applied system innovation

MDPI

Article

A SIRS-Type Dynamical Model for Generative-AI Data Pollution with Synthetic Seeding and Real-Data Preservation

Kavindi Subawickrama and Upeksha Perera *

Department of Mathematics, University of Kelaniya, No. 218, Kandy Road, Dalugama, Kelaniya 11600, Sri Lanka; kavindisubawickrama@gmail.com

* Correspondence: upeksha@kln.ac.lk

Abstract

Generative-AI systems are increasingly trained on datasets containing AI-generated content, creating recursive feedback that may cause progressive losses of tail mass, diversity, and effective variance in learned distributions–a phenomenon commonly associated with model collapse. Existing analyses are primarily empirical or based on discrete-generation mixture models, motivating the development of a continuous-time dynamical framework. We develop a SIRS-type compartmental model, in which a normalized training-data pool is divided into clean, AI-polluted, and flagged/recovered fractions, with synthetic seeding (σ) and real-data preservation (ρ). We establish well-posedness, positivity, boundedness, and positive invariance; derive the equilibria and an explicit contamination threshold governing the clean equilibrium in the unseeded system; and analyze local stability and global convergence using planar dynamical-systems methods. Numerical simulations and parameter sweeps illustrate the resulting dynamics. In addition, a proof-of-concept fit of the reduced contamination–seeding subsystem to independently reported web-content prevalence data reproduces the observed post-2022 growth trend ($R^2 = 0.977$). Contamination (β) and synthetic seeding (σ) increase equilibrium pollution, whereas detection (γ) and real-data preservation (ρ) reduce it. Within the parameter ranges investigated, increasing the cleaning rate (α) can indirectly increase equilibrium pollution by returning more data to the clean compartment, where it is again exposed to contamination. These results identify continuous real-data preservation and effective detection as key mechanisms for limiting synthetic-data pollution in generative-AI training ecosystems.

Keywords: model collapse; generative AI; synthetic data; SIRS compartmental model; data pollution; global stability; Bendixson–Dulac criterion; dynamical systems

Check for updates

Academic Editor: Friedhelm Schwenker

Received: 15 August 2026
Revised: 26 October 2026
Accepted: 31 October 2026
Published: 2 November 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Beginning with the emergence of generative artificial intelligence (AI), the generation of digital content has rapidly changed. Large language models (LLMs), diffusion-based image generators, and other modern generative-AI models are increasingly being trained and retrained on data ecosystems containing growing amounts of AI-generated content. As synthetic output is reused in later training cycles, learned distributions can progressively lose low-probability modes, diversity, or effective variance and become increasingly concentrated relative to the original real-data distribution. This phenomenon, commonly described as model collapse or AI autophagy, has been documented empirically and theoretically. Shumailov et al. [1] demonstrated that recursively trained generative models can degrade when synthetic data are repeatedly fed back into training, while Alemohammad et al. [2]

analyzed self-consuming generative loops and discussed mitigation through retention of real data and curation of synthetic content. At the ecosystem level, Pan et al. [3] considered the risk that large volumes of machine-generated text may pollute online information environments. Recent work has further characterized distinct aspects of recursive synthetic-data training: linguistic diversity can decline across successive synthetic-text generations [4]; mixed real–synthetic training can be studied through distributional error bounds [5]; corrective procedures can stabilize self-consuming loops [6]; and collapse rates can be quantified for basic distribution families [7]. Experimental results also show that outcomes depend strongly on the training workflow and, in particular, on whether real data are retained or progressively replaced [8]. Machine-generated-text detection and resampling therefore offer one possible curation mechanism [9].

Two related developments make the composition of future training corpora especially important. First, high-quality human-generated text is a finite resource. Based on projected training demand and estimates of the public human-text stock, Villalobos et al. [10] argued that frontier language-model training could reach data availability constraints between 2026 and 2032 if prevailing scaling trends continue. Second, synthetic text is already measurable in web corpora that are commonly used in language-model development. A Pew Research Center analysis of 490,000 Common Crawl webpages reported that 9.35% of sampled `.com` pages in January 2026 showed significant signs of AI authorship or editing [11]. Independently, Russell et al. [12] analyzed FineWeb-filtered Common Crawl data and reported that 27.5% of tokens in June 2026 and 31.1% in August 2026 were labelled as AI-generated. These quantities should not be interpreted as directly comparable estimates of a single web-wide contamination rate because they use different units of analysis, sampling frames, filters, and detection procedures. Taken together, however, they provide independent evidence that machine-generated material is becoming a non-negligible component of web-scale text corpora.

The measurement problem itself is nontrivial because synthetic-content detection is imperfect. In a comparative evaluation of 14 detection systems, Weber-Wulff et al. [13] found substantial false-positive and false-negative behavior and concluded that available tools were not uniformly reliable. At a more fundamental level, Sadasivan et al. [14] showed that paraphrasing and spoofing attacks can sharply reduce the effectiveness of several classes of AI-text detectors and reduce detector performance to the statistical distinguishability of human and machine-generated text. Consequently, the detection parameter γ in the present model should be interpreted as an effective detection-and-removal rate for a curation pipeline, not as the accuracy of a particular classifier. For the same reason, the external web-prevalence series used later in this paper is treated as a detector-dependent proxy rather than ground-truth labelling of all online content.

At a high level, the training data at each generation can be considered a blend of (i) original or verified human-generated content and (ii) synthetic content generated by models, introduced either deliberately to reduce labelling costs or supplement incomplete datasets, or unintentionally as AI-generated content accumulates on the open web. Most existing analyses of recursive training are discrete-generation models that trace mixture proportions or distributional changes across successive training rounds. In contrast, continuous-time compartmental models such as the Susceptible–Infected–Recovered (SIR) family and its extensions are well established in epidemiology and have also been adapted to rumor and misinformation propagation [15–18]. Importantly for the present formulation, epidemic models with persistent importation or immigration of infective individuals provide a classical analogue of exogenous seeding: Brauer and van den Driessche [19] showed that sustained external influx can remove the disease-free equilibrium and force persistence. Although the interpretation here is data pollution rather than infection, the

structural analogy motivates distinguishing endogenous contamination from an external synthetic-seeding mechanism. Compartmental models then provide explicit thresholds, equilibria, and stability results using standard nonlinear dynamical-systems tools [20,21].

The central contribution of this paper is a SIRS-type compartmental model of generative-AI data pollution applied to the composition of a training-data pool rather than to a population of individual agents. We split a normalized data pool into three compartments: clean or human-generated data $S(t)$, polluted AI-generated data $I(t)$, and flagged or recovering data $R(t)$ that is undergoing cleaning or verification. Two mechanisms specific to generative-AI pipelines which are absent from classical epidemic and rumor-spreading models are introduced:

- a synthetic seeding rate σ, representing the deliberate or incidental injection of model-generated content into the data pool, even when very little pollution is currently present; and
- a real-data preservation or injection rate ρ, representing the continuous effort to preserve or re-introduce verified real data into the training mixture, with proportional replacement of existing content.

The novelty lies in the continuous-time data-ecosystem formulation and its AI-specific flow structure rather than in the stability tools themselves, which are standard dynamical-systems methods. In particular, the model (i) introduces external synthetic seeding that removes the clean equilibrium whenever $\sigma > 0$, (ii) places real-data preservation and detection together in the unseeded pollution threshold, (iii) yields a closed-form positive-pollution equilibrium for the seeded system, and (iv) shows that faster recovery can increase equilibrium pollution over the investigated parameter range because recovered data re-enters the clean compartment and is again exposed to contamination.

Importantly, $I(t)$ represents the fraction of the training-data pool that is AI-polluted; it is not itself a direct metric of model collapse. The model therefore addresses the dynamics of synthetic-data pollution as a mechanism that may contribute to collapse. Direct collapse outcomes such as loss of tail coverage, reduced variance, distributional drift, or downstream performance degradation are outside the present state space and require additional state variables and task-specific empirical calibration. To provide an independent empirical check of the core pollution–seeding mechanism without conflating it with full pipeline calibration, we additionally fit a reduced two-parameter subsystem to published Common Crawl-derived estimates of AI-authored web content.

The resulting continuous-time SIRS-type system permits rigorous analysis of positivity, invariance, equilibria, and stability. When $\sigma = 0$, the model has a clean equilibrium and an interpretable threshold governing pollution elimination or persistence. When $\sigma > 0$, synthetic seeding eliminates the clean equilibrium and produces a unique positive-pollution equilibrium. We derive this equilibrium in closed form and investigate its dependence on contamination, seeding, detection, cleaning, and preservation.

The remainder of the paper is structured as follows. Section 2 develops the model: formulation, behavioral assumptions, well-posedness, positivity and boundedness, equilibrium analysis and local and global stability, followed by the numerical methodology and a reduced external calibration procedure. Section 3 presents the results: baseline time-domain simulation, numerical decay of a Volterra-type functional, single- and two-parameter sensitivity sweeps, heatmap/contour analyses of the steady-state polluted fraction, and the proof-of-concept external calibration result. Section 4 discusses the implications of these results for the design of generative-AI data pipelines and places them in context with the epidemiology and misinformation-modeling literature. Section 5 presents the conclusion and outlines directions for future work, including explicit diversity and variance metrics, and global sensitivity analysis.

2. Materials and Methods

2.1. Model Formulation

Motivated by empirical evidence of model collapse due to recursive training on synthetic data [1,2], we modify the classic SIRS compartmental model with two mechanisms tailored to generative-AI data ecosystems: a synthetic seeding term σ, which allows direct influx of AI-generated content without requiring interaction with existing polluted material, and a real-data preservation term ρ, representing continuous proportional replacement of polluted or recovering material by verified real data. The interpretation is AI-specific, but the mathematical role of persistent external seeding has a precedent in epidemic models with importation of infectives, where sustained influx can eliminate the disease-free equilibrium [19]. Analogously, when $\sigma > 0$ in the present system, a perfectly clean equilibrium is no longer available, and the long-run state necessarily contains positive pollution.

