a

StatMat
Jurnal Statistika dan Matematika

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

A Generalized Saturation Model and a Data-Saturation Theory for the Pricing of Artificial-Intelligence Training Data

Michael Mncedisi Willie[1,*]

[1]Policy, Research and Monitoring, Council for Medical Schemes, Pretoria, South Africa

***Correspondence author:** m.willie@medicalschemes.co.za

Abstract

The fees charged for access to data used to train and operate artificial-intelligence (AI) systems are widely observed to rise with the volume of data and then to flatten, a pattern this paper interprets as saturation. I propose a data-saturation theory of AI data pricing, grounded in the empirical regularity that model performance improves as a power law in data with diminishing returns, and I show that a rational, value-based fee schedule inherits this diminishing-returns structure and therefore saturates toward a ceiling. To represent the resulting curve I introduce the generalized data-saturation (GDS) model, a three-parameter saturation kernel that nests the Michaelis–Menten, Hill and stretched-exponential families as special cases and is governed by interpretable scale, sharpness and tail-curvature parameters. I develop the model's analytic properties the half-saturation volume, the marginal-fee and elasticity functions, and a saturation index and a full inferential framework based on nonlinear least squares, including identifiability conditions, the Jacobian, asymptotic normality and model selection by information criteria. A simulation study (synthetic data) shows that the ceiling, baseline and scale parameters are recovered with low bias and near-nominal confidence-interval coverage, while the shape parameters are estimable but weakly identified under high noise or small samples, a property reported transparently. The framework gives analysts a principled, parsimonious tool for describing and comparing AI data-fee schedules, and a vocabulary for reasoning about where, and how sharply, the value of additional data saturates.

Keywords: saturation model, nonlinear regression, diminishing returns, neural scaling laws, artificial intelligence.

MSC 2020 subject classifications: 62J02 (nonlinear regression); 62F12 (asymptotic properties of estimators); 62P20 (applications to economics).

ARTICLE INFO	
Submission received: 10 June 2026	Accepted: 31 August 2026
Revised: 16 June 2026	Published: 31 August 2026
Available on: https://doi.org/10.32493/sm.v8i2.60720	

a

StatMat
Jurnal Statistika dan Matematika

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

INTRODUCTION

The commercial supply of data for artificial-intelligence systems has grown into a substantial market, and with it the question of how such data should be priced. A recurring qualitative observation is that the fee a buyer is willing to pay, or that a supplier can command, rises steeply as the first tranches of data are acquired and then flattens: beyond some volume, additional records add little to the performance of the model they feed, and the price schedule reflects this. The phenomenon is intuitively one of saturation, but it has not, to my knowledge, been given an explicit statistical model or a theoretical justification rooted in how AI systems actually learn. This paper supplies both.

The motivation for treating data fees as a saturating function is not merely descriptive convenience. A large body of empirical work on the scaling of machine-learning systems has established that predictive performance improves as a power law in the quantity of training data, with an exponent that implies sharply diminishing returns (Hestness et al., 2017; Kaplan et al., 2020; Hoffmann et al., 2022; Sun et al., 2017). If the economic value of data derives from the performance it buys, and if performance gains decay as a power law, then the cumulative value of data and any fee schedule anchored to that value must be concave and bounded. Saturation, on this account, is not an empirical accident but a consequence of the learning curve. Section 2 makes this argument precise and states it as a small set of axioms, which together constitute what I call the data-saturation theory of AI data pricing.

Translating the theory into a usable statistical instrument requires a parametric family flexible enough to capture the range of saturation shapes that different data, tasks and markets might produce, yet parsimonious and interpretable enough to be estimated from modest samples and compared across settings. The contribution of this paper, beyond the theory itself, is the generalized data-saturation (GDS) model introduced in Section 3: a saturation kernel with a scale parameter, a sharpness parameter and a tail-curvature parameter that nests three classical families Michaelis–Menten, Hill and stretched-exponential as special cases, so that the data themselves can indicate which mechanism is operating. Section 4 develops the inferential apparatus, Section 5 reports a simulation study, Section 6 illustrates the method on a synthetic fee schedule, and Sections 7 and 8 discuss the approach and its limitations. Throughout, all numerical illustrations use simulated data and are labelled as such; the paper makes no empirical claim about the fees charged by any actual data supplier.

