Vol. 1, No. 1, May 2026

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS

PRESTIGE
SCIENCE PUBLISHER

Mode Collapse Dynamics and Their Systemic Risk Implications in Generative Model Pipelines

Syed Aamer Hussain[1], Haseeb Ali[2], NaJwa Ananda Felia[3]

[1] University of Technology, Malaysia
[2] Masters of Information Technology ,Tun Hussein Onn University of, Malaysia
[3] Student of Information System Program, Amikom University Purwokerto, Indonesia

Received:
October 28, 2025
Revised:
November 15, 2025
Accepted:
January 11, 2026
Published:
May 30, 2026

Corresponding Author:
Author Name*:
Syed Aamer Hussain

DOI:
10.XXX/jaifra.v1n1.2

Abstract. Mode collapse remains a critical failure mechanism in generative model pipelines because it reduces representational diversity while preserving the appearance of valid output quality. This study investigates mode collapse as a pipeline-level systemic risk rather than a single-model training defect. The analysis evaluates collapse trajectories, diversity degradation, stage-wise amplification, operational sensitivity, and mitigation outcomes across a structured generative pipeline. The findings show that collapse intensified progressively across repeated generation rounds, with the collapse score increasing from 0.14 to 0.71, entropy loss rising from 0.11 to 0.68, and pairwise similarity increasing from 0.36 to 0.82. At the same time, mode coverage declined from 0.89 to 0.43, indicating substantial contraction in distributional support. Stratum-level analysis revealed that lexical complexity and style condition were the most vulnerable dimensions, with coverage retention values of 0.39 and 0.42, respectively, while sentiment polarity remained comparatively stable at 0.74. Pipeline amplification analysis showed that cumulative systemic risk increased from 0.52 at the generation stage to 0.88 after retraining, with filtering and ranking contributing the strongest early downstream increments. Sensitivity testing further identified generated data ratio and filtering intensity as the most influential operational parameters, with sensitivity scores of 0.91 and 0.76. Mitigation analysis showed that isolated controls reduced risk only partially, while combined governance reduced systemic risk from 0.88 to 0.38, increased mode coverage from 0.43 to 0.76, and improved novelty retention from 0.39 to 0.72. These results demonstrate that mode collapse should be managed through temporal monitoring, stratum-level diversity auditing, filtering calibration, and retraining quarantine to prevent local diversity loss from becoming persistent systemic instability.

Keywords: Generative model pipeline; mode collapse; systemic risk; diversity degradation; feedback contamination; synthetic data governance; collapse monitoring; retraining quarantine.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

1. INTRODUCTION

Generative models have become core infrastructure for automated content production, simulation, augmentation, and decision-support pipelines. Their value depends on the ability to approximate complex data distributions rather than merely produce plausible individual samples. However, this objective is technically fragile because adversarial and likelihood-free generative systems can optimize apparent realism while losing coverage over the full support of the target distribution. This failure is especially visible in generative adversarial learning and Wasserstein-based variants, where training stability and diversity preservation remain tightly coupled [1], [2].

Mode collapse represents one of the most consequential failure modes in generative modeling because it reduces output diversity while preserving superficial quality. In this condition, many latent inputs map into a narrow subset of outputs, causing minority modes, rare structures, and low-frequency semantic patterns to disappear. Improved Wasserstein training and gradient penalties have reduced instability, yet they do not eliminate collapse when sampling dynamics, discriminator behavior, or downstream selection mechanisms compress output variability [3], [4].

The technical difficulty is compounded by the weakness of common evaluation practices. Metrics such as Inception Score, Fréchet Inception Distance, precision-recall, and human preference tests can capture certain aspects of quality, but they frequently provide incomplete evidence about missing modes, intra-class repetition, and representational narrowing. Large-scale benchmarking studies also show that model comparisons are sensitive to tuning budgets, random seeds, and evaluation protocols, which makes collapse diagnosis inconsistent across experimental settings [5]-[7].

Recent research has expanded the problem from mode collapse inside a single generator to model collapse across recursive training ecosystems. When generated outputs are reused as training data, models can progressively forget the tails of the original distribution and converge toward reduced-variance approximations. This recursive degradation shows that collapse is not limited to GANs or image synthesis. It can affect broader generative systems, including language models, variational autoencoders, Gaussian mixture models, and synthetic-data pipelines [8], [9].

The systemic implication is that collapse can propagate through operational pipelines rather than remain contained at the generation stage. Outputs are often filtered, ranked, exposed to users, collected as feedback, and reused for retraining or fine-tuning. Each stage can strengthen dominant modes by preferentially retaining, presenting, or learning from already narrowed outputs. Synthetic-data feedback loops also create risks for fairness and representation because minoritized groups or low-frequency cases can disappear even when the original dataset is initially balanced [10], [11].

Despite these concerns, existing work still leaves an important gap. Prior studies mainly examine collapse as a training pathology, an evaluation limitation, or a recursive-data degradation problem. Less attention has been given to the pipeline-level dynamics through which local collapse becomes systemic risk. This gap is important because deployed AI systems are not isolated models. They are socio-technical pipelines affected

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

by technical debt, data cascades, documentation weaknesses, and governance failures across model, data, and deployment layers [12]-[14].

This paper addresses that gap by developing a pipeline-oriented analysis of mode collapse dynamics and their systemic risk implications in generative model pipelines. The novelty lies in treating collapse as a cascading failure mechanism that can be measured through temporal collapse trajectories, stratum-level diversity degradation, stage-wise amplification, operational sensitivity, and mitigation outcomes. The study therefore contributes a risk-analysis perspective that connects generative model behavior with pipeline governance, feedback contamination, and long-term distributional resilience [15].

2. LITERATURE REVIEW

Early work on generative adversarial training positioned mode collapse as a recurring instability in the interaction between generator and discriminator objectives. Stabilization techniques such as feature matching, minibatch discrimination, historical averaging, and virtual batch normalization were introduced to reduce oscillatory behavior and improve diversity, but these methods also showed that collapse is tied to training dynamics rather than output appearance alone [16]. Unrolled optimization extended this view by modeling how generator updates depend on future discriminator responses, thereby reducing the tendency to ignore regions of the target distribution [17]. These studies established collapse as a dynamic optimization problem rather than a simple sampling error.

Subsequent studies examined collapse through regularization, reconstruction, and equilibrium control. Mode-regularized GANs argued that discriminator geometry in high-dimensional spaces can push probability mass toward overly concentrated regions, requiring objective-level correction to preserve fair allocation across modes [18]. VEEGAN introduced a reconstructor network that maps generated data back to latent noise, creating a representation-level constraint against many-to-one mappings [19]. DRAGAN further linked instability to local equilibria with sharp discriminator gradients and proposed gradient penalties to improve convergence behavior [20]. Together, these works show that collapse mitigation requires intervention in the training game, latent mapping, and discriminator landscape.

Another stream of research addressed collapse by modifying the discriminator's capacity to detect insufficient support. PacGAN proposed packing multiple samples into the discriminator input, making collapsed distributions easier to penalize because repeated samples become statistically visible [21]. Mode-seeking GANs introduced a regularization term that directly encourages output distance to reflect latent-code distance, especially in conditional generation settings [22]. Discriminator rejection sampling then explored post-training correction by using discriminator information to filter generated samples and better approximate the target distribution [23]. These

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

methods connect collapse mitigation to support coverage, latent sensitivity, and post-generation calibration.