We consider a normalized training data pool partitioned into three time-dependent fractions:

$$S(t) \quad \text{(clean/human data)}, \quad I(t) \quad \text{(AI-polluted data)}, \quad R(t) \quad \text{(flagged/recovering data)},$$

subject to $S(t) + I(t) + R(t) = 1$. The proposed model is,

$$\begin{aligned} \dot{S} &= \rho(1 - S) - \beta SI - \sigma S + \alpha R, \\ \dot{I} &= \beta SI + \sigma S - (\gamma + \rho)I, \\ \dot{R} &= \gamma I - (\alpha + \rho)R, \end{aligned} \tag{1}$$

where β, γ, α, and ρ are the contamination rate, detection rate, cleaning rate, and preservation rate, respectively (full interpretations are given in Table 1), and σ is the rate of synthetic seeding. In the present data-ecosystem interpretation, the term βSI does not represent literal contact between data items. It is a coarse-grained mixing/exposure term: when both clean material and AI-polluted or synthetic material are present, repeated web collection, dataset assembly, augmentation, and retraining create opportunities for clean portions of the future training pool to be replaced or combined with polluted content. The bilinear form assumes that the frequency of such opportunities scales with both fractions, so β should be interpreted as an effective mixing/contamination coefficient rather than an empirically calibrated infection rate. By contrast, σS represents direct synthetic seeding that can occur even when the current polluted fraction is negligible. This distinction is consistent with recent work emphasizing data curation and detection as mechanisms for controlling synthetic-content exposure [8,9].

Table 1. Parameter values and interpretations used in the model.

Parameter	Values Considered	Interpretation
β	0.8 to 1.2	Contamination intensity; higher values increase AI-contaminated data spread
σ	0.001 to 0.693	Synthetic-data injection rate; higher values indicate more AI-generated content
γ	0.105 to 0.693	Detection efficiency of contaminated data
α	0.5 to 1.0	Recovery rate through cleaning mechanisms
ρ	0.05 to 0.5	Preservation rate of authentic human-generated data

Qualitatively, R behaves as a holding stage between I and S such that polluted content can only be returned to the clean pool after passing through the detection process, while ρ

operates on all three simultaneously, continuously replacing polluted or recovering material with verified real data and thereby shifting the system toward the clean compartment.

Table 1 gives the parameter ranges considered in the full SIRS simulations. The full five-parameter set was not calibrated from an operating generative-AI training pipeline; these ranges were selected to cover light-to-heavy contamination and seeding, detection, cleaning, and preservation regimes rather than a single calibrated scenario. A separate proof-of-concept fit of the reduced pollution–seeding mechanism is described in Section 2.7.

2.1.1. Well-Posedness

We first examine the well-posedness before studying the equilibrium and stability properties. Each right-hand side of system (1) is a polynomial in S, I and R. Hence, the vector field is continuously differentiable and therefore locally Lipschitz in $\mathbb{R}^3$. By the Picard–Lindelöf theorem, every initial condition in $\mathbb{R}^3$ generates a unique local solution. The positivity and boundedness results established below imply that solutions starting in the feasible region remain in a bounded positively invariant set and therefore extend to all $t \geq 0$. Thus, for every feasible initial condition, system (1) has a unique solution defined for all $t \geq 0$.

2.1.2. Positivity

Proposition 1 (Positivity)**.** *If* $S(0) \geq 0$, $I(0) \geq 0$, *and* $R(0) \geq 0$, *then* $S(t) \geq 0$, $I(t) \geq 0$, *and* $R(t) \geq 0$ *for all* $t \geq 0$.

Proof. We examine the vector field on each coordinate boundary.

At $S = 0$, $\left.\frac{dS}{dt}\right|_{S=0} = \rho + \alpha R \geq 0$, so $S(t)$ cannot become negative.

At $I = 0$, $\left.\frac{dI}{dt}\right|_{I=0} = \sigma S \geq 0$, so $I(t)$ cannot become negative.

At $R = 0$, $\left.\frac{dR}{dt}\right|_{R=0} = \gamma I \geq 0$, so $R(t)$ cannot become negative.

Since the vector field points inward on each of the coordinate hyperplanes that bound the non-negative orthant, $\mathbb{R}^3_{\geq 0}$ is positively invariant. □

Proposition 2 (Invariance of the simplex)**.** *If* $S(0) + I(0) + R(0) = 1$, *then* $S(t) + I(t) + R(t) = 1$ *for all* $t \geq 0$.

Proof. Let $N(t) = S(t) + I(t) + R(t)$.

Differentiate $\dot{N}(t) = \dot{S} + \dot{I} + \dot{R}$

Summing the three equations of system (1), all bilinear and linear cross-terms cancel:

$$\dot{N} = \rho(1 - S) - \rho I - \rho R = \rho(1 - N).$$

This linear ODE has the explicit solution: $N(t) = 1 + (N(0) - 1)e^{-\rho t}$.

Since $N(0) = 1$, $N(t) = 1, \quad t \geq 0$.

Therefore, $S(t) + I(t) + R(t) = 1, \quad \forall t \geq 0$.

Hence, the normalized simplex $\Delta = \{(S, I, R) \in \mathbb{R}^3_{\geq 0} : S + I + R = 1\}$ is positively invariant. □

2.2. Boundedness and Feasible Region

From Proposition 2, $N(t) \to 1$ as $t \to \infty$ for any initial $N(0)$, and in particular $0 \leq N(t) \leq \max\{N(0), 1\}$. Combined with Proposition 1, trajectories with initial conditions in the operationally meaningful region remain in the closed, bounded, positively invariant region

$$\Omega = \{(S, I, R) \in \mathbb{R}^3_{\geq 0} : S + I + R \leq 1\}.$$

The normalized training-data model is restricted to the invariant simplex $\Delta = S + I + R = 1$, and all subsequent equilibrium and stability analysis is carried out on Δ.

In the present application, positivity means that the clean, polluted, and recovering fractions remain physically meaningful and cannot become negative. Boundedness and invariance mean that the training-data composition remains within a normalized feasible pool, so no compartment fraction can leave the admissible range or create data mass outside the modeled pool.

2.3. Equilibrium Points

The proposed SIRS model has equilibrium points that describe long-term states of the data ecosystem with a stable ratio of clean, AI-polluted, and recovering data over time. These equilibria are used to gain insight into the possible asymptotic behavior of the system and are used as the basis for the stability analysis that follows.

Equilibria are obtained by setting $\dot{S} = \dot{I} = \dot{R} = 0$.

(i) Pollution-free (clean) equilibrium ($\sigma = 0$); Setting $I = R = 0$ satisfies $\dot{I} = \dot{R} = 0$ identically. Substituting into $\dot{S}$ gives $\rho(1 - S) = 0$, so $S = 1$. Thus, when $\sigma = 0$,

$$E_0 = (1, 0, 0).$$

This is the pollution-free equilibrium. Note that when $\sigma > 0$, E_0 is not an equilibrium, since $\dot{I}|_{S=1,I=0} = \sigma \neq 0$: synthetic seeding rules out a perfectly clean steady state whenever it is active.

(ii) Positive equilibrium for $\sigma > 0$; From $\dot{R} = 0$,

$$R^* = \frac{\gamma I^*}{\alpha + \rho}.$$

From $\dot{I} = 0$,

$$\beta S^* I^* + \sigma S^* - (\gamma + \rho) I^* = 0 \Rightarrow S^* = \frac{(\gamma + \rho) I^*}{\beta I^* + \sigma}.$$

Substituting S^* and R^* into $\dot{S} = 0$ and using the invariance relation $S^* + I^* + R^* = 1$ (Proposition 2) yields a quadratic equation in I^*.

$$\rho\left(1 - \frac{(\gamma + \rho) I^*}{(\beta I^* + \sigma)}\right) - \frac{\beta(\gamma + \rho) I^{*2}}{(\beta I^* + \sigma)} - \frac{\sigma(\gamma + \rho) I^*}{(\beta I^* + \sigma)} + \frac{\alpha\gamma I^*}{(\alpha + \rho)} = 0.$$

Equivalently,

$$aI^{*2} + bI^* + c = 0,$$

with

$$I^* = \frac{-b + \sqrt{b^2 - 4ac}}{2a},$$

where

$$a = \beta\left(1 + \frac{\gamma}{\alpha + \rho}\right) > 0, \qquad b = \gamma + \rho + \sigma + \frac{\gamma\sigma}{\alpha + \rho} - \beta, \qquad c = -\sigma < 0.$$

Since $a > 0$ and $c < 0$, this quadratic has exactly one positive root $I^* > 0$ for any $\sigma > 0$, giving a unique positive equilibrium, $E^* = (S^*, I^*, R^*)$, where $S^* = 1 - \left(1 + \frac{\gamma}{\alpha+\rho}\right) I^*, R^* = \frac{\gamma I^*}{\alpha + \rho}$.