2. A data-saturation theory of AI data pricing

Let $x \geq 0$ denote the volume of data supplied, measured in whatever natural unit the market uses (records, tokens, labelled examples), and let $F(x)$ denote the fee associated with that volume. The theory rests on four assumptions, the first three of which formalise the qualitative observations above and the fourth of which supplies their mechanism.

Axiom 1 (monotonicity). F is non-decreasing in x: more data is never worth less. Axiom 2 (diminishing marginal fee). The marginal fee $F'(x)$ is eventually non-increasing, so the schedule is eventually concave.

a

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

Axiom 3 (bounded saturation). F is bounded above, with $F(x) \to \alpha < \infty$ as $x \to \infty$; the ceiling α is the saturation fee. Axiom 4 (value from a power-law learning curve). The instantaneous value of data is proportional to the marginal performance gain it produces, and that gain decays as a power law in cumulative data.

Axiom 4 is the substantive one, and it is where the scaling-law literature enters. Write the model's expected loss after absorbing volume x as L(x), and suppose, in line with the observed scaling regularities, that the attainable performance gain $G(x) = L(0) - L(x)$ approaches a finite ceiling $G\infty$ with a power-law deficit, $G(x) = G\infty[1 - (1 + \nu(x/\kappa)^{\lambda})^{-1/\nu}]$ for scale $\kappa > 0$, sharpness $\lambda > 0$ and curvature $\nu > 0$. If the fee is value-based, $F(x) = \varphi + \kappa_p G(x)$ for a baseline φ and a price-per-unit-value κ_p, then F inherits the same bounded, concave, power-law-saturating shape. The functional form in Axiom 4 is not arbitrary: it is the most parsimonious family that is consistent with Axioms 1–3, reduces to the learning-curve forms used in the scaling literature in limiting cases, and admits a closed-form half-saturation point. Section 3 adopts exactly this form.

Two remarks situate the theory. First, it is normative as well as descriptive: it states what a value-based fee schedule ought to look like if data value follows the scaling laws, and departures from saturation in observed fees are therefore informative, possibly signalling non-value-based pricing, bundling or market power. Second, the theory is agnostic about the source of diminishing returns; whether it arises from the statistical learning curve, from redundancy and coverage saturation in the data, or from both, the macroscopic consequence for the fee schedule is the same.

3. The generalized data-saturation (GDS) model

3.1 Definition and special cases

Adopting the form motivated in Section 2, I model the expected fee as

$$E[F(x)] = \mu(x; \theta) = \varphi + (\alpha - \varphi)\, g(x; \kappa, \lambda, \nu)\,,$$

where φ is the baseline fee at zero volume, α is the saturation ceiling, and the saturation kernel g maps $[0, \infty)$ monotonically onto $[0, 1)$,

$$g(x; \kappa, \lambda, \nu) = 1 - [1 + \nu\,(x/\kappa)^{\lambda}]^{-1/\nu}\,,\ \kappa, \lambda, \nu > 0\,.$$

The three kernel parameters have direct interpretations: κ sets the scale of data volume over which saturation occurs, λ governs the sharpness with which value accrues, and ν controls the curvature of the approach to the ceiling, that is, the heaviness of the saturation tail. The model is valuable precisely because it unifies three families that are usually treated separately, each recovered as a special case (Table 1). Setting $\nu = 1$ yields the Hill function $g = x^{\lambda}/(\kappa^{\lambda} + x^{\lambda})$; setting $\nu = \lambda = 1$ yields the Michaelis–Menten form $g = x/(\kappa + x)$; and letting $\nu \to 0$ yields the stretched-exponential (Weibull-type) form $g = 1 - \exp\{-(x/\kappa)^{\lambda}\}$. Because these are nested, the data can adjudicate among the mechanisms rather than the analyst having to choose one in advance.

a

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

Table 1. Special cases of the GDS kernel g(x; κ, λ, ν)