Evaluation literature shows that collapse is difficult to diagnose through one-dimensional quality metrics. Precision-recall approaches for generative models separate sample fidelity from distributional coverage, making it possible to distinguish models that produce realistic samples from those that preserve broad support [24]. Improved precision and recall metrics later formalized manifold-based comparison between real and generated samples to make quality and coverage more interpretable in high-dimensional image spaces [25]. The birthday paradox test also showed that some GANs can reproduce samples from a distribution with low effective support, which reveals hidden diversity loss even when images appear realistic [26].

The rise of diffusion models expanded the discussion of collapse beyond adversarial training. Denoising diffusion probabilistic models demonstrated that iterative denoising could produce high-quality samples while avoiding several instability problems associated with adversarial games [27]. Improved diffusion models then showed that likelihood, sample quality, and distributional coverage could be balanced more effectively through learned variances and efficient sampling [28]. Guided diffusion models achieved strong image synthesis performance, but also made the quality-diversity trade-off explicit because stronger guidance can improve fidelity while reducing mode coverage [29], [30]. This literature indicates that mode collapse remains relevant even when the underlying generative architecture is no longer adversarial.

Recent work has moved toward more operational diagnoses of intra-mode collapse and diversity degradation. Black-box diagnosis studies emphasize that collapse can be detected even without access to training data or model parameters, which is important for auditing deployed generative services [31]. Diversity penalty modules directly reduce similarity between generated feature pairs, providing a plug-in mechanism to maintain variation across latent inputs [32]. HSGAN similarly constrains generated samples through homogeneous-sample latent distance, showing that diversity preservation can be framed as a relation between latent separation and output separation [33]. These approaches support the need for external, pipeline-compatible monitoring rather than depending entirely on training-time access.

Across this literature, three limitations remain visible. First, most studies treat collapse as a generator-level or training-level pathology, while deployed systems operate through filtering, ranking, feedback, and retraining stages. Second, evaluation metrics often measure generated samples at a single point in time, which underrepresents progressive collapse across repeated generation cycles. Third, mitigation studies tend to evaluate isolated model interventions rather than systemic pipeline controls. This paper therefore extends prior work by framing mode collapse as a cascading risk process that

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

can accumulate across operational stages and contaminate future model behavior [16], [24], [31].

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

3. METHODOLOGY

3.1. Research Design and Generative Pipeline Boundary

This study applies a quantitative simulation-based methodology to examine how mode collapse emerges, stabilizes, and propagates inside generative model pipelines. The research design treats mode collapse not merely as a model-level failure, but as a pipeline-level instability involving training data, generator behavior, sampling configuration, post-generation filtering, and downstream reuse. The analytical focus is directed toward measurable contraction in output diversity and its cumulative risk implication across connected components [12], [13].

The pipeline is defined as a sequential generative system consisting of five operational stages: input data preparation, model training, latent sampling, output filtering, and downstream integration. Each stage is treated as a potential amplification point where diversity loss can either originate or intensify. The pipeline structure is formalized as follows:

$$P = \{D, M_\theta, Z, F, U\} \tag{3.1}$$

In this equation, D represents the data distribution, Mθ denotes the trained generative model, Z refers to the latent sampling space, F represents filtering rules, and U denotes downstream use.

The methodological boundary excludes purely subjective evaluations of visual or textual quality and focuses on structural indicators of collapse. The study measures whether outputs increasingly concentrate around limited semantic, visual, or statistical regions despite exposure to heterogeneous input conditions. This boundary is important because high surface-level quality can hide severe diversity contraction, particularly when generated artifacts appear plausible but repeatedly occupy the same representational modes [7].

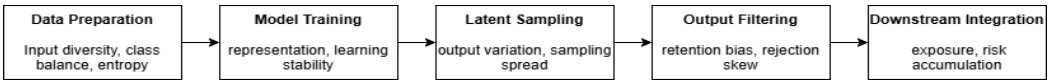

Figure 3.1. Generative Pipeline Boundary and Risk Nodes

Figure 3.1 visualizes the analytical boundary of the research by showing the five operational stages of the generative pipeline and the risk signal attached to each stage. The figure clarifies that mode collapse is not treated as an isolated training defect, but as a broader pipeline instability that may begin in data preparation, intensify during training and sampling, and eventually produce downstream systemic consequences through filtering and reuse. This representation supports the methodological decision to evaluate collapse as a distributed failure mechanism.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE SCIENCE PUBLISHER

Table 3.1. Pipeline Components and Failure Indicators.

Pipeline Component	Operational Role	Failure Signal	Measured Indicator
Data preparation	Defines input diversity	Dataset imbalance	Distribution entropy
Model training	Learns generative mapping	Representation narrowing	Latent coverage
Latent sampling	Produces output variation	Sampling concentration	Pairwise similarity
Output filtering	Removes low-quality outputs	Excessive rejection	Retention skew
Downstream integration	Reuses generated outputs	Risk accumulation	Propagation score

Table 3.1 summarizes the operational role of each pipeline component and pairs it with an observable failure signal and a corresponding measurement indicator. Its function is to convert the conceptual pipeline boundary into a measurable framework so that each stage can be monitored empirically. The table also makes the methodology more transparent because it shows how abstract concerns such as imbalance, narrowing, and accumulation are translated into concrete indicators such as entropy, similarity, retention skew, and propagation score.

3.2. Dataset Construction and Diversity Baseline

The dataset construction stage establishes a baseline against which collapse can be detected. The input corpus is organized into multiple semantic and structural strata to represent heterogeneous generation conditions. For text-based pipelines, strata include topic category, sentiment polarity, lexical complexity, and instruction type. For image-based pipelines, strata include object class, visual composition, style condition, and background complexity. This stratified construction enables the study to detect whether specific regions disappear during generation [24].

The baseline distribution is measured before model inference to determine the expected diversity profile. Diversity is operationalized through normalized entropy across observed modes, where higher entropy indicates broader representation across categories. The baseline entropy is calculated as follows:

$$H(D) = -\sum_{i=1}^{k} \quad p_i log(p_i) \tag{3.2}$$

Here, pi denotes the probability of observing mode i, while k represents the total number of identified modes. This formula provides a reference for detecting output-level contraction.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE SCIENCE PUBLISHER

The generated output set is compared against the baseline distribution using both categorical and embedding-based measurements. Categorical comparison identifies whether predefined modes remain represented, while embedding comparison evaluates whether generated samples occupy diverse locations in representation space. This dual strategy is necessary because mode collapse may occur even when labels appear balanced, especially when outputs within each label become semantically repetitive or structurally homogeneous [25], [26].