When $\sigma = 0$ and $\mathcal{R}_0 > 1$, the positive equilibrium is,

$$S^* = \frac{\gamma+\rho}{\beta}, \qquad I^* = \frac{1-\frac{\gamma+\rho}{\beta}}{1+\frac{\gamma}{\alpha+\rho}}, \qquad R^* = \frac{\gamma I^*}{\alpha+\rho}.$$

(iii) Threshold condition; For $\sigma = 0$, linearizing the infected equation at the clean equilibrium $E_0 = (1,0,0)$ gives $\dot{I} = (\beta - (\gamma+\rho))I$ giving the threshold,

$$\mathcal{R}_0 = \frac{\beta}{\gamma+\rho}.$$

When $\sigma = 0$, if $\mathcal{R}_0 < 1$, polluted data are asymptotically eliminated, and E_0 is stable. If $\mathcal{R}_0 > 1$, E_0 is unstable and a positive equilibrium exists. When $\sigma > 0$, E_0 is not an equilibrium. Therefore, $\mathcal{R}_0$ is not an extinction threshold for the seeded system, although it remains the threshold inherited from the corresponding unseeded model. This is similar to the traditional reproduction number threshold of epidemic theory [15–17], where persistent pollution is suppressed by both preservation (ρ) and detection (γ).

2.4. Local Stability

We use the standard linearization criterion for autonomous ordinary differential equations: local stability is determined from the eigenvalues of the Jacobian evaluated at an equilibrium [20,21]. Writing, $f_1 = \dot{S}$, $f_2 = \dot{I}$, $f_3 = \dot{R}$, the Jacobian of system (1) is,

$$J(S,I,R) = \begin{bmatrix} -\rho - \beta I - \sigma & -\beta S & \alpha \\ \beta I + \sigma & \beta S - (\gamma+\rho) & 0 \\ 0 & \gamma & -(\alpha+\rho) \end{bmatrix}.$$

Since the clean equilibrium $E_0 = (1,0,0)$ exists only when $\sigma = 0$, its Jacobian is,

$$J(E_0) = \begin{bmatrix} -\rho & -\beta & \alpha \\ 0 & \beta - (\gamma+\rho) & 0 \\ 0 & \gamma & -(\alpha+\rho) \end{bmatrix}.$$

Its characteristic polynomial is,

$$(\lambda+\rho)(\lambda+\alpha+\rho)\big(\lambda-\beta+\gamma+\rho\big) = 0,$$

so the eigenvalues are

$$\lambda_1 = -\rho, \quad \lambda_2 = -(\alpha+\rho), \quad \lambda_3 = \beta - (\gamma+\rho).$$

Thus, E_0 is locally asymptotically stable when $\mathcal{R}_0 < 1$ and unstable when $\mathcal{R}_0 > 1$.

Proposition 3. *When $\sigma = 0$, the clean equilibrium $E_0 = (1,0,0)$ is locally asymptotically stable if $\mathcal{R}_0 = \frac{\beta}{\gamma+\rho} < 1$, and unstable if $\mathcal{R}_0 > 1$.*

For the positive equilibrium, use the invariant relation $R = 1 - S - I$, to obtain the reduced planar system,

$$\dot{S} = (\rho+\alpha) - (\rho+\sigma+\alpha)S - \alpha I - \beta SI,$$
$$\dot{I} = \beta SI + \sigma S - (\gamma+\rho)I.$$

At $E^* = (S^*, I^*)$, the equilibrium relation $\dot{I} = 0$ gives $\beta S^* - (\gamma+\rho) = -\frac{\sigma S^*}{I^*}$.

Its Jacobian is,

$$J_r(E^*) = \begin{bmatrix} -(\rho+\sigma+\alpha) - \beta I^* & -\alpha - \beta S^* \\ \beta I^* + \sigma & -\dfrac{\sigma S^*}{I^*} \end{bmatrix}.$$

Therefore,

$$\operatorname{tr}(J_r(E^*)) = -(\rho+\sigma+\alpha) - \beta I^* - \frac{\sigma S^*}{I^*} < 0,$$

and

$$\det(J_r(E^*)) = (\rho+\sigma+\alpha+\beta I^*)\frac{\sigma S^*}{I^*} + (\alpha+\beta S^*)(\beta I^*+\sigma) > 0.$$

Hence, the positive equilibrium is locally asymptotically stable on the invariant simplex whenever it exists. In the full three-dimensional system, the transverse variable $N = S + I + R$ satisfies $\dot{N} = \rho(1-N)$, giving the additional transverse eigenvalue $-\rho < 0$.

2.5. Global Stability

The global arguments below use standard Lyapunov monotonicity and planar dynamical-systems results, including the Bendixson–Dulac criterion and the Poincaré–Bendixson theorem [20,21].

Global stability of the clean equilibrium: For $\sigma = 0$ and $\mathcal{R}_0 < 1$, define $V_0(I) = I$. Since $0 \le S \le 1$ on Δ,

$$\dot{V}_0 = [\beta S - (\gamma+\rho)]I \le [\beta - (\gamma+\rho)]I < 0$$

for $I > 0$, $I(t)$ decays to zero. Moreover, $\dot{R} + (\alpha+\rho)R = \gamma I$, so $R(t) \to 0$, and the invariance relation $S + I + R = 1$ then gives $S(t) \to 1$. Thus, $E_0 = (1,0,0)$ is globally asymptotically stable on Δ.

Proposition 4. *If $\sigma = 0$ and $\mathcal{R}_0 < 1$, the clean equilibrium $E_0 = (1,0,0)$ is globally asymptotically stable on Δ.*

Global convergence to the positive equilibrium: Because Δ is invariant and

$$S + I + R = 1,$$

we eliminate R by setting,

$$R = 1 - S - I.$$

The dynamics on Δ are therefore equivalent to the planar system

$$\begin{aligned} \dot{S} &= (\rho+\alpha) - (\rho+\sigma+\alpha)S - \alpha I - \beta SI, \\ \dot{I} &= \beta SI + \sigma S - (\gamma+\rho)I. \end{aligned} \tag{2}$$

Consider the open, simply connected region

$$D^\circ = \{(S,I) : S > 0,\ I > 0,\ S + I < 1\}.$$

For $\sigma > 0$, Section 2.3 establishes a unique equilibrium (S^*, I^*) in D°, and Section 2.4 establishes its local asymptotic stability.

Choose the Dulac function

$$B(S,I) = \frac{1}{SI}, \qquad (S,I) \in D^\circ. \tag{3}$$

Writing

$$f(S,I) = (\rho+\alpha) - (\rho+\sigma+\alpha)S - \alpha I - \beta SI,$$

and

$$g(S,I) = \beta SI + \sigma S - (\gamma+\rho)I,$$

we obtain

$$Bf = \frac{\rho+\alpha}{SI} - \frac{\rho+\sigma+\alpha}{I} - \frac{\alpha}{S} - \beta, \qquad Bg = \beta + \frac{\sigma}{I} - \frac{\gamma+\rho}{S}.$$

Hence,

$$\frac{\partial(Bf)}{\partial S} = -\frac{\rho+\alpha}{S^2 I} + \frac{\alpha}{S^2},$$

and

$$\frac{\partial(Bg)}{\partial I} = -\frac{\sigma}{I^2}.$$

Consequently,

$$\nabla\cdot(Bf,Bg) = \frac{\partial(Bf)}{\partial S} + \frac{\partial(Bg)}{\partial I} \quad = \frac{\alpha I - (\rho+\alpha)}{S^2 I} - \frac{\sigma}{I^2}. \tag{4}$$

For every $(S,I) \in D^\circ$, we have $0 < I < 1$.

Since $\rho > 0$ and $\alpha > 0$, $\quad \alpha I < \alpha < \rho + \alpha$, and therefore $\alpha I - (\rho+\alpha) < 0$.

Together with $\sigma > 0$, Equation (4) gives

$$\nabla\cdot(Bf,Bg) < 0, \qquad \forall\,(S,I) \in D^\circ. \tag{5}$$

Thus, by the Bendixson–Dulac criterion, system (2) admits no periodic orbit entirely contained in D°.

It remains to consider the boundary of the feasible region. On $I = 0$,

$$\dot{I} = \sigma S > 0, \qquad S > 0,$$

so no trajectory with $S > 0$ can remain on the boundary $I = 0$. On $S = 0$,

$$\dot{S} = \rho + \alpha(1-I) > 0,$$

because $0 \le I \le 1$. Hence trajectories cannot remain on $S = 0$. Finally, on $S + I = 1$, which is equivalent to $R = 0$,

$$\frac{d}{dt}(S+I) = -\dot{R} = -\gamma I < 0, \qquad I > 0.$$

Therefore, the boundary $S + I = 1$ is also not positively invariant except at the limiting boundary point corresponding to the clean state, which is not an equilibrium when $\sigma > 0$.

Hence, for $\sigma > 0$, every trajectory starting in D° remains in the interior and its ω-limit set cannot contain a boundary orbit or a periodic orbit. Since the system has a unique equilibrium E^* in the interior, the Poincaré–Bendixson theorem implies

$$(S(t), I(t)) \longrightarrow (S^*, I^*), \qquad t \to \infty.$$

Since $R(t) = 1 - S(t) - I(t)$, it follows that $R(t) \longrightarrow R^*$.

Therefore,

$$(S(t), I(t), R(t)) \longrightarrow (S^*, I^*, R^*), \qquad t \to \infty.$$

Thus, the unique positive equilibrium E^* is globally asymptotically stable in the relative interior of the invariant simplex Δ.

Proposition 5. *The unique positive equilibrium is globally asymptotically stable in the relative interior of the invariant simplex Δ when $\sigma > 0$. When $\sigma = 0$ and $\mathcal{R}_0 > 1$, the same Dulac function gives*

$$\nabla \cdot (Bf, Bg) = \frac{\alpha I - (\rho + \alpha)}{S^2 I} < 0$$

throughout D°. Moreover, $I = 0$ is invariant, while trajectories with $I(0) > 0$ remain in the interior. At E_0, the stable eigendirection is tangent to the invariant boundary $I = 0$; hence, the local stable manifold of E_0 is contained in this boundary. Therefore, no trajectory with $I(0) > 0$ can converge to E_0. Hence, by the Poincaré–Bendixson theorem, every trajectory with $I(0) > 0$ converges to the unique positive equilibrium E^*.

