CASE STUDY #02

Platform Enforcement Reliability Audit

EU Digital Services Act - Harmonised Transparency Reports

From 359.85 million reported enforcement measures to a risk-based human-review and quality-governance framework.

PLATFORMS BENCHMARKED	REPORTED MEASURES	AUTOMATION EXPOSURE
5	359.85M	85.24%

Prepared by

Rayan Braccio

Business Operations & Data Analytics Portfolio

July 2026

REPORT CONTENTS / CASE STUDY #02

Table of Contents

Platform Enforcement Reliability Audit / Risk Analytics Portfolio

EXECUTIVE SNAPSHOT / CASE STUDY #02

Executive Dashboard

REPORTED MEASURES **359.85M** Five-platform H2 2025 sample	AUTOMATION EXPOSURE **85.24%** Weighted end-to-end automated share	INTERNAL COMPLAINTS **24.77M** Submitted to platform mechanisms
REVERSED DECISIONS **10.80M** Resolved complaints reversed	RESOLVED REVERSAL **43.66%** Selected-sample resolved outcomes	HIGHEST REVIEW SCORE **65.1** TikTok - review need, not quality

01 / RISK ANALYTICS CASE STUDY

Executive Summary

A decision-ready view of enforcement scale, automation exposure, appeal outcomes and the highest-priority control surface.

REPORTED MEASURES **359.85M** Five-platform H2 2025 sample	AUTOMATION EXPOSURE **85.24%** Weighted Tab 6 share	INTERNAL COMPLAINTS **24.77M** Platform complaint mechanisms
REVERSED DECISIONS **10.80M** Resolved complaints reversed	RESOLVED REVERSAL **43.66%** Not an error rate	HIGHEST CATEGORY AUTOMATION **99.69%** YouTube - Protection of minors

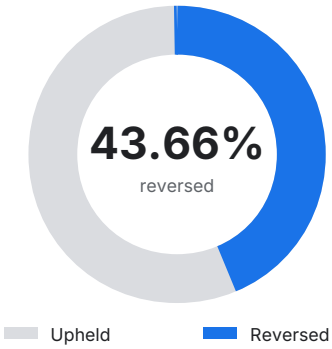
Executive readout

Scale concentration TikTok contributes 71.89% of reported measures and the highest normalized intensity at 1.45M measures per 1M active recipients.

Appeal dispersion Resolved reversal shares range from 11.78% at Snapchat to 70.65% at LinkedIn.

Control opportunity Prioritize category-sensitive human review where automation scale, decision severity and reversal outcomes converge.

Resolved internal complaints



Decision statement

Build an appeal-linked, category-sensitive quality loop. Increase human-review intensity where automated scale, action impact and reversal outcomes intersect; avoid blanket reductions in automation.

Interpretation and independence boundary

A reversal is not proof of an original error. It can reflect new evidence, changed context, policy interpretation or complaint eligibility. The methodology, formulas and weights were applied identically to all five platforms; no internal TikTok data or non-public knowledge was used.

Primary-source verification: 31/31 key metric checks passed before publication.

Source: Harmonised DSA transparency reports (H2 2025); calculations reproduced from the supplied machine-readable files.

02 / RISK ANALYTICS CASE STUDY

Business Context, Stakeholders & Questions

The analytical scope is designed for Platform Operations: identify exposure, define governance signals and connect findings to measurable controls.

Business context

The DSA harmonised reporting template creates a comparable public layer for enforcement actions, internal complaints, automated means, moderation resources and active-recipient metrics. The operational challenge is to translate regulatory disclosure into a practical QA and governance system without overstating what the data proves.

Stakeholder map

STAKEHOLDER	DECISION NEED	PRIMARY KPI
Executive / platform leadership	Set risk appetite and approve control priorities.	Review score; severe-action exposure.
Trust & Safety / Operations	Allocate queues, reviewers and escalation capacity.	Volume; SLA; reversal rate.
Policy / Model QA	Tune category thresholds and investigate defects.	Category precision; repeat defects.
Legal / Compliance	Validate reporting perimeter and DSA controls.	Completeness; reconciliation.
Analytics / BI	Operationalise monitoring and versioned definitions.	Refresh quality; metric drift.