Restriction	Resulting family	Kernel form	Origin
ν = 1, λ = 1	Michaelis–Menten	x / (κ + x)	Michaelis & Menten (1913)
ν = 1, λ free	Hill	x^λ / (κ^λ + x^λ)	Hill (1910)
ν → 0	Stretched-exponential	1 − exp{−(x/κ)^λ}	Weibull-type saturation
ν, λ free	Generalized (GDS)	1 − [1 + ν(x/κ)^λ]^(−1/ν)	this paper

Figure 1. The generalized data-saturation (GDS) kernel and its special cases

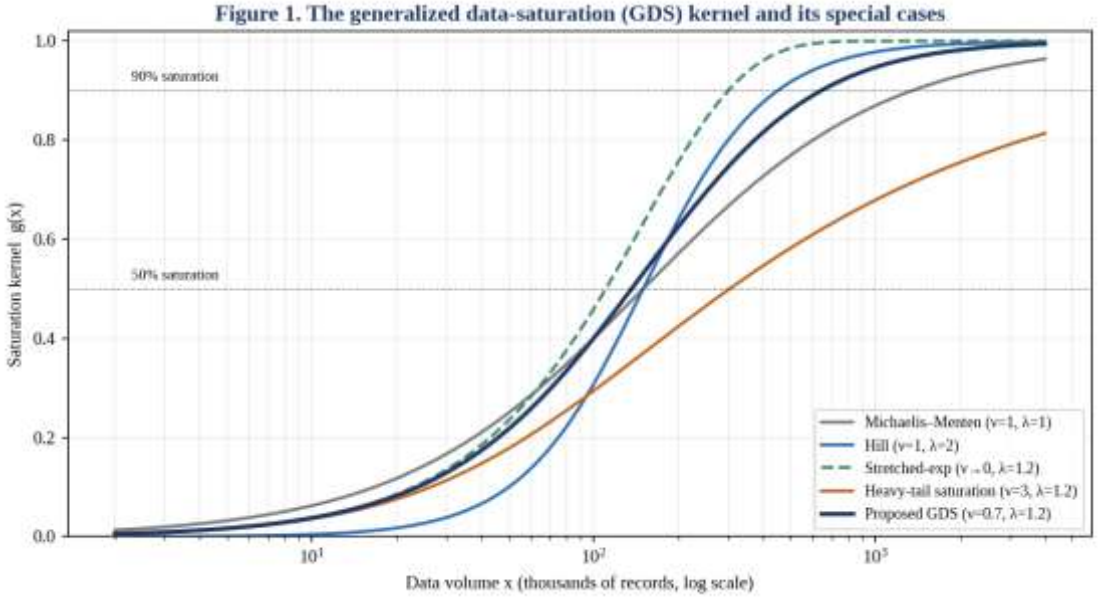

Figure 1. The GDS kernel for several parameter settings, showing how λ and ν interpolate between the Michaelis–Menten, Hill, stretched-exponential and heavy-tail saturation regimes.

Half-saturation, the saturation index and marginal fee

Solving g(x) = p for the volume at which a fraction p of the ceiling is reached gives a closed-form p-saturation volume,

$$x_p = \kappa \{ [(1 - p)^{-\nu} - 1] / \nu \}^{1/\lambda} ,$$

which specialises, when ν = 1, to the familiar half-saturation identity x_{0.5} = κ. I propose x_{0.9}, the volume at which ninety per cent of the ceiling fee is reached, as a saturation index summarising the practical point of diminishing returns; for the parameter values used in the simulation below it equals 643 (in thousands of records), against a half-saturation volume of 136. Differentiating the mean function gives the marginal fee and the fee elasticity,

$$\mu'(x) = (\alpha - \varphi) (\lambda u / x) (1 + \nu u)^{-(1+\nu)/\nu} , \; u = (x/\kappa)^{\lambda} , \; E(x) = x \mu'(x) / \mu(x) ,$$

which quantify, respectively, the incremental fee per additional unit of data and the percentage change in fee for a one-per-cent change in volume. Both decline toward zero as x grows, the formal expression of saturation, and are shown in Figure 4.

