account. 2. Implement remediation to prevent the "edit contact information" prompt from appearing for teen accounts at the time of conversion to a business account.	1. 10/31/2023 2. 11/30/2023†	Garett [redacted] (Software Engineer, IG SMB Business Growth) Yi [redacted] (Director, IG Core Ads Engineering)

† Based on the ongoing regulatory inquiry on the teen business accounts, the IG Product Counsel indicated PRIV Due to the dependency of Legal decision and the complexity of identifying the root cause, the IG SMB Business Growth team and the IG Product Counsel have requested PRIV

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504913
METATNAG-011-00622944

PX-01080, Page 4 of 28

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
2	**IG Teen Default Restrictive Settings:** Business Account Conversion	**Lack of Suppression on Data Collection for Users with Unknown Age** The IG SMB Growth team does not suppress the prompt to collect contact information when the users' age data is unavailable during the conversion to a business account. As a result, if the user is a teen, their information may be collected and displayed as public information, which may not fully meet Meta's external commitment.	High	The IG SMB Growth team will incorporate logic for the IG business account conversion flow to prevent the "edit contact information" prompt from appearing for accounts with missing age or where age data cannot be retrieved at the time of conversion.	10/15/2023	Garett [redacted] (Software Engineer, IG SMB Business Growth) Yi [redacted] (Director, IG Core Ads Engineering)

PX-01080, Page 5 of 28

Company confidential - for intended recipients only

CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504914
METATNAG-011-00622945

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
3	**IG Teen Default Restrictive Settings:** Website Embeds* - U.S.	**Inaccurate Implementation of Age Criteria for U.S. Teen's Default Privacy Setting** The IG Growth Acquisition team does not default to the more restrictive privacy setting of "opt-out" on Website Embeds for 16 and 17 years old U.S. teen public accounts, as required in the Launch Manager (LaMa). As a result, teens' posts and profiles may be publicly displayed outside the IG app.	High	The IG Growth Acquisition team will: 1. Review the age band difference between LaMa and current implementation, determine the logic based on current risk assessment and obtain cross-functional (XFN) approval. 2. Implement required changes based on the approval.	1. 10/31/2023 2. 11/30/2023†	Sona [redacted] (Product Manager, IG Growth Product) Po-Lung [redacted] (Software Engineering Manager, IG Growth Acquisition) Morgan [redacted] (Director, Product Management, IG Communities)

* The Website Embeds is an IG product feature that allows a content (e.g., post, reel) or profile of a public IG account to be shared by third-party website outside of IG (e.g., blogs and articles) using an embed code. Website Embeds allow the user's profile or content to be shared along with their username and a link to their Instagram profile.

† Due to the dependency on the XFN collaboration, the IG Growth Acquisition team has requested a 30-day extension to IA's standard 60-day remediation guideline for High risk observations.

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504915
METATNAG-011-00622946

PX-01080, Page 6 of 28

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
4	**IG Teen Default Restrictive Settings:** Website Embeds - U.S.	**Ineffective Implementation of U.S. Teen Default Privacy Setting** The IG Growth Acquisition team does not always default to the more restrictive privacy setting of "opt-out" on Website Embeds for the U.S. teen public accounts under 16 years old. As a result, teens' posts and profiles may be publicly displayed outside the IG app. For example, IA observed that out of 564K U.S. under 16 public accounts registered in May 2023, 43.7% (246K) were not pre-selected to "opt-out".	High	The IG Growth Acquisition team will: 1. Identify the root cause for the discrepancy in the "opt-out" default for Website Embeds. 2. Implement remediation to always pre-select Website Embeds to "opt-out" for intended users. 3. Establish monitoring to track effectiveness of "opt-out" as default for Website Embeds setting and set up automated alerts to notify oncall in case of anomalies.	1. 10/31/2023 2 & 3. 12/31/2023†	Sona [redacted] (Product Manager, IG Growth Product) Po-Lung [redacted] (Software Engineering Manager, IG Growth Acquisition) Morgan [redacted] (Director Product Management, IG Communities)

† Due to the dependency on XFN collaboration and the product decision made on observation #3, the IG Growth Acquisition team has requested a 60-day extension to IA's standard 60-day remediation guideline for High risk observations.