Business questions

Q1 How concentrated is enforcement volume after normalization by active-recipient metrics?

Q2 Where is end-to-end automated enforcement most exposed?

Q3 How do resolved complaint outcomes vary across platforms?

Q4 Which categories and restriction types create the highest potential review impact?

Q5 Where should quality assurance and human review be prioritized first?

> **Scope boundary**
>
> The dataset supports descriptive benchmarking and operational prioritisation. It does not identify unique affected users, causal error rates or case-level action-to-appeal relationships.

03 / RISK ANALYTICS CASE STUDY

Dataset Overview & Data Dictionary

A harmonised, platform-period dataset covering enforcement, appeals, automation, restriction mix and active-recipient proxies.

PLATFORMS	CORE TABLES	REPORTING WINDOW	ANALYTICAL GRAINS
5	55	H2 2025	3
VLOP reports	11 per platform	Aligned period	Platform / category / action

Master data dictionary

FIELD	TYPE	DEFINITION	ROLE
platform	Text	Reported service	Key
reporting_period	Text	Six-month harmonised period	Key
active_users	Numeric	AMAR or labelled proxy	Normalization
enforcement_volume	Numeric	Own-initiative T&C measures	Scale
automated_detection_volume	Numeric	End-to-end automated measures	Automation
complaints_received	Numeric	Internal complaints submitted	Appeals
appeals_resolved	Numeric	Upheld + partial + reversed	Appeals
decisions_reversed	Numeric	Resolved decisions reversed	Outcome
decision_type	Category	Visibility, monetary, service, account	Impact
content_category	Category	Top-level harmonised DSA category	Risk
data_quality_flag	Text	Provider-specific limitation	Control

Hierarchy rule

STATEMENT_CATEGORY_* rows are top-level groups. **KEYWORD_*** rows are subcategories. Summing both levels double-counts the same measures; only top-level categories are used for share calculations.

Processed outputs: platform_summary.csv, category_master.csv, decision_type_master.csv, data_quality_log.csv and source_registry.csv.

04 / RISK ANALYTICS CASE STUDY

Data Cleaning & Harmonisation

A reproducible control framework preserves source meaning while making cross-platform KPIs comparable.

Analysis workflow

1. **Ingest** — 5 provider reports
2. **Profile** — schema + notes
3. **Harmonise** — fields + taxonomy
4. **Reconcile** — totals + appeals
5. **Analyse** — KPIs + score
6. **Translate** — controls + roadmap

Cleaning decisions and controls

CONTROL	APPLIED LOGIC	WHY IT MATTERS
Schema validation	Confirm 11 expected tables and one reporting period per platform.	Prevents silent source drift.
Numeric normalization	Remove separators; preserve blank vs explicit zero.	Avoids invented data.
Category hierarchy	Tag top-level versus keyword subcategory rows.	Prevents double counting.
AMAR treatment	Use provider value; YouTube signed-in proxy only.	Preserves denominator meaning.
Appeal reconciliation	Resolved = upheld + partial + reversed.	Surfaces timing gaps.
Source exceptions	Attach provider-specific caveats to outputs.	Protects comparability.

Data-quality outcome

CORE KPI COVERAGE
100%
Six primary metrics across five platforms

ACTION RECONCILIATION
97.2%-101.2%
Provider totals vs decision columns

KNOWN CAVEATS
13
Documented and preserved

Important treatment note

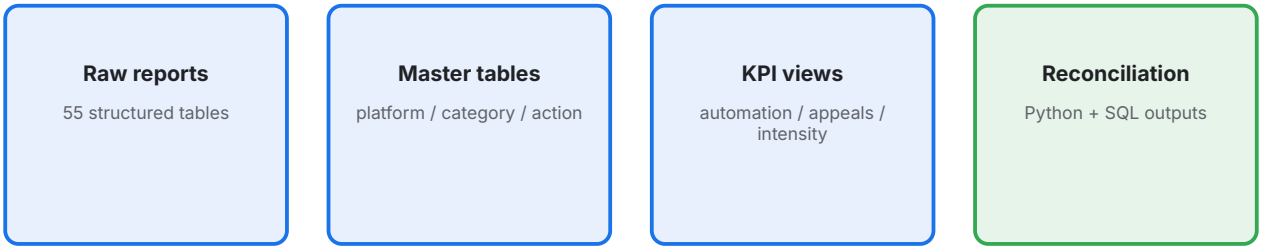
Blank cells are not automatically treated as zero. The provider context and total reconciliation determine whether a blank means zero, non-applicable or unavailable.