a

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

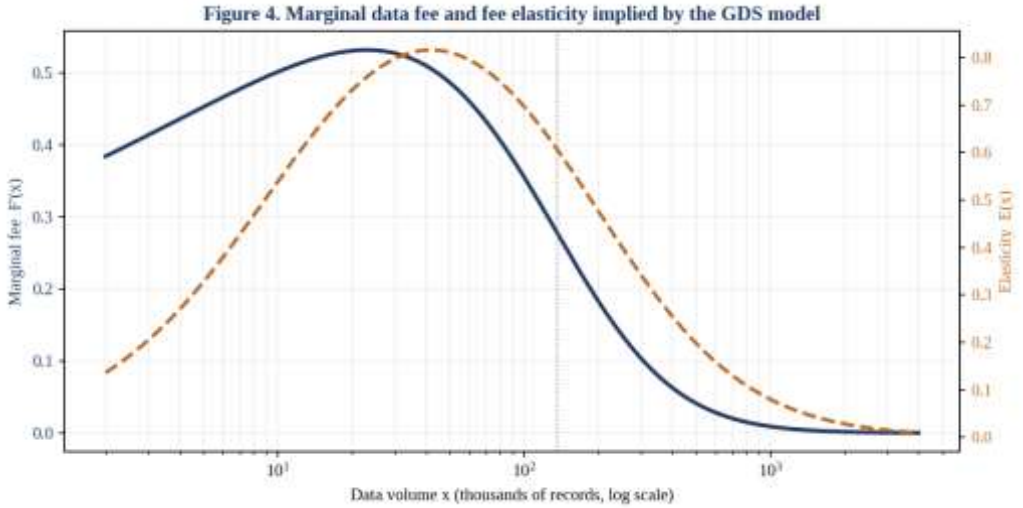

Figure 4. Marginal data fee μ'(x) and fee elasticity E(x) implied by the GDS model, both declining to zero as the schedule saturates.

Statistical inference

Model and estimation

Suppose fee observations (x_i, F_i), $i = 1, \ldots, n$, are generated as $F_i = \mu(x_i; \theta) + \varepsilon_i$ with $\theta = (\varphi, \alpha, \kappa, \lambda, \nu)^T$ and independent errors ε_i of mean zero and variance σ^2. The parameters are estimated by nonlinear least squares (NLS),

$$\hat{\theta} = arg\ min_{\theta}\ \Sigma\ (i=1 \ldots n)\ [\ F_i - \mu(x_i;\ \theta)\ ]^2\ ,$$

solved by Gauss–Newton or Levenberg–Marquardt iteration using the analytic Jacobian J with columns $\partial\mu/\partial\varphi = 1 - g$, $\partial\mu/\partial\alpha = g$, and $\partial\mu/\partial(\kappa, \lambda, \nu)$ obtained from $\partial g/\partial(\cdot)$; the kernel derivatives are given in the Appendix. Under the standard regularity conditions for nonlinear regression a compact parameter space, identifiability, and a Jacobian of full column rank at θ the NLS estimator is consistent and asymptotically normal (Jennrich, 1969; Wu, 1981; Seber & Wild, 1989; Bates & Watts, 1988), with

$$\sqrt{n}\ (\hat{\theta} - \theta) \rightarrow N(0,\ \sigma^2\ [n^{-1}\ J^T J]^{-1})\ ,$$

from which Wald confidence intervals and the standard errors used in Section 5 follow; profile-likelihood or bootstrap intervals (Efron & Tibshirani, 1993) are preferable when the curvature is appreciable, as it is for the shape parameters.

Identifiability

Two identifiability issues deserve emphasis, both inherent to flexible saturation models rather than peculiar to this one (Ratkowsky, 1983; Bates & Watts, 1988). First, α and κ are well identified only if the design spans the region around and beyond the half-saturation volume; a sample confined to the steeply rising part of the curve carries little information about the ceiling, and the ceiling and scale become correlated. Second, the sharpness λ and curvature ν jointly determine the shape near and beyond the inflection, and can partially trade off, so that separating them requires observations in the upper tail.

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

These features are not defects but guides to design: informative estimation of the full GDS model requires data extending well past x_{0.5} into the saturation tail.