Company confidential - for intended recipients only

CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504916
METATNAG-011-00622947

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
5	**IG Teen Default Restrictive Settings:** Website Embeds - Non-U.S.	**Lack of Privacy Setting and Restrictive Default for Non-U.S. Teens** The IG Growth Acquisition team does not provide the privacy setting and does not default to the more restrictive setting of "opt-out" on Website Embeds for non-U.S. teen public accounts. As a result, non-U.S. teen public accounts cannot control whether their content and profiles are displayed outside of IG, and do not have the most restrictive default setting.	High	The IG Growth Acquisition team will: 1. Determine if the Website Embeds setting and teen default should be rolled out to non-U.S. teens. Document the decision and rationale in the LaMa tool. 2. Based on #1 implement decision made in the LaMa tool.	1. 10/31/2023 2. 12/31/2023†	Sona [redacted] (Product Manager, IG Growth Product) Po-Lung [redacted] (Software Engineering Manager, IG Growth Acquisition) Morgan [redacted] (Director, Product Management, IG Communities)

† Due to the dependency on the XFN collaboration and the complexity of global roll-out, the IG Growth Acquisition team has requested a 60-day extension to IA's standard 60-day remediation guideline for High risk observations.

PX-01080, Page 8 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504917
METATNAG-011-00622948

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
6	**IGD Private Messaging Restrictions:** 1:1 Messaging	**Inconsistent Private Messaging Restriction Between Teens and Unconnected Adults*** The Messenger Well-being team does not always prevent unconnected adults from initiating a 1:1 thread with teens on IGD. As a result, teens may be exposed to unsolicited contact from unknown adults, which may not fully meet Meta's external commitment**. For example, on 6/10/2023, IA observed that out of 2.7M new threads initiated by an adult with a teen globally, 238K (9%) were from unconnected adults. ** Meta made an external commitment in the IG Blog Post on 3/17/2021 that unconnected adults cannot initiate a private 1:1 message with a teen on IGD.	High	The Messenger Well-being team will: 1. Establish thread initiation event-level logging and perform analysis to determine the effectiveness in unconnected adult - teen intervention in 1:1 threads. 2. Implement remediation to prevent the unconnected adults from initiating 1:1 threads with the teens. 3. Establish monitoring to track adult-side 1:1 messaging effectiveness and set up automated alerts to notify oncall in case of anomalies.	1. 9/30/2023 2. 10/31/2023 3. 11/30/2023†	Anand [redacted] (Product Manager, Well-being XFN Team) Felipe [redacted] (Software Engineer, MWB Foundations - Support & Care) Ananth [redacted] (Engineering Director, Trust ENG Leadership)

* Adults included in IGD messaging restriction testing are above 21 years old, including high risk adults.
† Due to the dependency on the XFN collaboration and the complexity of the investigation, the IG Messenger Well-being team has requested a 30-day extension to IA's standard 60-day remediation guideline for High risk observations.

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504918
METATNAG-011-00622949

PX-01080, Page 9 of 28

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
7	**IGD Private Messaging Restrictions:** Group Chats	**Inconsistent Group Chat Restriction Between Teens and Unconnected Adults** The Messenger Well-being team does not always prevent unconnected adults from initiating group threads with teens on IGD. As a result, teens may be exposed to unsolicited contact from unknown adults via group threads. For example, on 6/10/2023, IA observed that out of 315K teens added to group threads by adults globally, 35K (11%) were by unconnected adults.	High	The Messenger Well-being team will: 1. Collaborate with XFN teams to determine required risk coverage and acceptable tolerance for the unconnected adult - teen intervention in group threads. Document the decision. 2. Implement mitigation to limit the unconnected adult - teen interactions in group threads. 3. Establish monitoring to track adult-side group messaging effectiveness and set up automated alerts to notify oncall in case of anomalies.	1. 9/30/2023 2. 10/31/2023 3. 11/30/2023†	Anand [redacted] (Product Manager, Well-being XFN Team) Felipe [redacted] (Software Engineer, MWB Foundations - Support & Care) Ananth [redacted] (Engineering Director, Trust ENG Leadership)

† Due to the dependency on the XFN collaboration and the complexity of the investigation, the IG Messenger Well-being team has requested a 30-day extension to IA's standard 60-day remediation guideline for High risk observations.