2.6. Numerical Methods

No closed-form time-domain solution is used in the numerical analysis. Transient trajectories were obtained using ode45, whereas the closed-form equilibrium expressions were used for the steady-state parameter sweeps. The system was integrated numerically using MATLAB's `ode45` solver over $0 \leq t \leq T$, with output recorded at intervals of $\Delta t = 0.01$. The initial condition was $(S(0), I(0), R(0)) = (0.7, 0.2, 0.1)$, with RelTol $= 10^{-8}$, and AbsTol $= 10^{-10}$. This initial condition was chosen solely as an illustrative mixed state with nonzero mass in all three compartments and was not inferred from an operating generative-AI pipeline. The terminal time increased when necessary until $\|\dot{x}(T)\| < \varepsilon$, where $\varepsilon = 10^{-8}$. The steady-state fractions in the sweeps below are just the values that each trajectory will eventually reach once this transient is over.

The full five-parameter baseline was not calibrated from an operating generative-AI training pipeline. Therefore, the ranges in Table 1 were selected to represent qualitatively different dynamical regimes rather than to reproduce a particular calibrated scenario. The separate external calibration below is deliberately restricted to the two parameters identifiable in a reduced pollution–seeding subsystem and does not convert the baseline SIRS parameter set into an empirically calibrated pipeline model. The baseline parameter set used throughout the numerical study is,

$$\beta = 0.8, \qquad \sigma = 0.001, \qquad \gamma = 0.105, \qquad \alpha = 0.5, \qquad \rho = 0.105. \tag{6}$$

Single-parameter sweeps were performed by varying one parameter from the set $\{\alpha, \beta, \gamma, \sigma, \rho\}$ at a time, while keeping all remaining parameters fixed at their nominal values. The closed-form equilibrium expressions were used to compute steady-state values in the parameter sweeps. Time integration with ode45 was used to verify representative parameter combinations. This procedure was repeated for the grids in (β, ρ), (σ, ρ), (α, ρ), and (γ, ρ) to create heatmaps and contour plots of I^*, classifying parameter combinations as low-, moderate-, or high-pollution regimes. A near-clean state was defined numerically as $I^* < 10^{-3}$.

2.7. External Proof-of-Concept Calibration of the Reduced Pollution–Seeding Mechanism

To provide an independent empirical check of the contamination–seeding mechanism, we considered public estimates of AI-authored or substantially AI-edited web content reported by the Pew Research Center from English-language webpages sampled from the Common Crawl archive [11]. The study analyzed approximately 490,000 webpages and reports six-month averages by top-level domain. We used the `.com` observations from January 2023 to January 2026: 1.63, 2.76, 3.76, 4.70, 5.48, 6.64, and 9.35 percent. These data are used only as an external environmental proxy for the prevalence of synthetic content and are not treated as direct observations of the three compartments of an operating training pipeline. Because AI-text classifiers can exhibit false positives, false negatives, and

sensitivity to rewriting or paraphrasing [13,14], the reported percentages are treated as detector-dependent prevalence estimates rather than ground-truth labels.

To isolate the two mechanisms that can be identified from this single prevalence series, we considered the reduced case $R = 0$, $\gamma = \alpha = \rho = 0$, and $S = 1 - I$. System (1) then reduces to

$$\frac{dI}{dt} = \beta I(1 - I) + \sigma(1 - I), \tag{7}$$

where $I(t)$ is interpreted, for this calibration exercise only, as the observed fraction of webpages exhibiting substantial AI authorship or editing. The parameters β and σ were estimated by nonlinear least squares,

$$(\widehat{\beta}, \widehat{\sigma}) = \arg\min_{\beta,\sigma>0} \sum_j [I(t_j; \beta, \sigma) - I_j]^2, \tag{8}$$

using MATLAB R2019a (MathWorks, Natick, MA, USA) with `ode45` for time integration and `fminsearch` with a logarithmic parameter transformation to enforce positivity. The first observation, January 2023, was used as the initial condition. This exercise is a proof-of-concept calibration of the reduced contamination–seeding mechanism, not a calibration of γ, α, or ρ and not a validation of the full five-parameter SIRS model.

3. Results

3.1. Baseline Time Evolution

With the baseline parameters (6), the system (1) was integrated from the initial condition $(S(0), I(0), R(0)) = (0.7, 0.2, 0.1)$. The resulting trajectories of $S(t)$, $I(t)$ and $R(t)$ are shown in Figure 1. If the pool is contaminated and seeded, the clean-data fraction decreases, the polluted-data fraction increases, and the recovering fraction undergoes a short transient before all three variables approach their equilibrium values. All three compartments converge smoothly without oscillation to steady values within about ten time units, consistent with the analytically predicted steady endemic state $(S^*, I^*, R^*) \approx (0.26, 0.63, 0.11)$, and with the stability results for the baseline parameters, obtained in Sections 2.4 and 2.5.

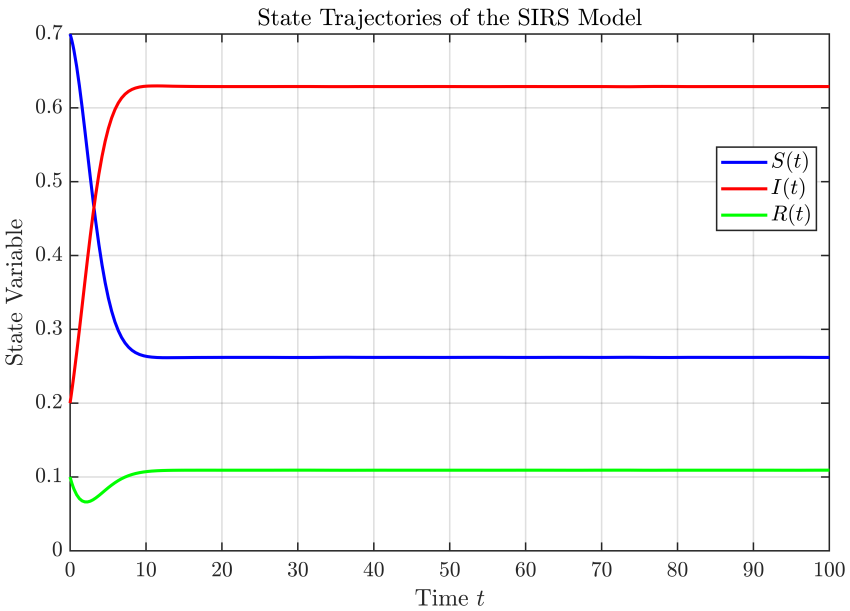


Figure 1. Time evolution of the SIRS model representing clean data $S(t)$, AI-polluted data $I(t)$, and recovered data $R(t)$.

3.2. Numerical Behavior of the Volterra-Type Functional

For an additional numerical illustration of convergence toward the positive equilibrium, we define the standard Volterra-type functional

$$V(S,I,R) = \left(S - S^* - S^* \ln \frac{S}{S^*}\right) + \left(I - I^* - I^* \ln \frac{I}{I^*}\right) + \left(R - R^* - R^* \ln \frac{R}{R^*}\right). \quad (9)$$

which is commonly used in the stability analysis of compartmental models [22,23]. Since $z - 1 - \ln z \geq 0$ for $z > 0$, we have $V \geq 0$, with $V = 0$ only at E^*. In this study, V is used only as a numerical diagnostic; the analytical global-convergence result is established independently in Section 2.5 using the Bendixson–Dulac criterion. The resulting numerical evolution is shown in Figure 2.

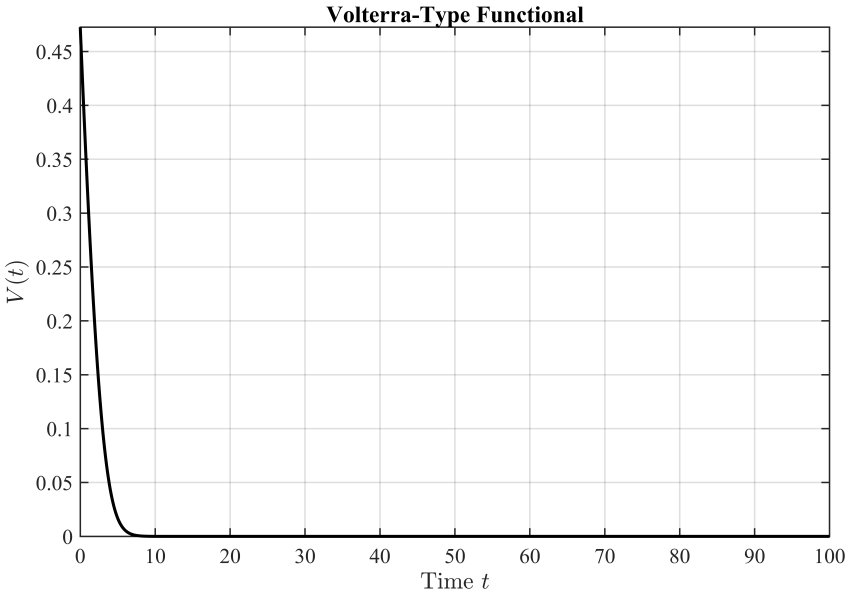


Figure 2. Numerical evolution of the Volterra-type functional $V(t)$ along the baseline trajectory.

The Volterra-type functional V defined in (9) was evaluated along the baseline numerical trajectory. For the baseline simulation, $V(t)$ decreases toward zero as the trajectory approaches the positive equilibrium. This observed decay is consistent with the global-convergence result in Proposition 5 but is not used as an independent Lyapunov proof.