05 / RISK ANALYTICS CASE STUDY

SQL Methodology

SQL functions as a validation layer for KPI definitions, hierarchy controls and cross-platform reconciliation.

Validation architecture



Representative SQL - platform benchmark

```
SELECT
  platform,
  ROUND(100.0 * automated_detection_volume / enforcement_volume, 2) AS automation_pct,
  ROUND(100.0 * decisions_reversed / appeals_resolved, 2) AS reversal_pct,
  ROUND(1000000.0 * enforcement_volume / active_users, 0) AS measures_per_million
FROM platform_summary
ORDER BY measures_per_million DESC;
```

Validation output

PLATFORM	MEASURES	AUTOMATION	REVERSAL	PER 1M
TikTok	258.69M	85.08%	46.16%	1,450,878
Instagram	55.92M	80.46%	39.85%	193,658
YouTube	29.95M	97.95%	29.15%	62,477*
Snapchat	11.28M	78.21%	11.78%	116,099
LinkedIn	4.01M	87.34%	70.65%	72,590

Reconciliation passed

Python and SQL reproduce 359,845,873 measures, 24,770,949 complaints and 10,799,092 reversed decisions. *YouTube normalization uses 479.4M signed-in accounts as a non-additive proxy.

06 / RISK ANALYTICS CASE STUDY

Exploratory Analysis I - Scale & Intensity

Raw volume and active-recipient-normalized intensity reveal materially different operational footprints.

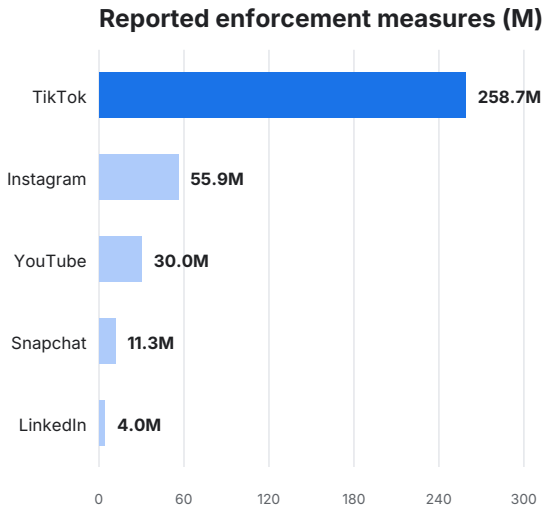


Figure 3. Reported volume.

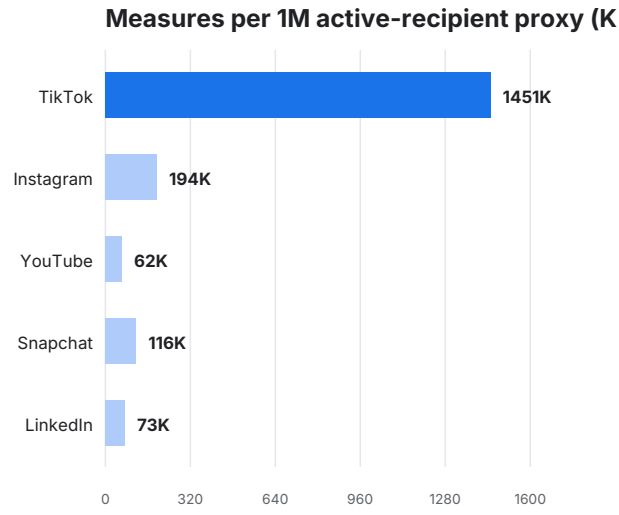


Figure 4. Normalized intensity.

Source: Harmonised DSA transparency reports (H2 2025); calculations reproduced from the supplied machine-readable files.