PX-01080, Page 10 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504919
METATNAG-011-00622950

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
8	**IGD Private Messaging Restrictions:** Messaging High Risk Adult	**Inconsistent Private Messaging Restriction Between Teens and High Risk Adults (HRA*)** The Messenger Well-being team does not always prevent teens from initiating 1:1 conversations on IGD with HRA they do not follow. As a result, teens may be exposed to potential negative experience in interactions with HRA. For example, on 7/3/2023, IA observed that out of 647K new threads initiated by teens with HRAs globally, there were 231K (36%) threads where the teen did not follow the HRA.	High	The Messenger Well-being team will: 1. Collaborate with XFN teams to determine required risk coverage and acceptable tolerance for the HRA - teen intervention. Document the decision. 2. Implement mitigation to limit HRA - teen or enforcement leakage to within acceptable tolerance. 3. Establish monitoring to track teen-side 1:1 messaging leakage and set up automated alerts to notify oncall in case of anomalies.	1. 9/30/2023 2. 10/31/2023 3. 11/30/2023†	Anand [redacted] (Product Manager, Well-being XFN Team) Felipe [redacted] (Software Engineer, MWB Foundations - Support & Care) Ananth [redacted] (Engineering Director, Trust ENG Leadership)

*High Risk Adults (HRA) heuristics on IGD is defined as:
- 7+ Negative Feedback eXperience (NFX) by teens against the adult in past 28 day (inc. 1+ direct NFX) OR 2 Child Sexual Abuse Material (CSAM)
- >3 year age gap
- Not followed by teen OR <= 10 mutual reciprocal.

Adults included in IGD messaging restriction testing are above 21 years old.

† Due to the dependency on the XFN collaboration and the complexity of the investigation, the IG Messenger Well-being team has requested a 30-day extension to IA's standard 60-day remediation guideline for High risk observations.

Company confidential - for intended recipients only

CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504920
METATNAG-011-00622951

DETAILED RESULTS AND MANAGEMENT ACTION PLANS

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
9	**IG Teen Default Restrictive Settings:** User Experience	**Inaccurate Visibility of Privacy Settings - Guides and Website Embeds** The IG Engineering team does not hide the Guides and the Website Embeds settings for private accounts. As a result, private users may mistakenly think that they can allow others to publicly share their post by creating Guides or Website Embeds.	Low	1. The IG Teen Safety team will make the Guides toggle unavailable in the User Interface (UI) for private accounts. 2. The IG Growth Acquisition team will make the Website Embeds toggle unavailable in the UI for private accounts.	1. Pending Validation 2. 10/31/2023	1. Yichen [redacted] (Software Engineer, IG Well-being Teen Safety) 2. Alex [redacted] (Software Engineer, IG Growth Acquisition) Conan [redacted] (Software Engineer, IG Growth Acquisition)