Model selection

Because the special cases of Table 1 are nested within the GDS model, the appropriate family can be chosen by likelihood-ratio tests between nested pairs, or, more generally and including the non-nested log-linear alternative, by the Akaike and Bayesian information criteria (Akaike, 1974; Schwarz, 1978; Burnham & Anderson, 2002). The strategy I recommend, and illustrate in Section 5, is to fit the full GDS model and the special cases, compare them by AIC, and adopt the most parsimonious form not decisively rejected, reserving the extra parameter ν for data that genuinely require it.

Simulation study

Design

To assess whether the GDS parameters can be recovered from realistic samples, I conducted a Monte-Carlo study. Fee schedules were generated from the GDS model with true parameters $\varphi = 5$, $\alpha = 120$, $\kappa = 150$, $\lambda = 1.2$ and $\nu = 0.7$, on a logarithmically spaced design of data volumes from 2 to 4 000 (thousand records), with independent Gaussian errors. Sample sizes $n \in \{20, 40, 80\}$ and error standard deviations $\sigma \in \{3, 8\}$ were crossed, and 300 replications were run per cell. Each replicate was fitted by bounded NLS, and for every parameter I recorded the relative bias, the root-mean-square error and the empirical coverage of the nominal 95% Wald interval. All data are synthetic.

5.2 Results

The ceiling α, baseline φ and scale κ are recovered accurately across every design cell, with absolute relative bias under 5.0% and interval coverage close to the nominal 95% even at n = 20 (Table 2). The shape parameters behave as the identifiability analysis predicts. When the design is reasonably large or the noise modest, λ and ν are also recovered well; but at the smallest sample with the largest noise their bias grows and their intervals become conservative, reflecting weak identifiability rather than inconsistency the estimator remains centred but uncertain, and the widened intervals over-cover. This pattern, visible in the boxplots of Figure 3, is reported transparently because it carries a practical lesson: the location and height of the saturation curve can be pinned down from modest data, but resolving its precise curvature demands either more observations or lower noise, particularly in the tail.

Table 2. Simulation results: relative bias (%) and 95% interval coverage (%) by sample size and noise. Synthetic data; 300 replications per cell.

n	σ	Quantity	φ	α	κ	λ	ν
20	3	bias %	-2.3	0.1	2.6	1.4	5.0
		coverage %	94	94	95	91	91
40	3	bias %	-1.8	0.1	1.3	0.4	2.6
		coverage %	95	94	95	95	93
80	3	bias %	-0.9	0.2	0.2	0.8	4.1

a

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

n	σ	Quantity	φ	α	κ	λ	ν
		coverage %	95	95	94	95	95
20	8	bias %	-4.8	2.6	0.2	26.3	113.7
		coverage %	94	96	93	100	100
40	8	bias %	-3.7	2.1	-0.9	16.6	79.9
		coverage %	92	97	90	99	100
80	8	bias %	-0.6	0.9	1.8	5.3	24.5
		coverage %	94	96	95	98	100

Figure 3. Sampling distribution of GDS estimates (n = 40, 300 replications): the ceiling, baseline and scale are well recovered; the shape parameters λ, ν are weakly identified under noise

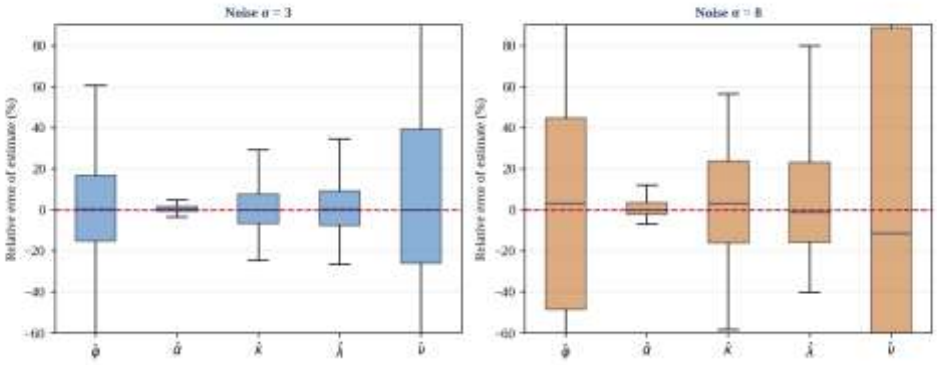

Figure 3. Sampling distributions of the GDS estimates at n = 40 over 300 replications, as relative errors. The ceiling, baseline and scale concentrate tightly on the truth; the shape parameters λ and ν are more dispersed, increasingly so at higher noise.