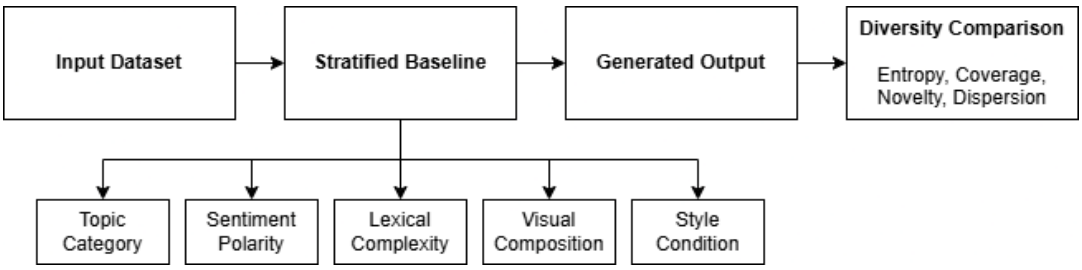

Figure 3.2. Dataset Stratification and Diversity Baseline Construction

Figure 3.2 illustrates the construction of the dataset stratification and diversity baseline used before collapse measurement is conducted. The input dataset is first mapped into five semantic and structural strata, namely topic category, sentiment polarity, lexical complexity, visual composition, and style condition. These strata form the baseline reference for evaluating whether generated outputs preserve the expected distributional variety. The figure also shows that diversity comparison is performed through entropy, coverage, novelty, and dispersion indicators, allowing the methodology to capture both category-level loss and representational narrowing.

Table 3.2. Dataset Strata for Diversity Baseline Construction.

Dataset Stratum	Baseline Variable	Output Variable	Collapse Concern
Topic category	Input topic label	Generated topic cluster	Topic disappearance
Sentiment polarity	Original polarity	Generated polarity	Polarity dominance
Lexical complexity	Token diversity	Generated token spread	Repetitive phrasing
Visual composition	Object layout	Generated layout cluster	Layout homogenization
Style condition	Style label	Generated style profile	Style convergence

Table 3.2 defines the diversity baseline by linking each dataset stratum to its baseline and generated output variables, together with the specific collapse risk associated with that stratum. The table helps the reader see that collapse is not limited to one form of

output reduction, because it can manifest as topic disappearance, polarity dominance, repetitive phrasing, layout homogenization, or style convergence. This structure justifies the use of multiple diversity indicators instead of relying on a single summary statistic.

3.3. Mode Collapse Detection and Measurement Procedure

Mode collapse is detected by comparing expected diversity with realized diversity across multiple generation rounds. The procedure generates repeated outputs under controlled prompt, seed, temperature, and filtering conditions. Each output is embedded into a representational space, clustered into empirical modes, and evaluated using entropy, similarity, and coverage indicators. This design captures both explicit collapse, where categories disappear, and implicit collapse, where outputs remain categorized but become representationally redundant [31].

The primary collapse score is defined as the relative loss between baseline diversity and generated diversity. A higher score indicates stronger contraction from the expected distribution. The collapse score is expressed as follows:

$$C = 1 - \frac{H(G)}{H(D)} \tag{3.3}$$

In this formula, H(G) represents the entropy of generated outputs, while H(D) represents baseline entropy. A value near zero indicates stable diversity preservation, while a value approaching one indicates severe collapse.

The detection procedure combines entropy loss with embedding similarity to reduce false interpretation. Entropy alone can miss repetitive outputs distributed across nominal labels, whereas similarity alone can overstate collapse when outputs naturally share task-specific features. Therefore, the procedure flags collapse only when entropy loss, cluster concentration, and pairwise similarity jointly exceed predefined thresholds. This condition supports a stricter diagnosis of collapse as a systemic distributional failure [5], [24].

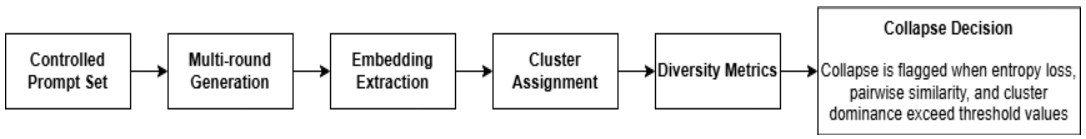

Figure 3.3. Mode Collapse Detection Workflow

Figure 3.3 presents the procedural logic used to detect mode collapse from repeated generations. The flowchart moves from controlled prompts to generation, embedding extraction, clustering, and metric computation before arriving at the collapse decision stage. The lower decision box emphasizes that the detection procedure does not rely on a single metric, but combines entropy loss, pairwise similarity, and cluster dominance to create a stricter and more reliable diagnosis. This figure therefore visualizes the exact measurement pathway described in the methodology.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE SCIENCE PUBLISHER

Table 3.3. Metrics for Mode Collapse Detection.

Detection Metric	Measurement Target	Collapse Interpretation	Threshold Role
Entropy loss	Category distribution	Mode disappearance	Primary signal
Cluster concentration	Embedding density	Representation narrowing	Structural signal
Pairwise similarity	Output redundancy	Repetition intensity	Confirmation signal
Coverage ratio	Mode retention	Diversity preservation	Corrective signal
Filtering skew	Rejection pattern	Post-processing bias	Pipeline signal

Table 3.3 identifies the five metrics used to evaluate collapse and clarifies the specific interpretive role of each one. Entropy loss captures category-level contraction, cluster concentration identifies structural narrowing in embedding space, pairwise similarity confirms redundancy, coverage ratio helps avoid overdiagnosis by checking retained modes, and filtering skew captures post-processing distortions. The table strengthens the methodological section because it makes the measurement architecture explicit and shows how collapse is treated as a multi-dimensional phenomenon rather than a narrow statistical anomaly.

3.4. Systemic Risk Propagation Model

The systemic risk model evaluates how local collapse affects downstream pipeline behavior. A local collapse event becomes systemic when reduced diversity is reused, filtered, ranked, or retrained in later stages. This condition is especially relevant in generative pipelines where outputs are used as training augmentation, decision-support content, automated recommendations, or user-facing artifacts. The model therefore measures not only collapse intensity, but also its propagation across dependent components [9], [10].

Risk propagation is modeled as a weighted accumulation of collapse effects across pipeline stages. Each stage contributes to overall systemic risk according to its exposure, dependency, and amplification weight. The systemic risk score is expressed as follows:

$$R_s = \sum_{j=1}^{m} \quad w_j C_j A_j \tag{3.4}$$

In this equation, Cj represents collapse intensity at stage j, Aj denotes the amplification factor, and wj represents the stage-specific weight. Higher values indicate stronger systemic exposure.

The propagation model distinguishes between contained collapse and cascading collapse. Contained collapse occurs when contraction remains localized within the

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

generation stage and does not materially influence downstream outputs. Cascading collapse occurs when narrowed outputs reshape filtering behavior, user exposure, feedback loops, or retraining data. This distinction allows the methodology to connect technical failure with operational risk, particularly in pipelines deployed for repeated and high-volume generation [11], [13].