Observed patterns

PLATFORM	REPORTED MEASURES	SHARE OF SAMPLE	MEASURES PER 1M
TikTok	258.69M	71.89%	1,450,878
Instagram	55.92M	15.54%	193,658
YouTube	29.95M	8.32%	62,477*
Snapchat	11.28M	3.14%	116,099
LinkedIn	4.01M	1.11%	72,590

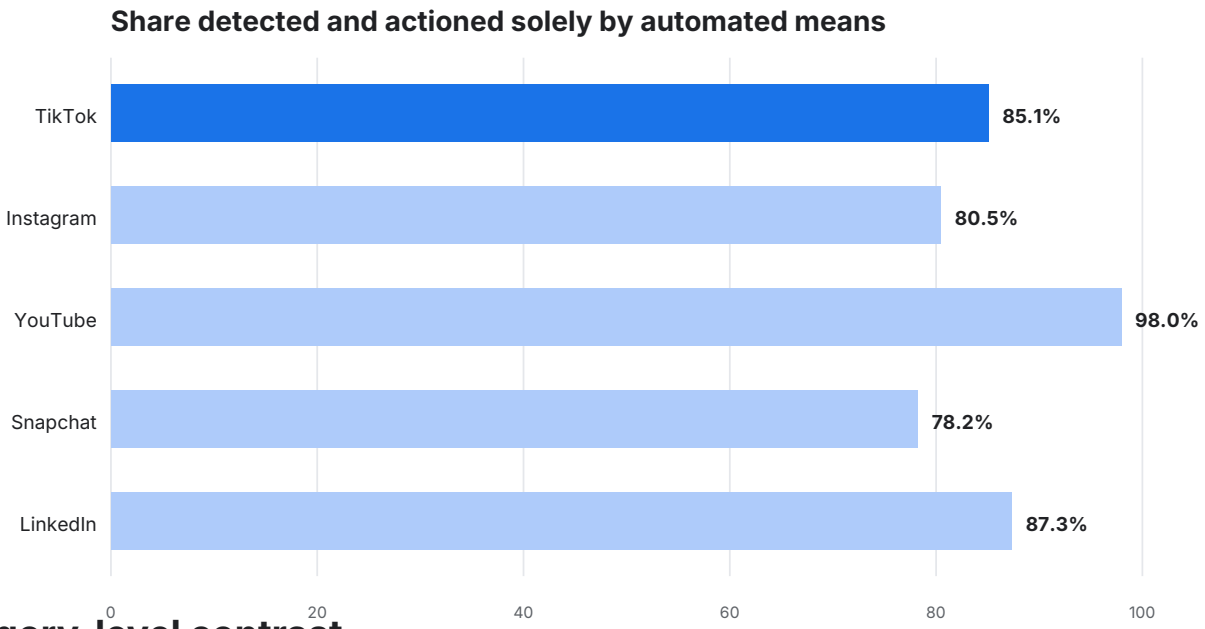
Interpretation

TikTok remains the highest signal by both raw volume and normalized intensity. The metric is enforcement intensity, not user incidence: one recipient or content item can generate multiple measures.

07 / RISK ANALYTICS CASE STUDY

Exploratory Analysis II - Automation

Aggregate automation is high across all five platforms, while category-level deployment remains highly differentiated.



Category-level contrast

PLATFORM / CATEGORY	MEASURES	AUTOMATED EXPOSURE	OPERATIONAL SIGNAL
Instagram - Civic discourse / elections	5.96M	2.45%	Predominantly non-automated
Instagram - Privacy violations	1.97M	3.32%	Context-sensitive review
LinkedIn - Cyber violence	23.5K	0.98%	Human-led handling
Snapchat - Cyber violence	618.8K	0.15%	Low end-to-end automation
YouTube - Protection of minors	11.16M	99.69%	High-scale automation