PX-01080, Page 12 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

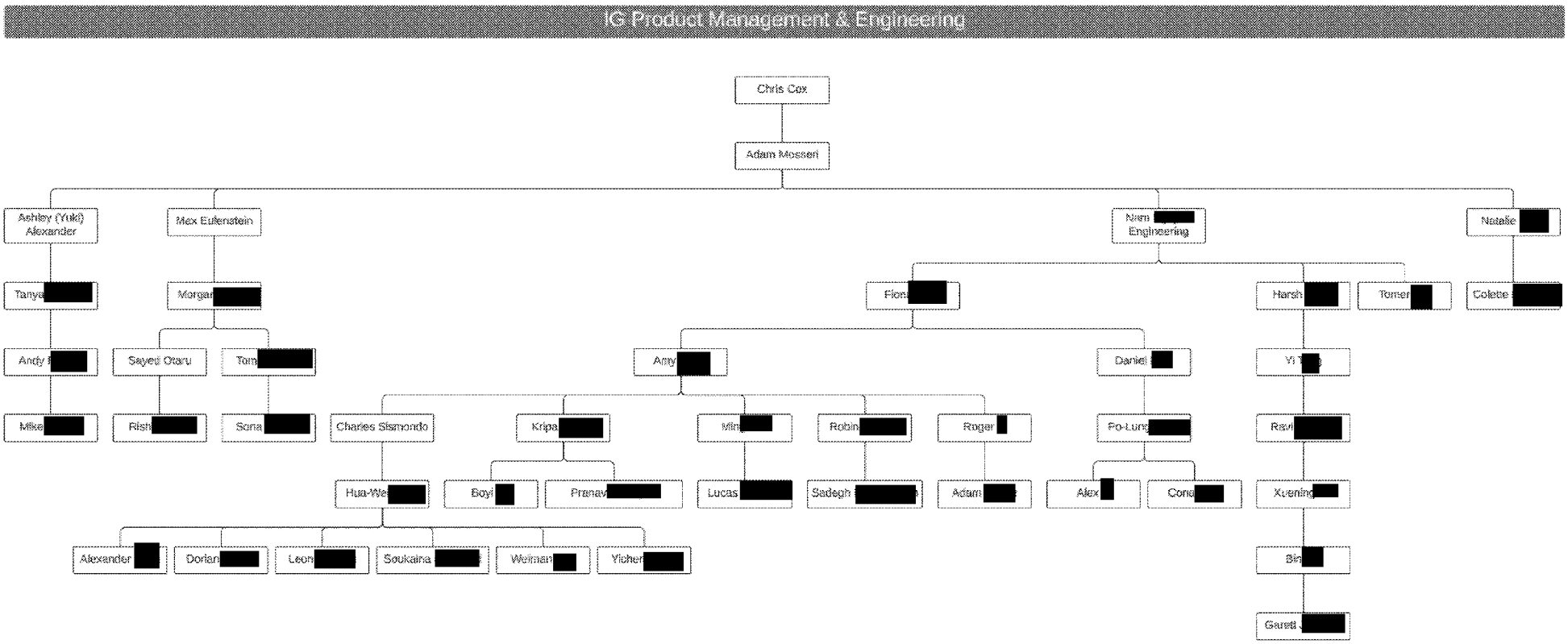
META3047MDL-034-00504921
METATNAG-011-00622952

APPENDIX I - MANAGEMENT ACTION PLANS - OBSERVATIONS (PHASE 1) - SUMMARY

Business Area	Observations	MAPs Total	MAPs Closed	MAPs Past Due
Age-based Gating, Filter and Demotion	1	1	1	0
Limiting Interactions	4	10	10	0
Sensitive Content Control	10	16	16	0
Total	**15**	**27**	**27**	**0**