5.3 Model selection

To check that the framework selects sensible families, I fitted the GDS model and its special cases to representative synthetic datasets and compared them by AIC. When the generating curvature is close to the Hill value (ν near 1), the parsimonious Hill model is correctly preferred and the full GDS model is within two AIC units of it, while the Michaelis–Menten and log-linear forms are decisively rejected (Table 3, Scenario A). When the generating mechanism is genuinely stretched-exponential (small ν), the flexible kernel and the stretched-exponential special case dominate the Michaelis–Menten and Hill forms (Scenario B). The information criteria thus reward the extra parameter only when the data require it, exactly as intended, and the GDS family serves as a diagnostic umbrella under which the operative mechanism can be identified.

Table 3. Model selection by ΔAIC on representative synthetic datasets (0 = best). Scenario A: ν near 1; Scenario B: ν = 0.25.

Model (parameters)	Scenario A ΔAIC	Scenario B ΔAIC
GDS proposed (5)	1.6	1.3
Hill (4)	0.0	0.0
Stretched-exponential (4)	4.3	3.5
Michaelis–Menten (3)	19.2	32.8
Log-linear (2)	119.6	

6. Illustrative application (synthetic)

a

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

Figure 2 shows the GDS model fitted to a single representative synthetic fee schedule, alongside a Michaelis–Menten and a log-linear fit for comparison. The estimated parameters recover the generating values closely ($\hat{\varphi} = 5.2$, $\hat{\alpha} = 122.2$, $\hat{\kappa} = 140.8$, $\hat{\lambda} = 1.35$, $\hat{\nu} = 0.66$), and the fitted curve identifies a clear half-saturation volume and a ninety-per-cent saturation index beyond which additional data command negligible incremental fees. The Michaelis–Menten curve, constrained to $\nu = \lambda = 1$, cannot reproduce the sigmoidal rise and saturates too gently, while the log-linear schedule, lacking any ceiling, diverges in the tail and would extrapolate to implausibly high fees at large volumes. The contrast underlines the practical payoff of a bounded, shape-flexible model when the object of interest is precisely where and how fees stop rising.

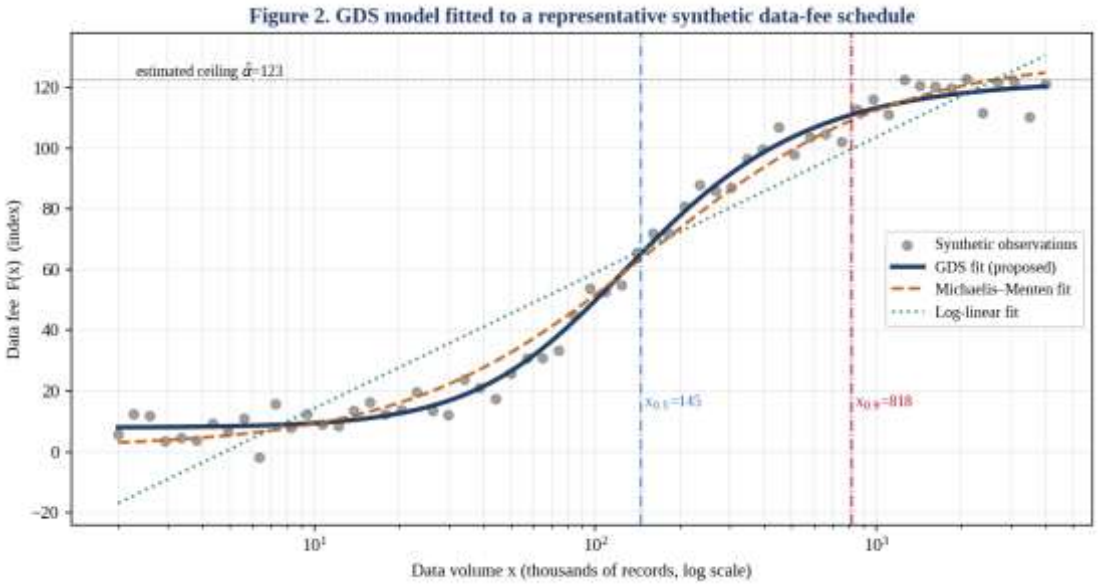

Figure 2. The GDS model (solid) fitted to a representative synthetic data-fee schedule, with Michaelis–Menten (dashed) and log-linear (dotted) fits, the estimated ceiling, and the half- and ninety-per-cent saturation volumes.