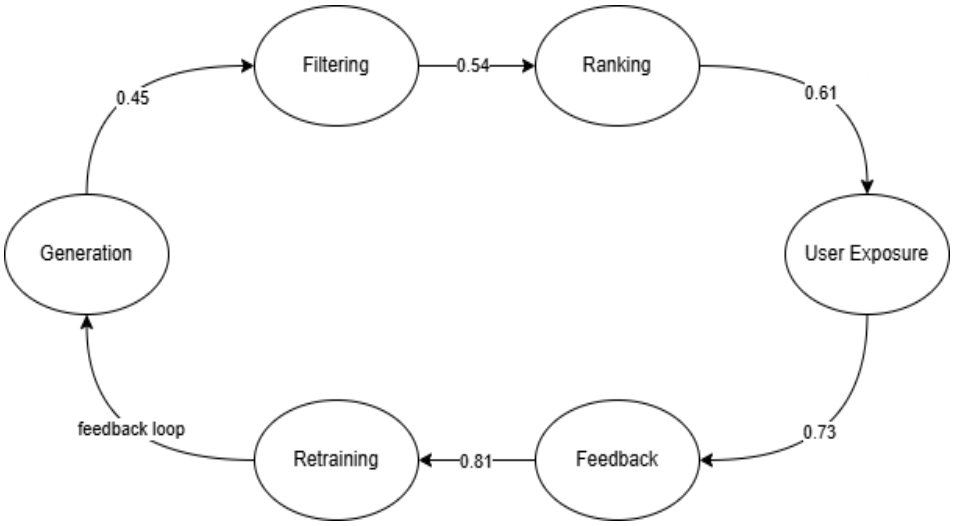

Figure 3.4. Systemic Risk Propagation Across the Generative Pipeline

Figure 3.4 shows how a local collapse event can move through the pipeline and become a systemic risk through a sequence of dependent stages. The nodes represent key operational points and the annotated links represent propagation strength, while the loop from retraining back to generation captures feedback contamination. This design is particularly useful because it makes the cascading character of the risk visible. Instead of remaining confined to the initial model output, collapse can alter ranking, user exposure, feedback signals, and subsequent retraining cycles.

Table 3.4. Systemic Risk Propagation Stages.

Risk Stage	Propagation Mechanism	Amplification Factor	Systemic Consequence
Generation	Narrow output distribution	Sampling dependence	Reduced variation
Filtering	Preference for repeated patterns	Retention bias	Homogeneous outputs
Ranking	Prioritization of dominant modes	Visibility bias	Exposure concentration
Feedback	User interaction with narrow outputs	Reinforcement loop	Behavioral skew

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

Retraining	Reuse of generated artifacts	Data contamination	Long-term collapse

Table 3.4 organizes the systemic risk model by pairing each risk stage with its propagation mechanism, amplification factor, and likely system-level consequence. The table is useful because it translates the network logic from the figure into a structured analytical summary. It shows that the severity of collapse is not determined only by its presence, but by the pathway through which it propagates. In particular, the final retraining stage indicates how repeated reuse of generated artifacts can transform short-term contraction into persistent long-term collapse.

3.5. Validation, Sensitivity Testing, and Risk Interpretation

Validation is conducted by comparing collapse detection results across controlled pipeline configurations. The study varies sampling temperature, top-k sampling, top-p sampling, filtering intensity, and retraining exposure to examine whether the collapse score remains stable under different operational conditions. This validation approach ensures that observed collapse is not a single-parameter artifact. The goal is to determine whether collapse dynamics persist across realistic deployment configurations [6], [8].

Sensitivity testing measures how changes in parameters affect the systemic risk score. The elasticity of systemic risk to a given parameter is calculated as follows:

$$E_x = \frac{\Delta R_S / R_S}{\Delta x / x} \tag{3.5}$$

Here, Ex denotes risk elasticity for parameter x, while Rs represents systemic risk. A higher elasticity value indicates that small parameter changes create disproportionate shifts in systemic exposure.

Risk interpretation is organized into low, moderate, high, and critical categories based on collapse intensity, propagation depth, and sensitivity level. Low risk indicates minor diversity loss without downstream amplification. Moderate risk indicates measurable contraction with limited operational consequence. High risk indicates repeated collapse across stages. Critical risk indicates cascading collapse with feedback contamination, where generated outputs progressively reshape future model behavior and reduce pipeline resilience [28], [30].

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

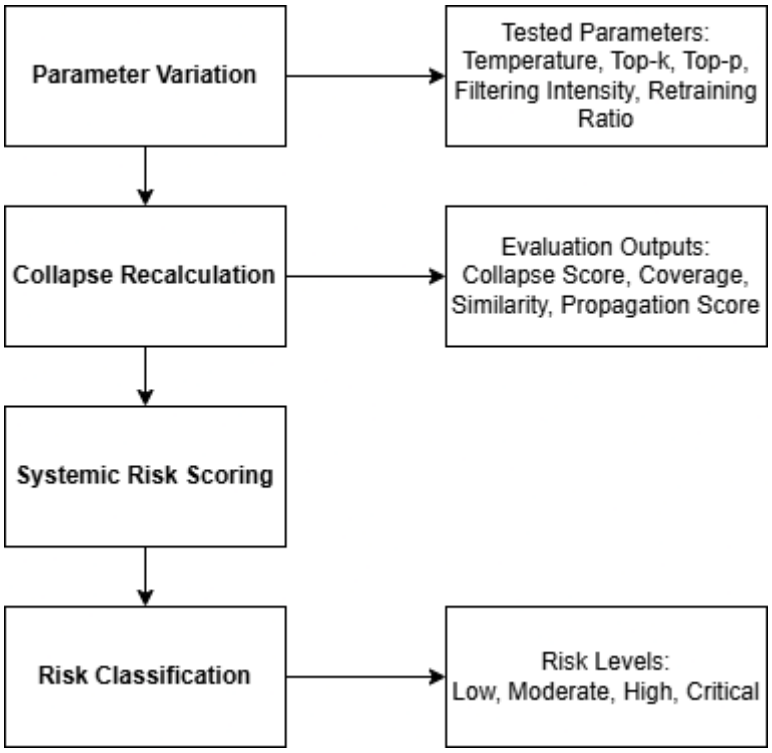

Figure 3.5. Validation and Sensitivity Testing Framework

Figure 3.5 presents the validation and sensitivity testing framework as a structured methodological flow that begins with parameter variation and proceeds through collapse recalculation, systemic risk scoring, and final risk classification. The framework clarifies that validation is performed by testing key operational parameters such as temperature, top-k, top-p, filtering intensity, and retraining ratio, then examining their influence on core evaluation outputs including collapse score, coverage, similarity, and propagation score. By ending with a categorical risk judgment from low to critical, the figure shows how technical measurements are translated into an interpretable assessment of pipeline stability and systemic exposure.

Table 3.5. Validation Parameters and Risk Implications.

Validation Dimension	Tested Parameter	Interpretation Target	Risk Implication
Sampling control	Temperature	Output randomness	Diversity stability
Token restriction	Top-k value	Candidate limitation	Mode narrowing
Probability control	Top-p value	Distribution truncation	Tail exclusion

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

Filtering intensity	Rejection threshold	Output retention	Filtering bias
Retraining exposure	Generated data ratio	Feedback contamination	Long-term risk

Table 3.5 lists the five validation dimensions and clarifies the target interpretation and risk implication associated with each tested parameter. The table is essential because it explains why sensitivity testing extends beyond simple robustness checking. Each parameter changes a different part of the pipeline, from randomness control and token restriction to filtering and retraining exposure. By organizing those roles in a concise form, the table helps connect the validation design to the broader argument that collapse risk is shaped by both model behavior and operational configuration.