PX-01080, Page 13 of 28

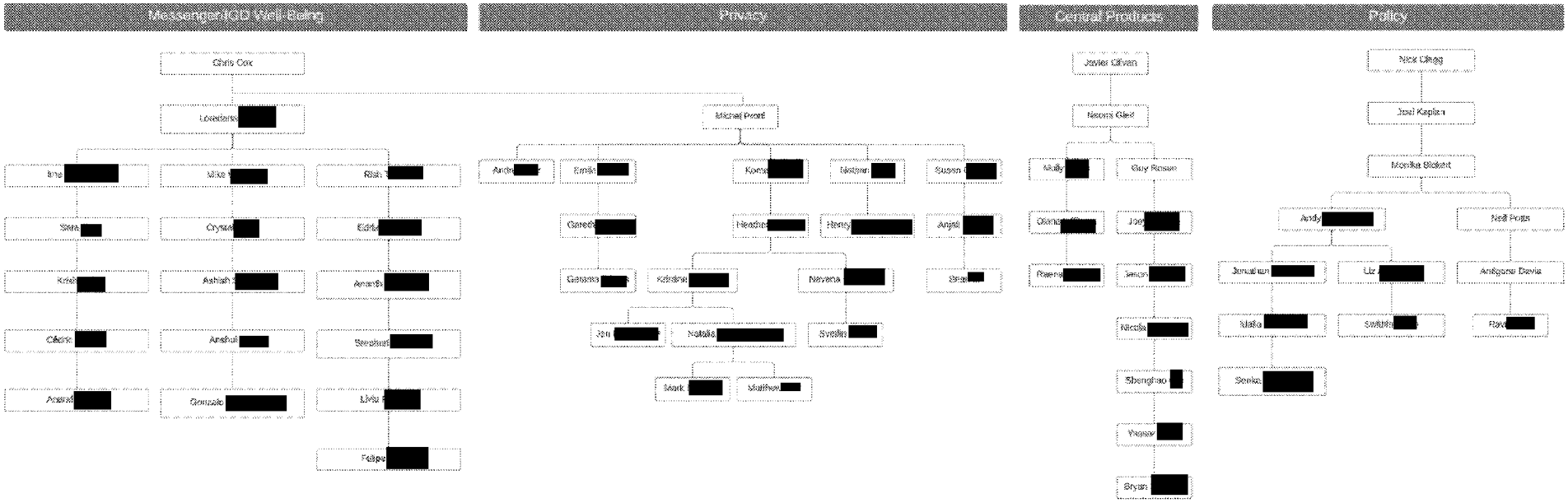
APPENDIX II - COMMUNICATION CHANNEL



Company confidential - for intended recipients only

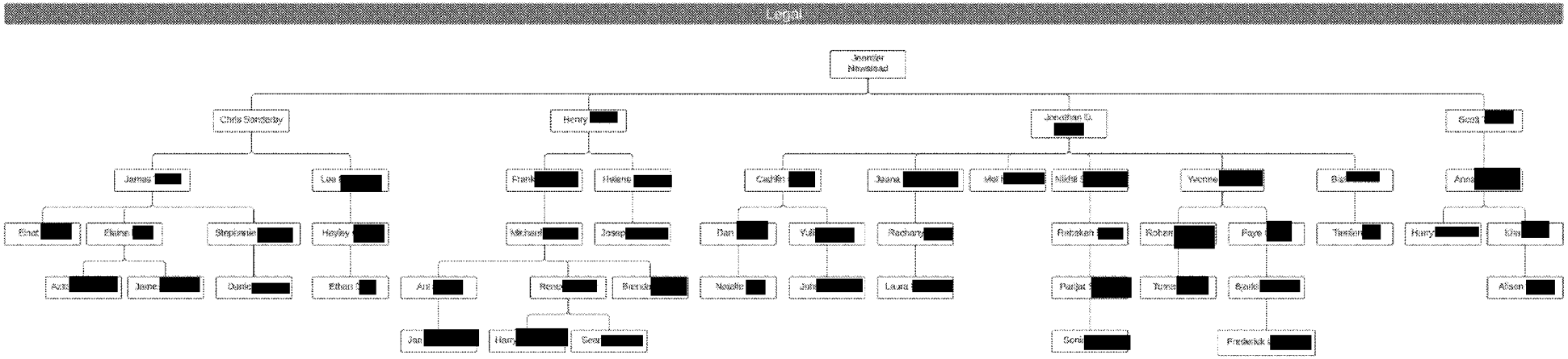
CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504923
METATNAG-011-00622954

APPENDIX II - COMMUNICATION CHANNEL



Company confidential - for intended recipients only

APPENDIX II - COMMUNICATION CHANNEL



PX-01080, Page 16 of 28

Company confidential - for intended recipients only

CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504925
METATNAG-011-00622956

APPENDIX III - AUDIT REPORT RATING GUIDELINES

Rating	Rating Criteria
Effective	Limited "Low risk" observations: Processes and controls provide reasonable assurance that risks are managed and objectives will be met.
Some Improvements Needed	Limited "Medium risk" observations or a number of "Low risk" observations: Processes and controls require improved effectiveness and efficiency to provide reasonable assurance that risks are managed and objectives will be met.
Significant Improvements Needed	Limited "High risk" observations or a number of "Medium risk" observations: Processes and controls include gaps that require significant improvements to provide reasonable assurance that risks are managed and objectives will be met.
	Several "High risk" observations: Processes and controls include critical gaps that require immediate improvement to provide reasonable assurance that risks are managed and objectives will be met.

PX-01080, Page 17 of 28

Company confidential - for intended recipients only

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504926
METATNAG-011-00622957

APPENDIX IV - OBSERVATION REMEDIATION GUIDELINES

Rating	Rating Definition	Remediation Expectations	AROC Communication
High Risk	Process and control gaps that may result in (i) significant deficiency or material weakness in internal controls, (ii) severe decrease in operational effectiveness/efficiency, (iii) violations in policy, laws or regulations, or (iv) damage to Meta's reputation leading to significant business impact.	Resolve issue within 60 days	Yes
Medium Risk	Process and control gaps that may result in (i) deficiency in internal controls, (ii) significant decrease in operational effectiveness/ efficiency, (iii) non-compliance with policy, laws or regulations, or (iv) damage to Meta's reputation leading to moderate business impact.	Target remediation within 120 days	Yes
Low Risk	Process and control gaps that may result in (i) minor non-compliance with policy, laws or regulations, (ii) deficiency in internal controls, or (iii) decrease in operational effectiveness/efficiency.	Target remediation within 180 days	No