Operational hypothesis to test

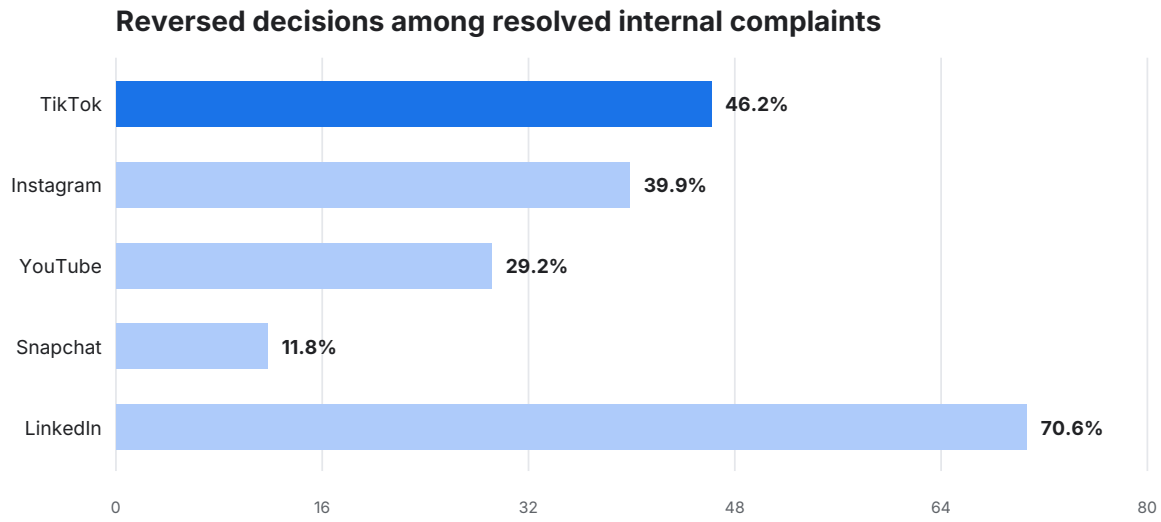
A platform-wide automation rate is too coarse for governance. Review thresholds should vary by category severity, ambiguity, evidence requirements and downstream action impact.

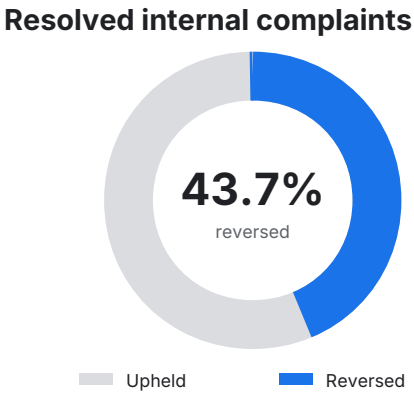
Source: Harmonised DSA transparency reports (H2 2025); calculations reproduced from the supplied machine-readable files.

08 / RISK ANALYTICS CASE STUDY

Exploratory Analysis III - Appeals

Resolved complaint outcomes show a wide cross-platform dispersion and a material quality-assurance workload.





Appeal readout

24.77M complaints were reported. 24.74M outcomes were resolved, of which 10.80M were reversed. LinkedIn has the highest reversal share among resolved complaints at 70.65%, followed by TikTok at 46.16% and Instagram at 39.85%.

Selection and timing effects

Complaint populations are self-selected. Instagram explicitly notes that complaints and restores can relate to decisions recorded in earlier reports. The resulting rates cannot be interpreted as platform-wide error rates.

Key quality signal

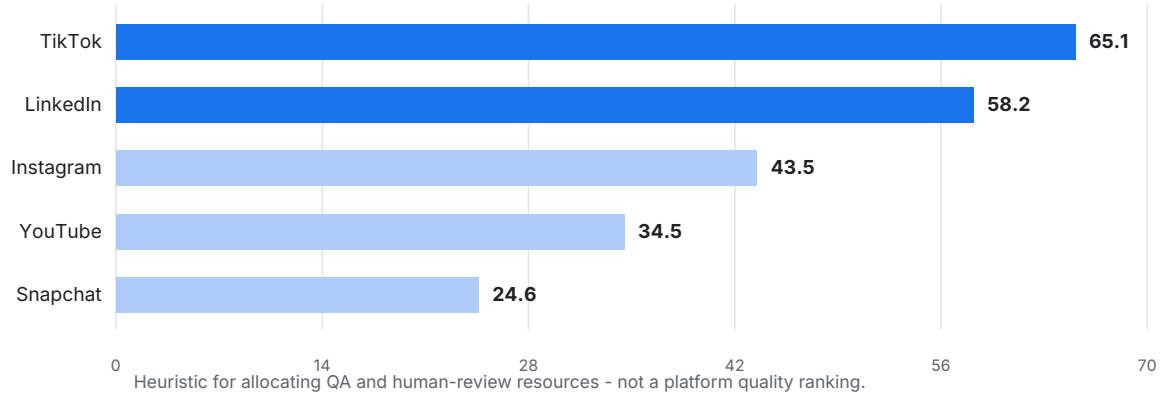
The variation is sufficient to prioritize investigation. It is not sufficient to rank platform fairness or decision accuracy without case-level, category-level and eligibility controls.