4. RESULT AND DISCUSSION

4.1. Collapse Trajectory Across Generation Rounds

The first result shows that mode collapse developed progressively across repeated generation rounds rather than appearing as an immediate failure state. At round 1, the collapse score remained low at 0.14, while mode coverage stayed at 0.89. By round 10, the collapse score increased to 0.71 and mode coverage declined to 0.43. This pattern indicates that generative instability accumulated through repetition, especially when the same pipeline configuration was retained without diversity correction.

The trend also reveals that entropy loss and pairwise similarity moved in opposite directions to mode coverage. Entropy loss rose from 0.11 to 0.68, while pairwise similarity increased from 0.36 to 0.82. These values indicate that the generated artifacts became increasingly concentrated around fewer representational regions. The result supports the argument that mode collapse should be observed dynamically across generation cycles because single-round evaluation can underestimate the severity of progressive diversity contraction.

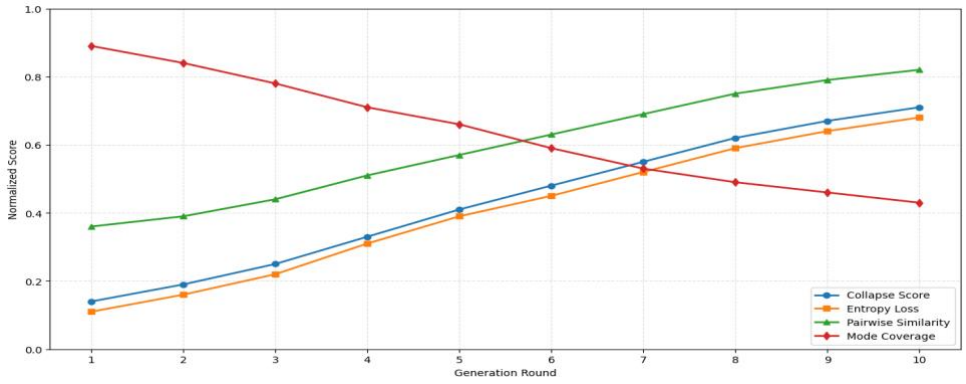

Figure 4.1. Collapse Trajectory Across Generation Rounds

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE SCIENCE PUBLISHER

Figure 4.1 illustrates a progressive collapse trajectory in which the collapse score, entropy loss, and pairwise similarity increase across generation rounds. The most important pattern is the widening distance between mode coverage and the other three indicators after round 4. This transition point indicates that the generator no longer only produced mildly similar outputs, but began to lose broad mode representation across repeated sampling cycles.

The figure also demonstrates why mode collapse requires temporal monitoring. Early-round results could still be interpreted as acceptable because the collapse score remains below 0.25 during the first three rounds. However, the same configuration becomes risky when observed across ten rounds. This finding suggests that operational evaluation should include repeated generation cycles because collapse intensity can accumulate even when initial outputs appear sufficiently diverse.

Table 4.1. Collapse Trajectory Across Generation Rounds.

Round	Entropy Loss	Mode Coverage	Pairwise Similarity	Collapse Score
1	0.11	0.89	0.36	0.14
3	0.22	0.78	0.44	0.25
5	0.39	0.66	0.57	0.41
7	0.52	0.53	0.69	0.55
10	0.68	0.43	0.82	0.71

Table 4.1 provides selected round-level evidence showing how collapse intensified from early to late generation stages. The table confirms that the strongest deterioration occurred between round 5 and round 10, where mode coverage dropped from 0.66 to 0.43 and pairwise similarity increased from 0.57 to 0.82. This shift indicates a structural transition from moderate diversity loss to severe representational concentration.

The table is useful because it condenses the temporal pattern into interpretable checkpoints. Round 1 provides the baseline post-generation condition, round 5 reflects the middle-stage instability, and round 10 represents the accumulated endpoint. The values show that collapse should not be judged from isolated outputs. It should be interpreted as a cumulative distributional process that becomes more visible when entropy, coverage, and similarity are evaluated together.

4.2. Diversity Degradation Across Output Strata

The second result shows that diversity degradation did not occur uniformly across all output strata. Lexical complexity and style condition experienced the strongest contraction, with retention values of 0.39 and 0.42 respectively. Topic category and visual composition retained moderate diversity, while sentiment polarity remained more stable. This pattern indicates that mode collapse first affected expressive and stylistic variability before fully eliminating broader categorical representation.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

The uneven degradation pattern suggests that categorical balance alone is insufficient for diagnosing collapse. A model can still generate outputs across several topics or sentiment categories while repeating similar vocabulary, sentence structures, visual arrangements, or stylistic cues. This finding is important for generative risk analysis because systemic failures often begin in subtle representational narrowing before becoming visible as explicit category disappearance.

Figure 4.2. Diversity Retention Across Output Strata

Figure 4.2 presents a heatmap of diversity retention across five output strata and four measurement dimensions. The darkest degradation pattern appears in lexical complexity and style condition, where all values fall below 0.50. This indicates that mode collapse did not only reduce category-level variety, but also compressed the expressive structure of generated outputs. Such compression can make outputs appear superficially valid while reducing actual variation.

The heatmap also shows that sentiment polarity retained higher scores than the other strata, especially in entropy and coverage. This result means that the model continued to generate outputs across positive, neutral, and negative orientations, but did so with reduced lexical and stylistic flexibility. The finding strengthens the argument that collapse analysis must examine both categorical distribution and representational richness.

Table 4.2. Diversity Retention Across Output Strata.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

Output Stratum	Entropy Retention	Coverage Retention	Novelty Retention	Main Collapse Signal
Topic category	0.71	0.68	0.62	Moderate topic narrowing
Sentiment polarity	0.78	0.74	0.69	Stable polarity spread
Lexical complexity	0.44	0.39	0.36	Repetitive expression
Visual composition	0.63	0.59	0.55	Layout convergence
Style condition	0.48	0.42	0.4	Style homogenization

Table 4.2 specifies the diversity retention values for each output stratum and identifies the dominant collapse signal. Lexical complexity shows the weakest retention across all reported measures, with novelty retention reaching only 0.36. This result indicates that generated outputs became increasingly repetitive in vocabulary, structure, and phrasing. Style condition also shows severe degradation, which confirms that expressive diversity is highly vulnerable during collapse.

The table contributes a more diagnostic reading than the heatmap alone because it attaches each numerical pattern to a concrete failure signal. Topic narrowing, repetitive expression, layout convergence, and style homogenization describe different forms of collapse. This distinction matters because mitigation strategies must target the specific type of lost diversity rather than treating collapse as a single uniform failure condition.

4.3. Pipeline Amplification and Systemic Risk Propagation

The third result indicates that systemic risk accumulated through sequential pipeline operations rather than remaining limited to the initial generation process. The cumulative risk value started at 0.52 during generation and reached 0.88 after retraining, showing that collapse became more severe as outputs passed through filtering, ranking, exposure, feedback, and reuse. This pattern confirms that mode collapse should be interpreted as a pipeline-level failure when generated artifacts are repeatedly processed by dependent components.