PX-01080, Page 18 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504927
METATNAG-011-00622958

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
1	Limiting Interactions	**Inconsistent filtering of IIC account recommendation** The IG Well-being Teen Safety team does not always limit teen accounts from being recommended to the IIC potential violators* and vice versa. As a result, teens' accounts may be recommended to potential violators and vice versa via account recommendation surfaces. For example, on a sampled day tested, on the Accounts You May Follow (AYMF) surface, IA identified: 1) 7.2% (~1.4M out of ~19M) of potential violator accounts in AYMF were recommended to at least one teen. 2) 4.0% (~3.4M out of ~85M) of AYMF recommendations to potential violators were recommending teen accounts. * This control applies to IIC Medium potential violators, defined by the product team as users above 18 without verified tag, having 1 or more Child Sexual Exploitation (CSE) violation in past 90 days or a Community Integrity (CI) classifier for IIC score score greater than or equal to 0.3.	High	1. Perform validation testing upon MaCSA migration and investigate any recommendation leakage of IIC potential violator accounts to teen accounts and vice versa. 2. Investigate any potential delays in the propagation of predicted age information from Core Dimensions and mitigate discrepancy, if needed. 3. Implement remediation to further limit exposure between teens and potential violators through AYMF recommendations.	1&2. 2/28/2023 3. 3/31/2023†	Weiman[redacted] (Software Engineer, IG Well-being Teen Safety) Soukaina[redacted] (Software Engineer, IG Well-being Teen Safety)

† Due to complexity of identifying the root cause of the leakage and the migration from the original IIC classifier to the new MaCSA encryption resilient detection method, the IGWB Teen Safety team has requested a 30-day extension to IA's standard 60-day remediation guideline for High risk observations.

Company confidential - for intended recipients only

PX-01080, Page 19 of 28

CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504928
METATNAG-011-00622959

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
2	Limiting Interactions	**Inconsistent follow list hiding** The IG Well-being Teen Safety team does not always hide teen accounts from IIC potential violators* and vice versa on the "followers and following" surface. As a result, teen accounts may be displayed to potential violators and vice versa via the follow lists. For example, on 2022-11-25, IA observed that out of ~2.5M IIC potential violators who viewed the follow list surface, 37% (924K) were shown at least one unconnected teen account. * This control applies to IIC Medium potential violators	High	1. Perform validation testing upon MaCSA migration and investigate teen accessibility to discover IIC potential violators via follower and following lists of other accounts and vice versa. 2. Investigate any potential delays in the propagation of predicted age information from Core Dimensions and mitigate discrepancy, if needed. 3. Implement remediation to further limit teens' exposure to IIC potential violator accounts via follow list and vice versa.	1&2. 2/28/2023 3. 3/31/2023†	Weiman [redacted] (Software Engineer, IG Well-being Teen Safety) Soukaina [redacted] (Software Engineer, IG Well-being Teen Safety)

† Due to complexity of identifying the root cause of the leakage and the migration from the original IIC classifier to the new MaCSA encryption resilient detection method, the IGWB Teen Safety team has requested a 30-day extension to IA's standard 60-day remediation guideline for High risk observations.

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504929
METATNAG-011-00622960

PX-01080, Page 20 of 28

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
3	Limiting Interactions	**Inconsistent effectiveness of IIC content recommendation filtering** The IG Well-being Teen Safety team does not always limit recommendations of teen created content to the IIC potential violators* and vice versa. As a result, teens' content may be recommended to IIC potential violators via content recommendation surfaces and vice versa. For a sampled week tested: 1) 2.2% (~27K out of ~1.2M) of IIC potential violator created content on the Explore surface was recommended to at least one teen, accumulating ~466K teen impressions. 2) 1.6% (~83K out of ~5.2M) of content recommended to IIC potential violators on the Explore surface were created by teens, accumulating ~200K potential violator impressions. ** This control applies to IIC Medium potential violators.*	Medium	1. Perform validation testing upon Malicious Child Safety Actor (MaCSA) model migration and investigate any leakage of IIC potential violator created content to teens on the recommendation surfaces, and vice versa. 2. Implement remediation to further limit teens' exposure to IIC potential violator created content, and vice versa.	1. 2/28/2023 2. 3/31/2023	Weiman [redacted] (Software Engineer, IG Well-being Teen Safety) Soukaina [redacted] (Software Engineer, IG Well-being Teen Safety)