09 / RISK ANALYTICS CASE STUDY

Key Findings & Review Priority

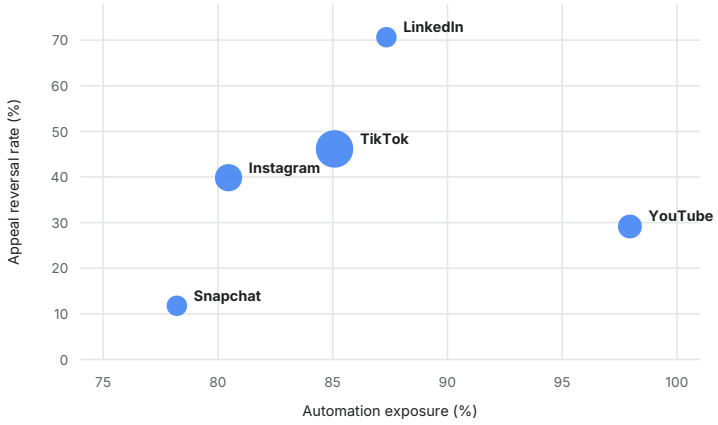
A transparent heuristic concentrates QA attention without claiming predictive or regulatory authority.

Enforcement Review Priority Score



Heuristic for allocating QA and human-review resources - not a platform quality ranking.

Operational review priority map



Four findings that should shape action

1

Scale-driven priority

TikTok combines dominant volume, high automation and a 46.16% reversal share.

2

Impact-driven priority

LinkedIn combines a 70.65% reversal share with a 92.03% account/service restriction mix.

3

Data-aware priority

Instagram requires targeted controls and explicit treatment of demotion and cross-service caveats.

4

Category priority

YouTube and Snapchat still require category-level QA despite lower aggregate scores.

Priority rule

Score = 30% normalized intensity + 25% automation exposure + 30% reversal rate + 15% severe-restriction mix. It ranks review need, not platform quality.

Source: Harmonised DSA transparency reports (H2 2025); calculations reproduced from the supplied machine-readable files.

10 / RISK ANALYTICS CASE STUDY

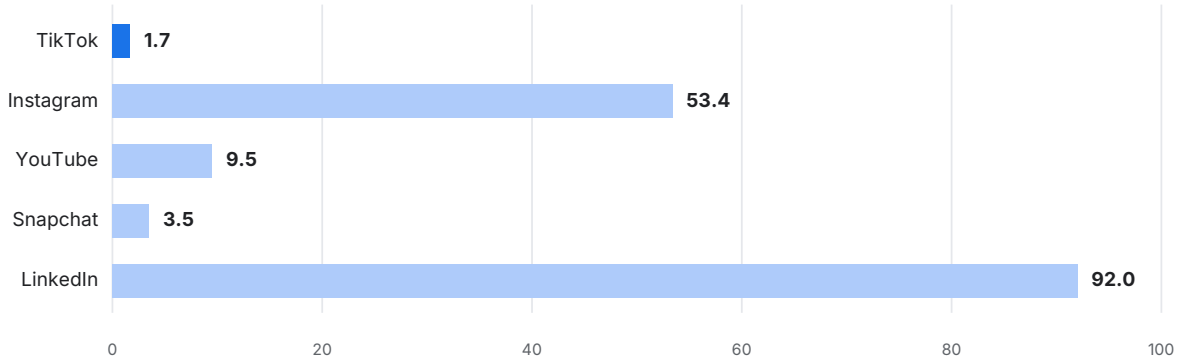
Category & Decision Mix

Enforcement is highly concentrated, but the dominant risk category and restriction type differ materially by platform.