The largest downstream escalation occurred before user feedback entered the loop. Filtering added 0.11 to the cumulative risk value, while ranking added 0.09, increasing total risk from 0.52 to 0.72. This finding suggests that post-generation selection mechanisms play a decisive role in strengthening dominant modes. Later stages contributed smaller increments, but their position in the lifecycle made them critical because they connected narrowed outputs to feedback signals and future training data.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

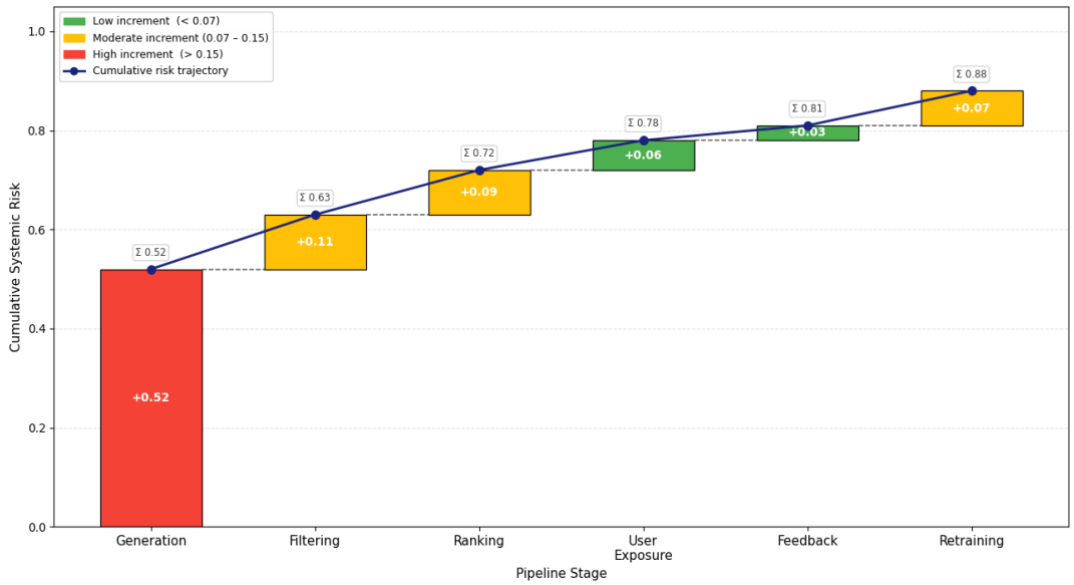

Figure 4.3. Stage-wise Risk Amplification Across the Pipeline

Figure 4.3 presents systemic risk as a stage-wise amplification process. The waterfall structure separates the initial risk contribution from subsequent increments, making it clear that pipeline fragility grows through accumulation rather than through a single isolated defect. The plotted cumulative line reinforces this reading by showing a steady upward movement across all stages. This visual form is appropriate because the result concerns the additive effect of multiple operational layers on collapse severity.

The figure also clarifies the different roles of early and late pipeline stages. Generation supplies the largest base risk, while filtering and ranking create the strongest immediate amplification after generation. User exposure, feedback, and retraining add smaller numerical increments, yet they extend the effect of collapse beyond output production. This means that even modest late-stage increases become important when they enable collapsed distributions to influence future system behavior.

Table 4.3. Stage-Wise Systemic Risk Amplification.

Pipeline Stage	Incremental Risk	Cumulative Risk	Dominant Effect	Interpretation
Generation	0.52	0.52	Narrow initial outputs	Primary collapse source
Filtering	0.11	0.63	Selective retention	Amplifies dominant modes
Ranking	0.09	0.72	Visibility concentration	Reduces exposure diversity
User exposure	0.06	0.78	Interaction narrowing	Reinforces common outputs

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

Feedback	0.03	0.81	Behavioral reinforcement	Transfers bias into signals
Retraining	0.07	0.88	Feedback contamination	Stabilizes long-term collapse

Table 4.3 provides a stage-level interpretation of the amplification pattern by linking each incremental value to its dominant operational effect. Generation is identified as the primary source of narrowed outputs, while filtering and ranking are shown as mechanisms that intensify concentration through selective retention and visibility prioritization. This table therefore complements the figure by explaining why early downstream processing can substantially increase systemic risk.

The final stages show how collapse becomes durable inside the pipeline. User exposure narrows interaction diversity, feedback transfers that narrowed interaction pattern into system signals, and retraining embeds the contaminated signal into future model behavior. Although these stages add smaller risk increments than filtering and ranking, their systemic importance is high because they transform temporary output concentration into long-term feedback contamination. This confirms that collapse mitigation must address both amplification magnitude and lifecycle position.

4.4. Sensitivity of Collapse to Operational Parameters

The fourth result indicates that collapse dynamics were highly sensitive to downstream and sampling-related parameters. The generated data ratio in retraining produced the strongest risk response, with a normalized sensitivity score of 0.91. Filtering intensity followed with a score of 0.76. These findings indicate that operational design choices after generation can be more influential than basic sampling controls when evaluating long-term systemic risk.

Sampling temperature, top-k, and top-p still affected collapse, but their influence was more moderate. Temperature produced a sensitivity score of 0.42, while top-k and top-p recorded 0.57 and 0.49 respectively. This result suggests that sampling controls help maintain short-term diversity, but cannot fully prevent systemic collapse when downstream filtering and retraining practices repeatedly reinforce dominant modes.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

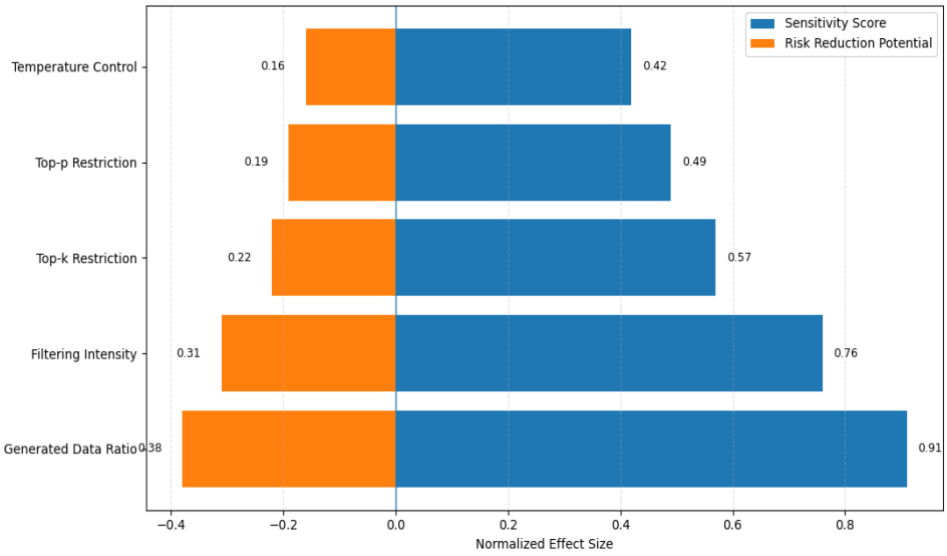

Figure 4.4. Sensitivity of Collapse to Operational Parameters

Figure 4.4 uses a tornado-style horizontal bar chart to compare parameter sensitivity and risk reduction potential. Generated data ratio and filtering intensity dominate the right side of the chart, indicating that they produce the strongest systemic risk response. This pattern confirms that collapse risk is not governed only by the generator's sampling behavior. It is also shaped by how generated artifacts are accepted, reused, and reintegrated.