3.3. Single-Parameter Sensitivity Analysis

The steady-state values (S^*, I^*, R^*) from sweeping each of the parameters alone around the baseline (6) are reported in Figures 3 and 4. For $\sigma > 0$, the equilibrium satisfies $I^* > 0$. We therefore classify $I^* < 10^{-3}$ as near-clean and larger values as persistent pollution.

Increasing β causes a gradual decrease in S^* while a gradual increase in I^* occurs, corresponding to a greater influx of new polluted data into the compartment being recovered; R^* is increasing with I^* as more data enters the recovered compartment. Under γ the pattern is reversed: S^* rises about linearly with γ as more and more polluted data enters the recovering compartment, I^* falls steeply, nearly monotonically, with γ, and R^* is non-monotonic, increasing with γ until it becomes sufficiently large that recovered data are purged back to S faster than they enter from I.

For α, S^* only increases slightly for the swept range and R^* decreases steadily as expected to drain the recovering compartment into S. Within the parameter range examined, I^* increases with α. In this model, increasing α transfers data more rapidly from R to S, thereby enlarging the fraction exposed to the contamination term βSI.

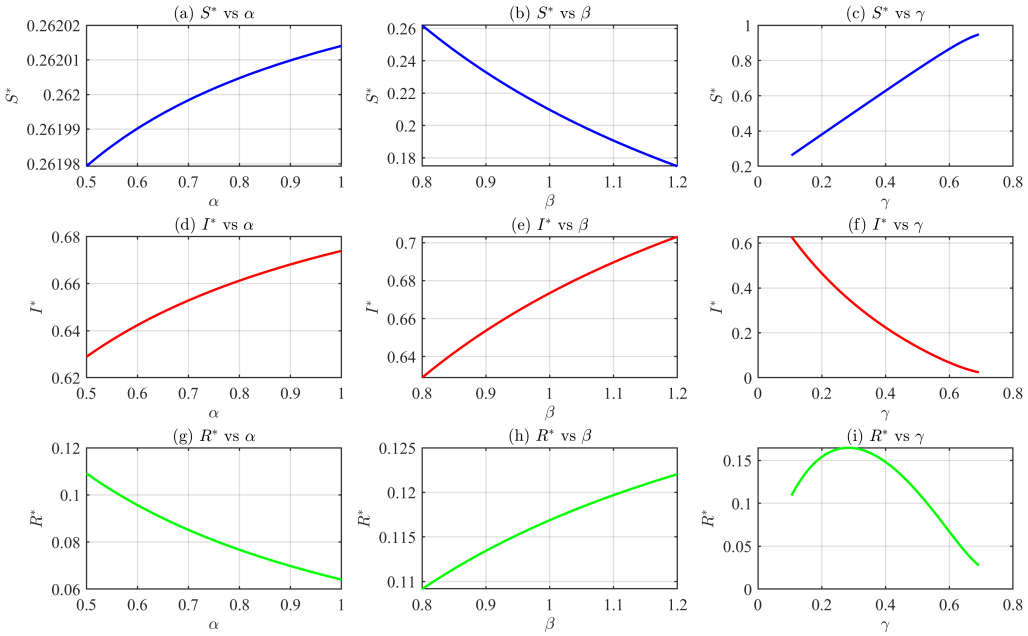


Figure 3. Single-parameter equilibrium sweeps: (**a**–**c**) S^* versus α, β, and γ; (**d**–**f**) I^* versus α, β, and γ; and (**g**–**i**) R^* versus α, β, and γ, respectively.

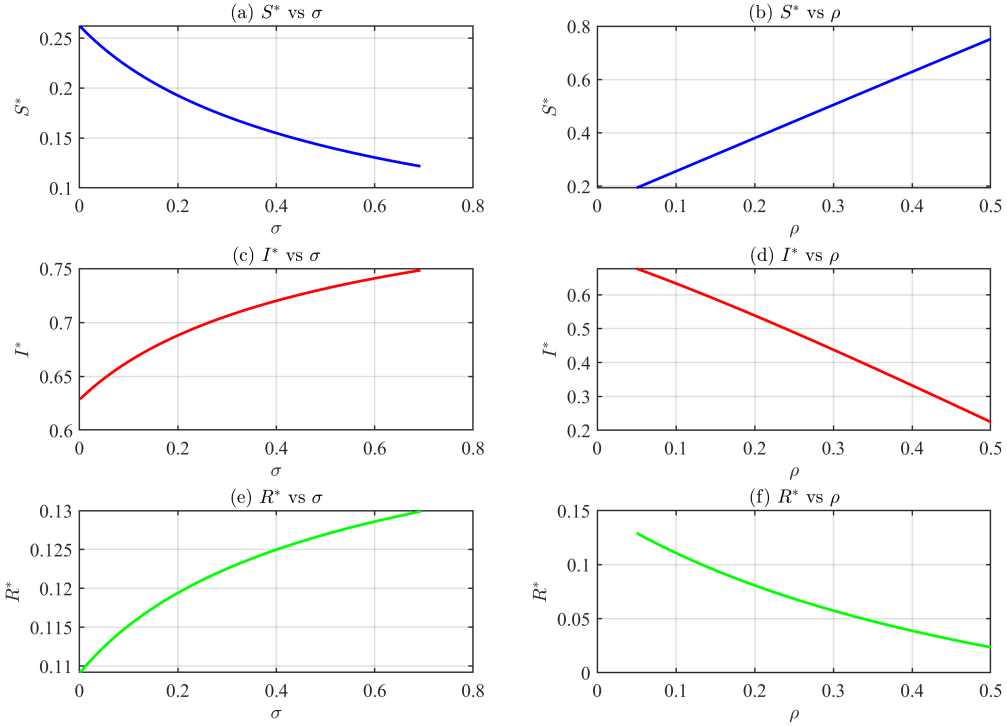


Figure 4. Single-parameter equilibrium sweeps: (**a**,**b**) S^* versus σ and ρ; (**c**,**d**) I^* versus σ and ρ; and (**e**,**f**) R^* versus σ and ρ, respectively.

This distinguishes α sharply from γ and ρ, both of which act on I in a way that nets out to reducing I^* and it is consistent with the threshold $\mathcal{R}_0 = \frac{\beta}{\gamma+\rho}$, derived in Section 2.3, which likewise contains no dependence on α.

All three steady states shift slightly over the range of parameters considered under σ: S^* decreases marginally, I^* and R^* increase marginally, showing that the effect of synthetic seeding is rather small compared to the rest of the parameter range under consideration. In contrast, under ρ, the response is pronounced: S^* increases quickly and almost linearly with increasing ρ; whereas I^* decreases accordingly, and R^* again exhibits a non-monotonic

response, a rise followed by a fall, as more and more data are recovered back to S through R, and ρ appears in the outflow term $(\alpha + \rho)R$ as well as the inflow to S.

In all sweeps, β and σ raise the steady-state fraction of the polluted population I^*, while γ and ρ lower it. The data returned by α, which we consider a "recovery" mechanism, increases the steady-state fraction of the polluted population I^* indirectly, since the data it brings back to S is now available for contamination. S^* and R^* mirror this behavior in each case, as required by $S^* + I^* + R^* = 1$. For the ranges examined with $\sigma > 0$, the model always possesses a positive-pollution equilibrium. Moreover, increasing α alone is not a mitigating mechanism over the investigated parameter range; effective pollution reduction instead requires sufficiently strong detection, preservation, or reductions in contamination and seeding.

3.4. Two-Parameter Analysis (Heatmaps and Contours)

Two-parameter sweeps were conducted to investigate the interaction between the parameters (β, ρ), (σ, ρ), (α, ρ), and (γ, ρ) while keeping all other parameters at their standard values. For each pair, the heatmap displays the entire intensity of I^* over the parameter space and the contour plots display specified levels of I^*, allowing low-, moderate-, and high-pollution regions to be compared.

Figure 5 shows the overall impact of the contamination rate β, synthetic seeding rate σ, and real-data preservation rate ρ on I^*.

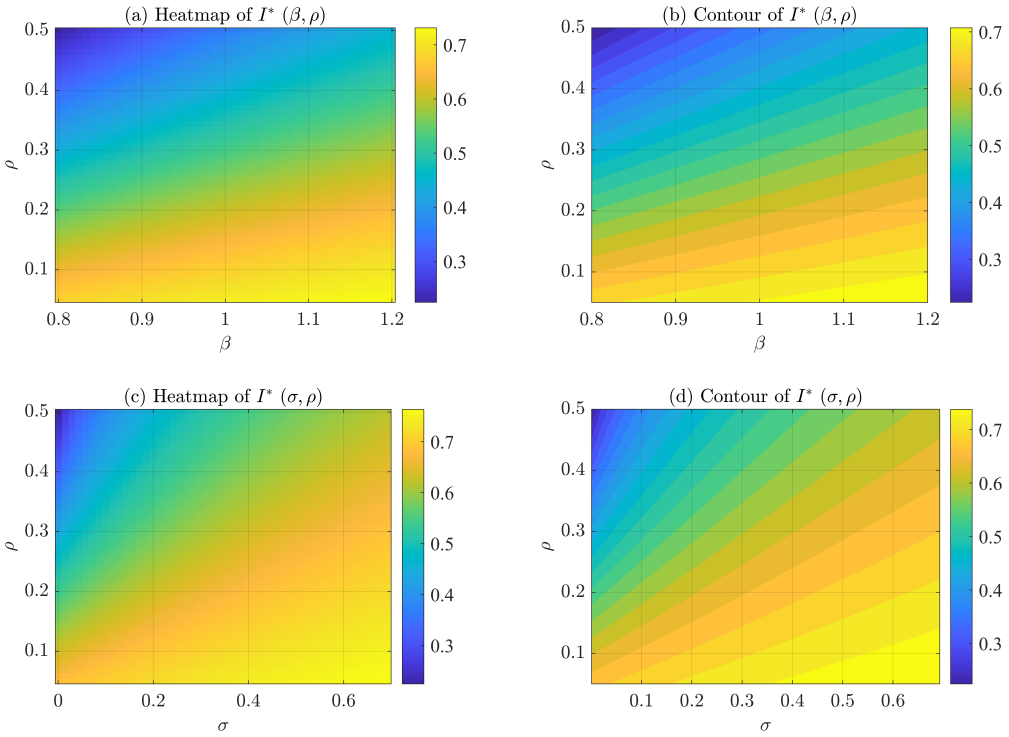


Figure 5. Two-parameter equilibrium pollution: (**a**) heatmap and (**b**) contour plot of I^* for (β, ρ); (**c**) heatmap and (**d**) contour plot of I^* for (σ, ρ).