Top-level enforcement category mix (% of reported measures)

	Minors	Scams / fraud	Cyber violence	Harmful speech	Other T&C
TikTok	33.2%	3.6%	18.5%	6.5%	23.7%
Instagram	1.0%	16.6%	0.4%	0.4%	58.2%
YouTube	37.3%	0.2%	23.4%	14.9%	5.8%
Snapchat	0.8%	0.5%	5.5%	0.6%	88.2%
LinkedIn	0.0%	92.2%	0.6%	1.3%	5.1%

Reported account, service and monetary restrictions



Interpretation

LinkedIn is overwhelmingly concentrated in scams/fraud and inauthentic-account enforcement. TikTok is concentrated in protection of minors, broad T&C violations and cyber violence. Broad catch-all categories dominate Instagram and Snapchat, limiting diagnostic precision.

Taxonomy guardrail

Top-level categories are mutually analysed; keyword subcategories are retained for drill-down only.

11 / RISK ANALYTICS CASE STUDY

Recommendation Matrix

Each recommendation is linked to an observed signal, an operational owner and a measurable control outcome.

ID	RECOMMENDATION	EVIDENCE	OWNER	OPERATIONAL DESIGN	SUCCESS KPI	PRIORITY
R1	Appeal-linked quality loop	10.80M reversals	QA / Policy	Link appeals to policy, model, action and cohort.	Reversal; repeat defect	High
R2	Risk-tiered human review	85.24% automation	T&S Ops	Raise review thresholds for severe actions.	Severe-action precision	High
R3	Category-specific thresholds	0.15%-99.69% category range	Model / Policy	Tune sampling and confidence by category.	Category precision / recall	High
R4	Reporting reconciliation	13 documented caveats	Data / Compliance	Schema, hierarchy and total controls.	Reconciliation pass rate	High
R5	Governance dashboard	Cross-platform fragmentation	Ops Analytics	Weekly ops + monthly risk review.	SLA; drift; closure	Medium

Measurement guardrails

GUARDRAIL	REQUIRED CONTROL
Appeal linkage	Separate complaint eligibility, new evidence and policy reinterpretation from confirmed defects.
Severity calibration	Weight account, service and monetary restrictions above low-impact visibility actions.
Cost accounting	Track reviewer time, queue latency and operational cost alongside quality gains.
Fairness and consistency	Monitor policy and language cohorts; investigate systematic divergence.
Stop criteria	Pause controls that create excessive latency without measurable reliability gains.

Recommended starting point

Launch R1, R2 and R4 together. R1 identifies recurring defects, R2 protects high-impact decisions, and R4 ensures the governance dashboard is based on reconciled data.

12 / RISK ANALYTICS CASE STUDY

Monitoring Framework & 90-Day Roadmap

The operating model connects public reporting, quality controls, investigation queues and executive governance.

Three-layer monitoring framework

INPUT LAYER DSA reports | appeals | automation | AMAR | taxonomy

CONTROL LAYER Schema | hierarchy | reconciliation | versioning | exceptions

REVIEW LAYER Weekly operations | monthly risk | quarterly governance

90-day execution sequence

TIMING	PHASE	DELIVERABLE
Days 0-30	Instrument	Publish definitions; map appeals; establish severity weights; implement data-quality tests.
Days 31-60	Pilot	Risk-based samples; category thresholds; root-cause reversed severe actions.
Days 61-90	Evaluate & scale	Measure quality lift, queue cost and SLA; scale only positive controls.

Decision cadence

CADENCE	FOCUS	CORE METRICS
Weekly	Execution health	Volume, queue SLA, appeal backlog, data exceptions
Monthly	Risk governance	Automation drift, reversal, severe actions, corrective actions
Quarterly	Methodology	Taxonomy changes, audit findings, thresholds, source versions

> **Operating principle**
> Escalate based on evidence, impact and repeatability - not raw volume alone.