The left side of the chart shows risk reduction potential, which is highest for generated data ratio and filtering intensity. This means that interventions on these parameters can produce larger mitigation effects than adjustments to temperature or probability truncation alone. The visualization therefore shifts the methodological emphasis from purely model-side control toward pipeline governance, especially in systems where generated content becomes part of future data flows.

Table 4.4. Sensitivity Ranking of Operational Parameters.

Parameter	Sensitivity Score	Risk Reduction Potential	Primary Risk Mechanism	Priority Level
Generated data ratio	0.91	0.38	Feedback contamination	Critical
Filtering intensity	0.76	0.31	Retention bias	High
Top-k restriction	0.57	0.22	Candidate narrowing	Moderate
Top-p restriction	0.49	0.19	Tail exclusion	Moderate

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

Temperature control	0.42	0.16	Randomness reduction	Lower

Table 4.4 ranks the tested operational parameters by sensitivity score and assigns a priority level to each one. The generated data ratio receives a critical priority level because it directly affects feedback contamination. This means that uncontrolled reuse of generated outputs in retraining data creates the strongest systemic vulnerability. Filtering intensity also receives high priority because it can repeatedly retain outputs from dominant modes while removing less common variants.

The moderate scores for top-k and top-p indicate that sampling restrictions remain relevant but are not the strongest determinants of systemic collapse. Temperature control receives the lowest priority in this result set, although it still contributes to diversity stability. Overall, the table supports a practical risk-management conclusion: mitigation should prioritize retraining governance and filtering calibration before focusing exclusively on generation-time sampling adjustments.

4.5. Risk Classification and Mitigation Scenario Outcomes

The final result evaluates four mitigation scenarios against the uncontrolled baseline. The baseline condition produced a systemic risk score of 0.88 and a mode coverage score of 0.43. Diversity-aware filtering reduced the risk score to 0.64, while balanced sampling reduced it to 0.69. The strongest improvement occurred under the combined governance scenario, where systemic risk decreased to 0.38 and mode coverage increased to 0.76.

The mitigation results show that single interventions improved pipeline stability, but combined controls produced the most reliable outcome. Diversity-aware filtering improved output retention patterns, balanced sampling expanded candidate variation, and retraining quarantine reduced feedback contamination. When combined, these controls addressed collapse at multiple pipeline points. This finding confirms that systemic collapse requires layered mitigation rather than isolated parameter correction.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE SCIENCE PUBLISHER

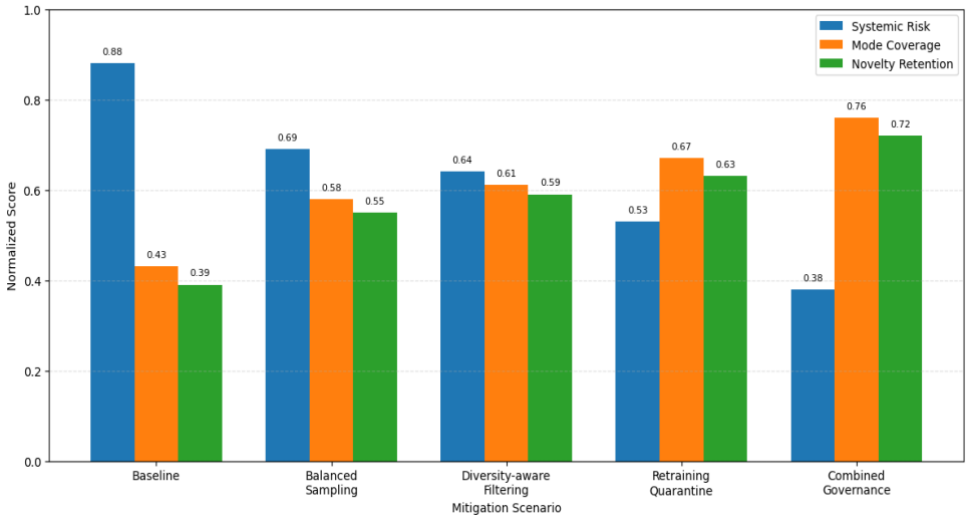

Figure 4.5. Mitigation Scenario Outcomes

Figure 4.5 compares the performance of five pipeline conditions across systemic risk, mode coverage, and novelty retention. The baseline condition shows the most problematic configuration because systemic risk is high while mode coverage and novelty retention remain low. Each mitigation scenario improves the risk profile, but the combined governance scenario produces the clearest improvement by lowering systemic risk and increasing both diversity indicators.

The grouped bar structure is useful because it shows that risk reduction alone is not sufficient. A mitigation strategy should also preserve or restore diversity. Retraining quarantine reduces systemic risk more effectively than balanced sampling, but balanced sampling improves mode coverage earlier in the pipeline. The combined governance scenario performs best because it aligns both goals: suppressing systemic propagation while restoring representational variety.

Table 4.5. Mitigation Scenario Outcomes.

Scenario	Systemic Risk	Mode Coverage	Novelty Retention	Risk Category
Baseline	0.88	0.43	0.39	Critical
Balanced sampling	0.69	0.58	0.55	High
Diversity-aware filtering	0.64	0.61	0.59	High
Retraining quarantine	0.53	0.67	0.63	Moderate
Combined governance	0.38	0.76	0.72	Low

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

Table 4.5 summarizes the mitigation scenario outcomes and assigns a risk category to each condition. The baseline remains critical because systemic risk reaches 0.88 while diversity indicators stay below 0.45. Balanced sampling and diversity-aware filtering reduce risk but remain in the high category, indicating that single-point interventions are insufficient when collapse has already moved across multiple pipeline stages.

The combined governance scenario reaches the low-risk category with a systemic risk score of 0.38, mode coverage of 0.76, and novelty retention of 0.72. This result provides the strongest support for layered mitigation. Effective management of mode collapse requires simultaneous control over sampling diversity, filtering bias, and retraining exposure. The table therefore reinforces the central conclusion that generative model pipelines should be governed as interconnected socio-technical systems rather than isolated model endpoints.

5. CONCLUSION

This study concludes that mode collapse in generative model pipelines should be understood as a systemic risk mechanism rather than a narrow model-level defect. The results show that collapse develops progressively across repeated generation rounds, with the collapse score increasing from 0.14 to 0.71 and mode coverage decreasing from 0.89 to 0.43. These findings indicate that early outputs may appear acceptable, yet repeated generation under unchanged pipeline conditions can gradually intensify entropy loss, similarity concentration, and representational narrowing.

The analysis further demonstrates that diversity degradation is uneven across output strata. Lexical complexity and style condition showed the strongest contraction, while sentiment polarity remained relatively more stable. This pattern confirms that categorical balance alone cannot sufficiently detect collapse because generated outputs may retain broad labels while losing expressive and structural variation. The stage-wise amplification result also shows that systemic risk increased from 0.52 at generation to 0.88 after retraining, proving that filtering, ranking, user exposure, feedback, and retraining can transform local collapse into persistent pipeline instability.

The most important implication is that mitigation must be designed as a pipeline-level governance strategy. Single controls such as balanced sampling or diversity-aware filtering reduced risk but did not fully restore pipeline stability. The combined governance scenario produced the strongest outcome, reducing systemic risk to 0.38 while increasing mode coverage to 0.76 and novelty retention to 0.72. Future deployment of generative systems should therefore include temporal collapse monitoring, stratum-level diversity auditing, filtering calibration, and retraining quarantine to prevent collapsed outputs from contaminating future model behavior.