In Figure 5a,b, a high β and a low ρ produce the highest steady-state pollution, while the contour lines illustrate a clear trade-off: in order to compensate for a high β, the pollution can be reduced by increasing the value of ρ, and thereby moving the system from the region of high pollution to the low-pollution or near-clean region, according to the criterion $I^* < 10^{-3}$. Figure 5c,d indicates that, for the ranges of interest, the effect of the seeding rate σ is relatively small compared to the effect of the preservation rate ρ: the steady-state pollution level is strongly dominated by the preservation rate and higher

values of ρ for the preservation rate result in uniformly lower levels of pollution in the system states.

Figure 6 shows how the steady-state level of AI-polluted data changes with the cleaning rate α, the detection rate γ, and the preservation rate ρ.

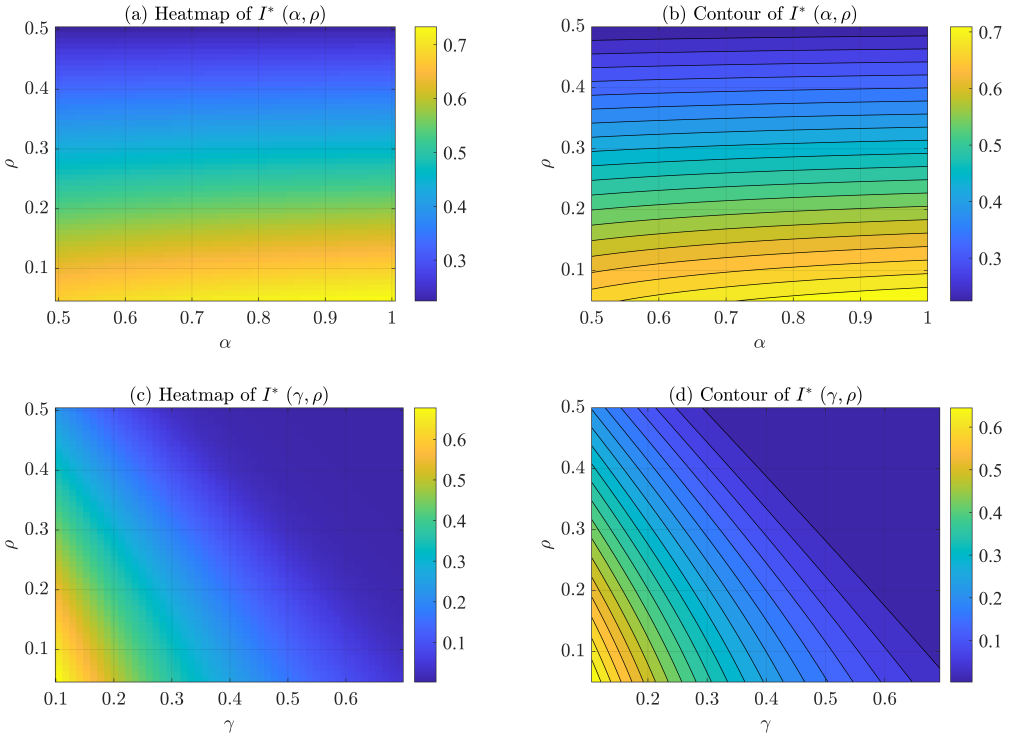


Figure 6. Two-parameter equilibrium pollution: (**a**) heatmap and (**b**) contour plot of I^* for (α, ρ); (**c**) heatmap and (**d**) contour plot of I^* for (γ, ρ).

Figure 6a,b shows the (α, ρ) interaction. Increasing ρ moves the heatmap toward lower I^* across the entire α range, whereas increasing α at fixed ρ produces a modest increase in I^*. The nearly horizontal contour structure confirms that preservation dominates the variation in this panel and that faster recovery alone does not mitigate pollution. Figure 6c,d shows the (γ, ρ) interaction. Here both increasing detection and increasing preservation reduce I^*, with the lowest-pollution region occurring when both are large. The downward trend across both axes demonstrates that γ and ρ act cooperatively, unlike the (α, ρ) case. Thus, over the parameter region shown, preservation is the strongest common mitigating direction, while the effect of α reflects redistribution from R back to S and renewed exposure to the contamination pathway.

3.5. State Trajectory Analysis

To further illustrate the global dynamics of the model, the baseline simulation was performed again with a number of different initial conditions, keeping the baseline parameter values unchanged. This analysis shows that the convergence presented in Section 3.1 is not only applicable to a specific initial condition, but that the temporal evolution of each compartment can be analyzed individually without having to look at a single overlaid trajectory.

Figure 7 shows that, despite different initial distributions of clean, AI-polluted, and recovering data, all trajectories converge to the same positive equilibrium. Although the transient dynamics vary with the initial condition, the long-term behavior remains unchanged. This provides numerical evidence of convergence from the selected initial conditions and is consistent with the global-convergence result in Proposition 5.

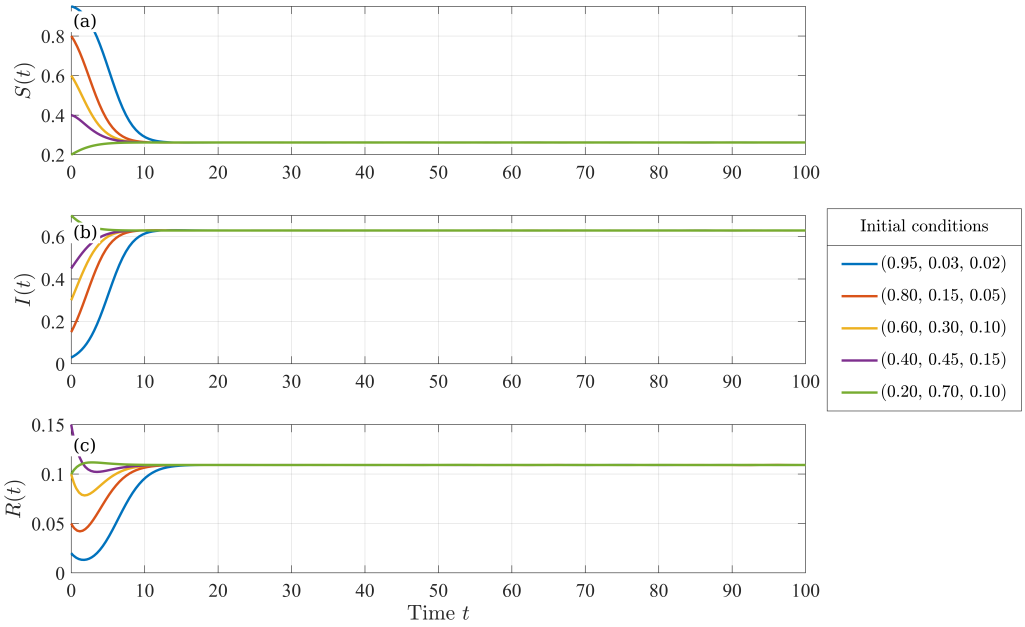


Figure 7. Time evolution for several initial conditions under the baseline parameter values: (**a**) clean-data fraction $S(t)$, (**b**) AI-polluted fraction $I(t)$, and (**c**) recovering/flagged fraction $R(t)$.

3.6. Phase Portrait Analysis

In order to gain further insight into the dynamical behavior of the system, the qualitative relationships between the state variables were investigated using phase portraits.

Figure 8 illustrates the convergence of the selected trajectory toward the positive equilibrium. The projections show no oscillation or closed orbit along this trajectory. Global convergence follows from the analytical results, not from the phase portrait alone.

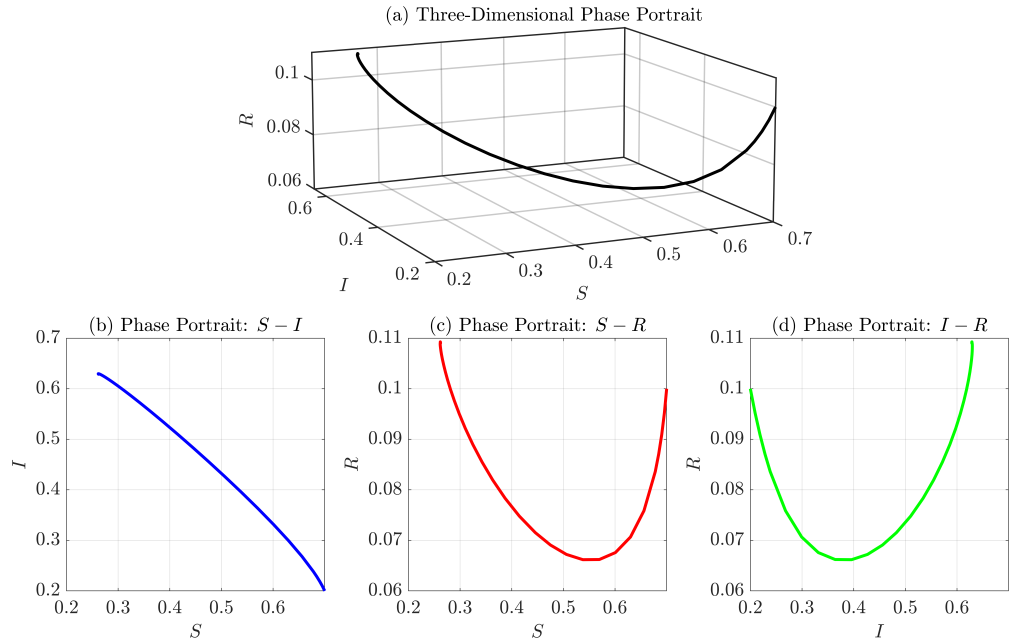


Figure 8. Phase portraits for the baseline trajectory: (**a**) three-dimensional (S, I, R) trajectory; (**b**) S–I projection; (**c**) S–R projection; and (**d**) I–R projection.