13 / RISK ANALYTICS CASE STUDY

Risks, Limitations & Learning

The analysis is useful for prioritisation, but its constraints determine which claims are safe and which require new evidence.

Risks and limitations

LIMITATION	IMPLICATION	REQUIRED NEXT STEP
No case-level action-to-appeal key	Cannot estimate a true decision error rate.	Add event-level linkage and outcome reason.
Self-selected complaints	Appeal population is not representative.	Model eligibility and propensity to appeal.
Measures are not unique users	Intensity can exceed one action per recipient.	Add unique content / account denominators.
YouTube AMAR methodology	Signed-in and signed-out counts are non-additive.	Keep proxy flag and sensitivity analysis.
Instagram logging caveats	Demotion undercount; some combined Meta fields.	Exclude affected fields from precise ranking.
Snap automation scope	Tab 8 includes approvals and differs from Tab 6.	Use Tab 6 for cross-platform exposure.
Provider-defined taxonomies	Same label can represent different product scopes.	Maintain source-specific mapping and notes.
Prior-employer familiarity	TikTok familiarity could be perceived as interpretive bias.	Use identical formulas and weights; no internal data or non-public knowledge.

Lessons learned

LESSON	APPLICATION
Harmonised does not mean identical	Preserve provider caveats in every comparison.
A reversal is an investigation trigger	Do not equate it with a confirmed original error.
Category granularity matters	Platform-wide automation rates hide human-review design.
Transparent heuristics are useful	A simple score can allocate resources without overstating model sophistication.
Validation improves trust	Reconcile Python, SQL and source totals before presentation.

> **Future improvement**
>
> Extend the project with DSA Transparency Database statements of reasons, event-level samples and longitudinal reporting periods.

14 / RISK ANALYTICS CASE STUDY

Technical Appendix

KPI definitions, reproducibility controls and the analytical boundary used in this case study.

KPI definitions

KPI	FORMULA	VERIFIED RESULT
Reported measures	SUM(enforcement_volume)	359,845,873
Automation exposure	SUM(automated) / SUM(measures)	85.24%
Resolved reversal rate	SUM(reversed) / SUM(resolved)	43.66%
Measures per 1M	measures / active proxy x 1M	62,477 to 1,450,878
Severe restriction mix	severe columns / measures	1.69% to 92.03%
Review Priority Score	30% intensity + 25% automation + 30% reversal + 15% severity	24.6 to 65.1

Reproducibility checklist

CONTROL	RESULT
Input inventory	5 providers; 55 core tables; aligned H2 2025 period
Hierarchy integrity	Top-level categories separated from keyword subcategories
Appeal reconciliation	24,736,361 resolved = upheld + partial + reversed
Primary-source verification	31/31 report-to-master checks passed
Cross-tool validation	Python extraction and SQL definitions reconcile
Lineage	Source file, official URL and treatment retained
Artifacts	CSV masters, XLSX workbook, SQL, Python, PDF and PPTX

Primary source set

European Commission harmonised template; TikTok, Meta/Instagram, Google/YouTube, LinkedIn and Snap transparency reports published in February-April 2026. Key totals were traced directly to the primary report rows before publication.

> **Repository note**
> All processed outputs retain source lineage and provider-specific treatment notes.

15 / RISK ANALYTICS CASE STUDY

Executive Takeaways

Five conclusions for a decision-maker reviewing platform enforcement reliability.

01	Automation is the operating default: 85.24% of the selected sample was reported as end-to-end automated under Tab 6.
02	Scale and risk are not the same: TikTok dominates volume, while LinkedIn combines the highest reversal rate with account-level impact.
03	Appeals are a QA signal, not an error label: 10.80M reversals justify investigation but do not prove 10.80M original mistakes.
04	Category-specific governance is essential: automation exposure ranges from near zero to nearly 100% depending on policy category.
05	The best next step is an operational control loop linking appeals, policy categories, models and restriction severity to targeted human review.

Thank you for reading.

Prepared by

Rayan Braccio

Business Operations & Data Analytics Portfolio

rayanbraccio.com | github.com/rayanbraccio

Final decision statement

Use the Enforcement Review Priority Score as an orchestration layer for investigation and resource allocation. Validate the score with event-level data before using it for production decisions.