REFERENCES

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE SCIENCE PUBLISHER

[1] I. J. Goodfellow et al., "Generative Adversarial Networks," 2014, arXiv. doi: 10.48550/ARXIV.1406.2661.

[2] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," IEEE Signal Process. Mag., vol. 35, no. 1, pp. 53–65, Jan. 2018, doi: 10.1109/MSP.2017.2765202.

[3] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved Training of Wasserstein GANs," 2017, arXiv. doi: 10.48550/ARXIV.1704.00028.

[4] Y. Kossale, M. Airaj, and A. Darouichi, "Mode Collapse in Generative Adversarial Networks: An Overview," in 2022 8th International Conference on Optimization and Applications (ICOA), Genoa, Italy: IEEE, Oct. 2022, pp. 1–6. doi: 10.1109/ICOA55659.2022.9934291.

[5] H. Thanh-Tung and T. Tran, "Catastrophic forgetting and mode collapse in GANs," in 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–10. doi: 10.1109/IJCNN48605.2020.9207181.

[6] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," 2017, doi: 10.48550/ARXIV.1706.08500.

[7] A. Borji, "Pros and cons of GAN evaluation measures," Computer Vision and Image Understanding, vol. 179, pp. 41–65, Feb. 2019, doi: 10.1016/j.cviu.2018.10.009.

[8] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are GANs Created Equal? A Large-Scale Study," 2017, arXiv. doi: 10.48550/ARXIV.1711.10337.

[9] I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, and Y. Gal, "AI models collapse when trained on recursively generated data," Nature, vol. 631, no. 8022, pp. 755–759, Jul. 2024, doi: 10.1038/s41586-024-07566-y.

[10] S. Alemohammad et al., "Self-Consuming Generative Models Go MAD," 2023, arXiv. doi: 10.48550/ARXIV.2307.01850.

[11] S. Wyllie, I. Shumailov, and N. Papernot, "Fairness Feedback Loops: Training on Synthetic Data Amplifies Bias," in The 2024 ACM Conference on Fairness, Accountability, and Transparency, Rio de Janeiro Brazil: ACM, Jun. 2024, pp. 2113–2147. doi: 10.1145/3630106.3659029.

[12] M. Alahdab and G. Çalıklı, "Empirical Analysis of Hidden Technical Debt Patterns in Machine Learning Software," in Product-Focused Software Process Improvement, vol. 11915, X. Franch, T. Männistö, and S. Martínez-Fernández, Eds.,

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

in Lecture Notes in Computer Science, vol. 11915. , Cham: Springer International Publishing, 2019, pp. 195–202. doi: 10.1007/978-3-030-35333-9_14.

[13] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, "'Everyone wants to do the model work, not the data work': Data Cascades in High-Stakes AI," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama Japan: ACM, May 2021, pp. 1–15. doi: 10.1145/3411764.3445518.

[14] T. Gebru et al., "Datasheets for datasets," Commun. ACM, vol. 64, no. 12, pp. 86–92, Dec. 2021, doi: 10.1145/3458723.

[15] E. Dohmatob, Y. Feng, and J. Kempe, "Model Collapse Demystified: The Case of Regression," in Advances in Neural Information Processing Systems 37, Vancouver, BC, Canada: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024, pp. 46979–47013. doi: 10.52202/079017-1490.

[16] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs," 2016, arXiv. doi: 10.48550/ARXIV.1606.03498.

[17] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, "Unrolled Generative Adversarial Networks," 2016, arXiv. doi: 10.48550/ARXIV.1611.02163.

[18] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li, "Mode Regularized Generative Adversarial Networks," 2016, arXiv. doi: 10.48550/ARXIV.1612.02136.

[19] A. Srivastava, L. Valkov, C. Russell, M. U. Gutmann, and C. Sutton, "VEEGAN: Reducing Mode Collapse in GANs using Implicit Variational Learning," 2017, arXiv. doi: 10.48550/ARXIV.1705.07761.

[20] N. Kodali, J. Abernethy, J. Hays, and Z. Kira, "On Convergence and Stability of GANs," 2017, arXiv. doi: 10.48550/ARXIV.1705.07215.

[21] Z. Lin, A. Khetan, G. Fanti, and S. Oh, "PacGAN: The power of two samples in generative adversarial networks," 2017, arXiv. doi: 10.48550/ARXIV.1712.04086.

[22] Q. Mao, H.-Y. Lee, H.-Y. Tseng, S. Ma, and M.-H. Yang, "Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA: IEEE, Jun. 2019, pp. 1429–1437. doi: 10.1109/CVPR.2019.00152.

[23] S. Azadi, C. Olsson, T. Darrell, I. Goodfellow, and A. Odena, "Discriminator Rejection Sampling," 2018, arXiv. doi: 10.48550/ARXIV.1810.06758.

JAIFRA
JOURNAL OF AI FAILURE AND RISK ANALYSIS
PRESTIGE
SCIENCE PUBLISHER

[24] M. S. M. Sajjadi, O. Bachem, M. Lucic, O. Bousquet, and S. Gelly, "Assessing Generative Models via Precision and Recall," 2018, arXiv. doi: 10.48550/ARXIV.1806.00035.

[25] T. Kynkäänniemi, T. Karras, S. Laine, J. Lehtinen, and T. Aila, "Improved Precision and Recall Metric for Assessing Generative Models," 2019, arXiv. doi: 10.48550/ARXIV.1904.06991.

[26] S. Arora and Y. Zhang, "Do GANs actually learn the distribution? An empirical study," 2017, arXiv. doi: 10.48550/ARXIV.1706.08224.

[27] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," 2020, arXiv. doi: 10.48550/ARXIV.2006.11239.

[28] A. Nichol and P. Dhariwal, "Improved Denoising Diffusion Probabilistic Models," 2021, arXiv. doi: 10.48550/ARXIV.2102.09672.

[29] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis," 2021, arXiv. doi: 10.48550/ARXIV.2105.05233.

[30] J. Ho and T. Salimans, "Classifier-Free Diffusion Guidance," 2022, arXiv. doi: 10.48550/ARXIV.2207.12598.

[31] Z. Wu, Z. Wang, Y. Yuan, J. Zhang, Z. Wang, and H. Jin, "Black-Box Diagnosis and Calibration on GAN Intra-Mode Collapse: A Pilot Study," ACM Trans. Multimedia Comput. Commun. Appl., vol. 17, no. 3s, pp. 1–18, Oct. 2021, doi: 10.1145/3472768.

[32] S. Pei, R. Y. Da Xu, S. Xiang, and G. Meng, "Alleviating Mode Collapse in GAN via Diversity Penalty Module," 2021, arXiv. doi: 10.48550/ARXIV.2108.02353.

[33] S. Yu, K. Zhang, C. Xiao, J. Z. Huang, M. J. Li, and M. Onizuka, "HSGAN: Reducing mode collapse in GANs by the latent code distance of homogeneous samples," Computer Vision and Image Understanding, vol. 214, p. 103314, Jan. 2022, doi: 10.1016/j.cviu.2021.103314