PX-01080, Page 21 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504930
METATNAG-011-00622961

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
4	Limiting Interactions	**Undocumented exemption of accounts** The IG Well-being Teen Safety team exempted verified accounts from being classified as IIC potential violators without documented assessment of the risk or rationale. As a result, potential violating verified accounts may not be classified as IIC potential violators.	Medium	1. Reassess the blanket exemption on verified accounts and document the decision and rationale. 2. Implement MaCSA model for potential violator identification based on the reassessment decision.	3/31/2023	Weiman [redacted] (Software Engineer, IG Well-being Teen Safety) Soukaina [redacted] (Software Engineer, IG Well-being Teen Safety) Swithin [redacted] (Product Policy Manager, IG Product Policy)

PX-01080, Page 22 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504931
METATNAG-011-00622962

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
5	Sensitive Content Control (SCC 2.0)	**Inconsistent non-recommendable content flagging** The IG Relevance Integrity team does not always flag non-recommendable media when the classifier scores exceed predetermined thresholds. As a result, users may be exposed to sensitive content via content recommendation surfaces. For a sampled day tested, IA identified 0.2% of content with nudity classifier scores exceeding the integrity threshold that was not flagged as non-recommendable therefore, not enforced.	Medium	1. Investigate and remediate the leakage of content with nudity exceeding the predetermined non-recommendable threshold. 2. Set up monitoring and alert on the volume of content exceeding the non-recommendable thresholds.	3/31/2023	Lucas [redacted] (Software Engineer, IG Relevance Integrity) Bryan [redacted] (Software Engineer, XI)
6	Sensitive Content Control (SCC 2.0)	**Lack of monitoring on content filtering leakage** The IG Relevance Integrity team does not monitor the leakage level of non-recommendable content getting recommended to users. As a result, the team may not promptly identify issues if an excessive number of sensitive content is displayed on recommendation surfaces.	Medium	Set up monitoring and alert on the non-recommendable content prevalence across all SCC settings.	3/31/2023	Lucas [redacted] (Software Engineer, IG Relevance Integrity)

PX-01080, Page 23 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504932
METATNAG-011-00622963

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
7	Sensitive Content Control (SCC 2.0)	**Lack of monitoring on account filtering leakage** The IG Well-being Connections Safety team does not monitor the leakage level of non-recommendable accounts getting recommended on the AYMF surface. As a result, the team may not promptly identify issues if an excessive number of sensitive accounts are displayed on AYMF.	Medium	Set up automated alerts to monitor filtering leakage and maintain an on-call runbook.	3/31/2023	Boyi[redacted] (Software Engineer, IG Well-being Connections Safety)
8	Sensitive Content Control (SCC 2.0)	**Missing alert on non-recommendable account prevalence** The IG Well-being Connections Safety team has not set up automated alerts to monitor non-recommendable account prevalence on the AYMF surface. As a result, if the prevalence of potential violating accounts on the surface increases, the issue may not be identified and resolved promptly.	Medium	1. Investigate discrepancies in impressions count for users with the 'Allow' setting on the 'IG AYMF non-rec prevalence' dashboard. 2. Set up automated alerts to monitor non-recommendable account prevalence on AYMF. 3. Develop and maintain an on-call runbook to address and remediate alerts.	3/31/2023	Boyi[redacted] (Software Engineer, IG Well-being Connections Safety)

PX-01080, Page 24 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504933
METATNAG-011-00622964

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
9	Sensitive Content Control (SCC 2.0)	**Lack of Logging of Content Demotion** The IG Well-being Enforcements & Classifications Platform team does not log the real-time demotion score* and dynamic threshold for non-recommendable content. As a result, the demotions applied cannot be measured or compared against the original expectation. * demotion score indicates the expected demotion strength	Medium	1. Migrate late stage demotion onto Soft Action Kaleidoscopic Enforcement (SAKE) platform. 2. Implement logging of the real-time demotion score and dynamic threshold.	1. 4/30/2023 2. 5/31/2023†	Adam [redacted] (Software Engineering Manager, IG Well-being Enforcements & Classifications Platform)
10	Sensitive Content Control (SCC 2.0)	**Inconsistent content filtering and demotion monitoring** The IG Well-being Enforcements & Classifications Platform team does not always investigate and address filtering and demotion performance alerts promptly. As a result, content filtering/demotion issues may not be identified or resolved promptly.	Medium	1. Implement and maintain an on-call shift summary tracker to: - log alerts with assigned oncall information, and - document relevant details on the issue/resolution to provide context for future alerts. 2. Establish SLA to close alerts promptly.	3/31/2023	Adam [redacted] (Software Engineering Manager, IG Well-being Enforcements & Classifications Platform)