3.7. External Proof-of-Concept Calibration Result

Fitting the reduced model (7) to the seven .com observations from January 2023 to January 2026 yielded

$$\widehat{\beta} = 0.3394 \text{ yr}^{-1}, \qquad \widehat{\sigma} = 0.00962 \text{ yr}^{-1}.$$

The fitted trajectory reproduced the observed increase in the prevalence proxy with $R^2 = 0.977$ and a root-mean-square error of 0.359 percentage points (Figure 9). The agreement provides an external check that the reduced contamination–seeding mechanism can reproduce the observed growth trend with only two fitted parameters. It does not establish that the full SIRS parameter values are empirically calibrated, because the external series does not separately observe detection, recovery, or deliberate real-data preservation.

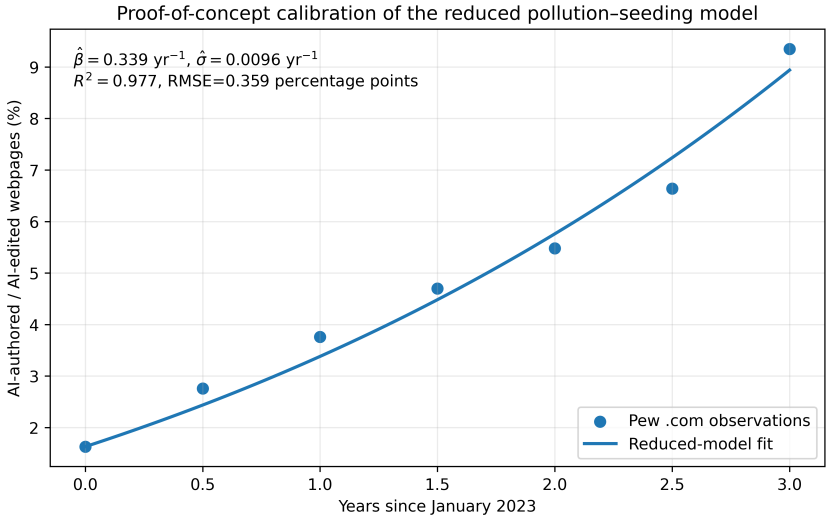


Figure 9. Proof-of-concept calibration of the reduced pollution–seeding subsystem against the percentage of .com webpages showing significant signs of AI editing or authorship in the Pew Research Center Common Crawl analysis, January 2023–January 2026 [11]. Points represent the reported six-month averages and the curve represents the fitted reduced model.

3.8. Summary of Findings

Over the parameter ranges investigated, ρ produces the largest reduction in the equilibrium polluted fraction, with the contamination rate, β, and synthetic seeding rate, σ, being the main contributors to pollution. Detection (γ) offers complementary and cooperative mitigation in addition to preservation. By contrast, recovery rate (α) does not act as a mitigating factor: it works as expected to reduce R^*, but it also increases I^* because the data it returns to the susceptible pool is now available for contamination, pointing out again that rapid recovery is not, on its own, equivalent to reducing pollution, and that this asymmetry is already reflected in the threshold $\mathcal{R}_0 = \frac{\beta}{\gamma+\rho}$, which depends on γ and ρ but not on α. The numerical trajectories are consistent with convergence toward the analytically derived equilibrium. The threshold $\mathcal{R}_0$ applies to the unseeded system, whereas the seeded system has a positive equilibrium for every $\sigma > 0$. Thus, while preservation and detection can substantially reduce equilibrium pollution, a perfectly clean equilibrium is impossible as long as synthetic seeding remains active.

4. Discussion

The proposed continuous-time SIRS-type model provides an analytically tractable framework for studying generative-AI data pollution, a mechanism that may contribute to model collapse. Recent work measures collapse more directly through quantities such as

linguistic diversity, distributional distance, variance contraction, or downstream performance [4,5,7,8]. In contrast, $I(t)$ in the present model is a pollution-burden variable rather than a direct predictor of those collapse outcomes. The reduced external calibration in Section 3.7 provides a limited empirical check of the contamination–seeding growth mechanism: two fitted parameters reproduce an independently observed rise in AI-authored web content with $R^2 = 0.977$. The wider literature supports the relevance of both sides of this mechanism: human-generated text may become a binding training resource under continued scaling [10], while independent Common Crawl analyses report substantial and increasing machine-generated content in web-scale corpora [11,12]. At the same time, detector reliability remains imperfect [13,14], so these prevalence estimates should be interpreted as method-dependent indicators rather than exact measurements of a universal contamination fraction. Because the calibration observations describe the wider web rather than an instrumented training pipeline, the result supports the plausibility of the reduced mechanism but does not validate the complete SIRS model. Accordingly, the analytical results establish conditions for persistence or suppression of synthetic-data pollution, while a quantitative map from $I(t)$ to tail coverage, variance, or model quality requires additional state variables and empirical calibration. The threshold quantity $\mathcal{R}_0 = \frac{\beta}{\gamma+\rho}$ plays a role analogous to the basic reproduction number in classical epidemic theory [15–17,22,23]: for the unseeded system ($\sigma = 0$), if $\mathcal{R}_0 < 1$ then polluted content is eliminated and the clean equilibrium is globally stable, whereas if $\mathcal{R}_0 > 1$ pollution persists at a computable positive equilibrium. Notably, $\mathcal{R}_0$ depends on γ and ρ but not on α, a structural feature that anticipates the counterintuitive equilibrium finding discussed below. The seeding term also has a close mathematical analogue in epidemic models with importation: sustained external influx can remove the disease-free equilibrium [19]. In the present AI setting, this means that when synthetic seeding remains active, policy design should target an acceptably low long-run pollution level rather than exact elimination.

The two-parameter analysis sets the relative importance of policy levers, and also reveals one that behaves differently than its label suggests. The parameters that are most difficult to control directly are contamination (β) and seeding (σ), as these parameters depend on the wider data ecosystem, e.g., the amount of AI-generated content on the open web. The two mechanisms that are directly under the control of an organization running a training pipeline, and which, in our numerical study, are the first two that are effective in suppressing pollution, are detection (γ) and real-data preservation (ρ). The recovery rate (α) is likewise under an organization's control, but the model shows it should not be treated as interchangeable with γ and ρ: because α governs the rate at which recovering data re-enters the clean data pool rather than removing polluted data outright, increasing α enlarges the pool available for renewed contamination through the βSI and σS terms and, over the parameter ranges investigated here, increases the equilibrium polluted fraction rather than reducing it (Sections 3.3 and 3.4). The dependence on α is indirect rather than explicit, since the I-equation contains no α term; changes in I^* therefore occur through the effect of α on S^*. Practically, it implies that fast recovery of flagged data is not, by itself, a safeguard against long-run contamination: recovered data that re-enters circulation without adequate detection or preservation capacity in place becomes available to be polluted again. This dominance of ρ in the two-parameter heatmaps thus suggests that, in this continuous-time compartmental scenario, the most effective mitigation strategy is continuous reinjection of verified real data, not just recovery rate alone, as was qualitatively recommended by the model-collapse literature [1,2,8,9] and quantified within the proposed model, with the identification of the recovery rate as a lever that must be paired with detection or preservation to be net beneficial.

There is a need to be clear about what this type of model can and cannot tell us. In exchange for the analytical tractability that made the equilibrium and stability results of Section 2 possible, one has to make certain compromises in collapsing the entire, highly heterogeneous data ecosystem into three aggregate compartments, such as what sources or platforms might be contributing polluted content, how exposure rates across different sources or platforms vary, and any network effects that govern how synthetic material might propagate. This is the cost of closed-form thresholds rather than simulation-only conclusions, and is the same trade-off made by analogous low-dimensional treatments of misinformation spread [18]. The full five-parameter SIRS baseline should therefore not be interpreted as a calibrated operational model. The proof-of-concept external fit estimates only β and σ in a deliberately reduced two-state setting; the available external series cannot identify γ, α, or ρ, nor does it observe clean, polluted, and recovering fractions inside an actual training pipeline. Accordingly, the full-model sweeps in Section 3 still represent qualitative scenarios rather than a fitted operating point, including the direction of the α effect. Within the investigated parameter range, the direction of the α-effect follows from the model structure, although its practical magnitude has not been empirically validated. The sweeps presented here also do not sample all five parameters together, and a complete global sweep remains to be carried out, which would be valuable to see how the effect of α on pollution affects the entire parameter space, as opposed to only the region of parameter space used in the baseline sweep presented here.