† Due to complexity of migration to SAKE, collaborative development platform for integrity enforcements across Instagram, the IGWB Enforcements & Classification Platform team has requested a 30-day extension to IA's standard 120-day remediation guideline for Medium risk observations.

PX-01080, Page 25 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504934
METATNAG-011-00622965

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
11	Sensitive Content Control (SCC 2.0)	**Lack of input monitoring of P(Severe) v2 model** The IG Well-being Teen Safety team does not monitor the input signal coverage of the P(Severe) v2 model* to assess its reasonableness. As a result, low coverage of specific signals (e.g., profile picture signal) may go undetected, leading to deteriorated model output. * P(Severe) v2 model is the machine learning model that uses profile element approach to identify violating and non-recommendable profiles.	Medium	Set up monitoring and alert on ML input signals for the P(Severe) v2 model in production.	3/31/2023	Pranav [redacted] (Software Engineer, Machine Learning, IG Well-being Teen Safety)
12	Sensitive Content Control (SCC 2.0)	**Lack of performance monitoring of P(Severe) v2 model** The IG Well-being Teen Safety team does not monitor the performance (e.g., precision, recall) of the P(Severe) v2 model. As a result, deteriorating model performance over time may go undetected.	Medium	1. Set up monitoring and alert on performance (Precision/Recall) of the P(Severe) v2 model. 2. Establish model refresh cadence based on performance monitoring of P(Severe) v2 model.	3/31/2023	Pranav [redacted] (Software Engineer, Machine Learning, IG Well-being Teen Safety)

PX-01080, Page 26 of 28

Company confidential - for intended recipients only

CONFIDENTIAL
CONFIDENTIAL
META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017
META3047MDL-034-00504935
METATNAG-011-00622966

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
13	Sensitive Content Control (SCC 2.0)	**Incomplete logging of account demotion** The IG Well-being Discovery Safety team does not completely log account demotion on the search surface. As a result, the effectiveness of account demotion cannot be measured or compared against the original expectation.	Medium	Set up logging of pagination subrequests in the account demotion logging table.	3/31/2023	Sadegh [redacted] (Research Scientist, IG Well-being Discovery Safety)
14	Sensitive Content Control (SCC 2.0)	**Lack of account demotion monitoring on Search surface** The IG Well-being Discovery Safety team does not monitor demotion volume on the Search surface. As a result, account demotion leakage may not be identified and resolved promptly.	Medium	Set up monitoring and alert for the Search surface on: - Demotion volume - Demotion rate, and - Prevalence of non-recommendable accounts on Search across SCC settings.	3/31/2023	Sadegh [redacted] (Research Scientist, IG Well-being Discovery Safety)

PX-01080, Page 27 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504936
METATNAG-011-00622967

APPENDIX V - MANAGEMENT ACTION PLANS - CLOSED OBSERVATIONS (PHASE 1)

#	Business Area	Observation/Risk Description	Rating	Management Action Plan	Due Date	Owner(s)
15	Age-based Gating, Filtering and Demotion	**Lack of monitoring on age-based recommendation filtering** The IG Well-being Teen Safety team does not monitor the input signal coverage and volume for age-based filtering of the potential violating Adult Nudity and Sexual Activity (ANSA) & RGS content. As a result, anomalies of filtering may not be detected.	Medium	1. Set up monitoring and alert to track filtering volumes of ANSA & RGS content and get notifications on significant changes. 2. Create a dataswarm pipeline to measure input signal coverage and set up monitoring alert on the coverage.	1. CLOSED 2. 2/28/2023	Alexander [redacted] (Software Engineer, IG Well-being Teen Safety)

PX-01080, Page 28 of 28

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL021
METATNAG_REPROD_VOL017

META3047MDL-034-00504937
METATNAG-011-00622968