5. Conclusions

We have developed a three-compartment SIRS-type dynamical model of generative-AI data pollution, which decomposes a normalized training data pool into clean, AI-polluted, and flagged/recovering parts, and adds two generative-AI pipeline-specific mechanisms: synthetic seeding (σ) and real-data preservation (ρ). We have established well-posedness, positivity, and invariance of the feasible simplex. We derived the clean equilibrium for the unseeded system and closed-form expressions for the positive equilibrium. The threshold $\mathcal{R}_0 = \frac{\beta}{\gamma+\rho}$ determines the stability of the clean equilibrium when $\sigma = 0$. Local stability and global convergence of the positive equilibrium were established using planar dynamical-systems analysis. Numerical simulations illustrated the analytical results. A separate proof-of-concept fit of the reduced contamination–seeding subsystem to independently reported web-content prevalence data yielded $\widehat{\beta} = 0.3394\ \mathrm{yr}^{-1}$ and $\widehat{\sigma} = 0.00962\ \mathrm{yr}^{-1}$ with $R^2 = 0.977$; this provides an external check of the reduced mechanism rather than calibration of the full SIRS system. Under the illustrative baseline parameter set, the positive equilibrium is approximately $(S^*, I^*, R^*) = (0.262, 0.629, 0.109)$. Holding the remaining baseline parameters fixed, increasing ρ from 0.05 to 0.50 reduces I^* from approximately 0.677 to 0.224, while increasing γ from 0.105 to 0.693 reduces I^* from approximately 0.629 to 0.024. These values are model-based scenario results rather than calibrated predictions. Over the parameter ranges investigated, detection and real-data preservation reduced equilibrium pollution, whereas contamination and synthetic seeding increased it.

This separation between truly mitigating parameters and α implies that, for pipeline design, curators of generative-AI training datasets can use the framework as a scenario-screening tool to compare the relative effects of stronger provenance/detection filters, greater retention of verified real data, and tighter synthetic-data controls. It should not be interpreted as prescribing calibrated operational rates. The results also show that speeding up the recovery of flagged data is not, by itself, sufficient to reduce long-run pollution. The model translates qualitative recommendations concerning real-data preservation and synthetic-data control into explicit equilibrium relationships and a threshold condition for the unseeded system.

The model invites three natural extensions. First, coupling the polluted compartment I to auxiliary state variables tracking tail-mass, diversity, variance, or downstream performance would allow the framework to model-collapse outcomes directly rather than using pollution as a precursor mechanism. Second, sampling the five parameters β, σ, γ, α, and ρ simultaneously, instead of sequentially, would determine if α's effect on pollution is only found in the parameter regimes used here, or if it is found throughout the entire parameter space. Third, relaxing the well-mixed, deterministic structure (by adding stochastic noise, or using a network structure, as in related extensions to rumor-spreading models [18]) would enable the framework to reflect the fact that different data sources are also heterogeneous, which the present aggregate compartments ignore.

Author Contributions: Conceptualization, U.P. and K.S.; methodology, U.P. and K.S.; software, K.S.; validation, K.S. and U.P.; formal analysis, K.S. and U.P.; investigation, K.S.; writing—original draft preparation, K.S.; writing—review and editing, U.P.; visualization, K.S.; supervision, U.P.; project administration, U.P.; funding acquisition, U.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Kelaniya Research Council, grant number RC/2026/PPCR01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The MATLAB code and numerical data used to generate the original model results are available at the following GitHub repository: https://github.com/kavindisubawickrama-blip/SIRS-GAI-Data-Pollution, accessed on 11 August 2026. The external calibration data are publicly reported by the Pew Research Center [11]; the values used for the fit and the MATLAB calibration script are included with the revised submission source files.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Shumailov, I.; Shumaylov, Z.; Zhao, Y.; Papernot, N.; Anderson, R.; Gal, Y. AI models collapse when trained on recursively generated data. *Nature* **2024**, *631*, 755–759. . Correction in *Nature* **2025**, *640*, E6. [CrossRef] [PubMed]
2. Alemohammad, S.; Casco-Rodriguez, J.; Luzi, L.; Humayun, A.I.; Babaei, H.; LeJeune, D.; Siahkoohi, A.; Baraniuk, R.G. Self-Consuming Generative Models Go MAD. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
3. Pan, Y.; Pan, L.; Chen, W.; Nakov, P.; Kan, M.Y.; Wang, W.Y. On the Risk of Misinformation Pollution with Large Language Models. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023*; Association for Computational Linguistics: Singapore, 2023; pp. 1389–1403. [CrossRef]
4. Guo, Y.; Shang, G.; Vazirgiannis, M.; Clavel, C. The Curious Decline of Linguistic Diversity: Training Language Models on Synthetic Text. In *Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2024*; Association for Computational Linguistics: Mexico City, Mexico, 2024; pp. 3589–3604. [CrossRef]
5. Fu, S.; Zhang, S.; Wang, Y.; Tian, X.; Tao, D. Towards Theoretical Understandings of Self-Consuming Generative Models. In Proceedings of the 41st International Conference on Machine Learning, PMLR, Vienna, Austria, 21–27 July 2024 ; Volume 235, pp. 14228–14255.
6. Gillman, N.; Freeman, M.; Aggarwal, D.; Hsu, C.H.; Luo, C.; Tian, Y.; Sun, C. Self-Correcting Self-Consuming Loops for Generative Model Training. In Proceedings of the 41st International Conference on Machine Learning, PMLR, Vienna, Austria, 21–27 July 2024; Volume 235, pp. 15646–15677.
7. Suresh, A.T.; Thangaraj, A.; Khandavally, A.N.K. Rate of Model Collapse in Recursive Training. In Proceedings of the 28th International Conference on Artificial Intelligence and Statistics, PMLR, Mai Khao, Thailand, 3–5 May 2025; Volume 258, pp. 1396–1404.

8. Kazdan, J.; Schaeffer, R.; Dey, A.; Gerstgrasser, M.; Rafailov, R.; Donoho, D.L.; Koyejo, S. Collapse or Thrive: Perils and Promises of Synthetic Data in a Self-Generating World. In Proceedings of the 42nd International Conference on Machine Learning, PMLR, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 29469–29494.
9. Drayson, G.; Yilmaz, E.; Lampos, V. Machine-Generated Text Detection Prevents Language Model Collapse. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Singapore, 2025, pp. 29657–29673. [CrossRef]
10. Villalobos, P.; Ho, A.; Sevilla, J.; Besiroglu, T.; Heim, L.; Hobbhahn, M. Position: Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data. In Proceedings of the 41st International Conference on Machine Learning, PMLR, Vienna, Austria, 21–27 July 2024; Volume 235, pp. 49523–49544.
11. Bestvater, S.; Smith, A.; TerBush, C.; Baronavski, C.; Chavda, J. How Much of the Internet Is Written with AI? Pew Research Center Data Labs. 20 August 2026. Available online: https://www.pewresearch.org/data-labs/2026/08/20/how-much-of-the-internet-is-written-with-ai/ (accessed on 15 October 2026).
12. Russell, J.; Glickenhaus, B.; Thai, K.; Wieting, J.; Iyyer, M.; Spero, M.; Emi, B. How Much Is an AI Token Worth? Scaling Laws for Wild AI-Generated Web Text. *arXiv* **2026**, arXiv:2609.40295.
13. Weber-Wulff, D.; Anohina-Naumeca, A.; Bjelobaba, S.; Foltýnek, T.; Guerrero-Dib, J.; Popoola, O.; Šigut, P.; Waddington, L. Testing of Detection Tools for AI-Generated Text. *Int. J. Educ. Integr.* **2023**, *19*, 26. [CrossRef]
14. Sadasivan, V.S.; Kumar, A.; Balasubramanian, S.; Wang, W.; Feizi, S. Can AI-Generated Text Be Reliably Detected? Stress Testing AI Text Detectors Under Various Attacks. *Trans. Mach. Learn. Res.* **2025**. Available online: https://openreview.net/forum?id=OOgsAZdFOt (accessed on 30 October 2026).
15. Kermack, W.O.; McKendrick, A.G. A Contribution to the Mathematical Theory of Epidemics. *Proc. R. Soc. Lond. Ser. A* **1927**, *115*, 700–721. [CrossRef]
16. Hethcote, H.W. The Mathematics of Infectious Diseases. *SIAM Rev.* **2000**, *42*, 599–653. [CrossRef]
17. Korobeinikov, A.; Wake, G.C. Lyapunov functions and global stability for SIR, SIRS, and SIS epidemiological models. *Appl. Math. Lett.* **2002**, *15*, 955–960. [CrossRef]
18. Zhao, L.; Wang, J.; Chen, Y.; Wang, Q.; Cheng, J.; Cui, H. SIHR rumor spreading model in social networks. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 2444–2453. [CrossRef]
19. Brauer, F.; van den Driessche, P. Models for Transmission of Disease with Immigration of Infectives. *Math. Biosci.* **2001**, *171*, 143–154. [CrossRef] [PubMed]
20. Perko, L. *Differential Equations and Dynamical Systems*, 3rd ed.; Texts in Applied Mathematics ; Springer: New York, NY, USA, 2001; Volume 7. [CrossRef]
21. Brauer, F.; Castillo-Chavez, C.; Feng, Z. *Mathematical Models in Epidemiology*; Texts in Applied Mathematics; Springer: New York, NY, USA, 2019; Volume 69. [CrossRef]
22. Korobeinikov, A. Lyapunov functions and global properties for SEIR and SEIS epidemic models. *Math. Med. Biol. J. IMA* **2004**, *21*, 75–83. [CrossRef]
23. Korobeinikov, A.; Maini, P.K. A Lyapunov Function and Global Properties for SIR and SEIR Epidemiological Models with Nonlinear Incidence. *Math. Biosci. Eng.* **2004**, *1*, 57–60. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.