7. Discussion

The data-saturation theory and the GDS model together offer a compact way to think about, and to measure, the pricing of AI training data. The theory explains why saturation should be expected rather than merely noting that it occurs: if the value of data is tethered to the performance it produces, and performance follows the diminishing-returns scaling laws now well documented in machine learning, then a value-based fee schedule is necessarily bounded and concave. The model turns that qualitative expectation into estimable quantities a ceiling, a half-saturation volume, a saturation index, and an elasticity curve each of which has a clear interpretation for a buyer, a supplier or a regulator assessing whether observed fees are consistent with value-based pricing.

The unification the model provides is more than mathematical tidiness. Because Michaelis–Menten, Hill and stretched-exponential saturation are nested within it, an analyst need not commit in advance to one mechanism; the estimated curvature parameter ν locates the data within that spectrum, and the information criteria indicate whether the additional flexibility is warranted. The simulation evidence is

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

encouraging on the quantities that matter most for pricing the ceiling and the scale are recovered robustly while being candid about the cost of flexibility, namely that the curvature is resolved only with richer data. This is a familiar trade-off in nonlinear modelling, and the explicit identifiability analysis turns it into design guidance rather than a hidden hazard.

8. Limitations and future work

Several limitations frame the contribution. The empirical illustrations are entirely synthetic; the paper proposes and validates a model and a theory, but does not estimate them on observed market data, which would be the natural next step were a suitable dataset of volume–fee pairs to be assembled. The weak identifiability of the shape parameters under noise, documented above, means that confident statements about curvature require informative designs reaching into the saturation tail. The error structure assumed here is additive and homoscedastic; data fees may instead exhibit multiplicative or volume-dependent variance, which a weighted or log-scale formulation, or a generalized-least-squares extension, would accommodate. The model is also static, describing a single cross-section of volume against fee; where repeated schedules are observed across suppliers or over time, a nonlinear mixed-effects formulation with random saturation parameters (Pinheiro & Bates, 2000; Davidian & Giltinan, 1995) would borrow strength across units and quantify heterogeneity in saturation behaviour. Finally, the theory treats fees as value-based; relaxing that assumption to incorporate strategic or bundled pricing is a worthwhile theoretical extension. None of these qualifications undermines the core contribution, but each marks a direction in which it could be developed.

9. Conclusion

This paper has proposed a data-saturation theory for the pricing of artificial-intelligence training data, deriving the saturation of data fees from the diminishing-returns scaling laws that govern machine-learning performance, and has introduced the generalized data-saturation model to give that theory an estimable, interpretable and unifying statistical form. The model nests three classical saturation families, admits closed-form expressions for the half-saturation volume, the saturation index, the marginal fee and the elasticity, and is supported by a standard nonlinear-regression inferential framework whose behaviour was examined in a simulation study. The ceiling, baseline and scale of a data-fee schedule are shown to be recoverable from modest synthetic samples, while the finer curvature requires richer data a limitation stated plainly and turned into design guidance. The framework equips analysts with a principled vocabulary and a practical tool for describing where, and how sharply, the value of additional data saturates, and for judging whether observed fee schedules are consistent with the way AI systems actually learn.

Appendix. Kernel derivatives

a

StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
15th article, Pages: 278 – 287
DOI: https://doi.org/10.32493/sm.v8i2.60720

Writing u = (x/κ)^{λ} and h = 1 + ν u, the saturation kernel is g = 1 − h^{−1/ν}, and its partial derivatives, required for the Jacobian of μ(x; θ) = φ + (α − φ) g, are

$$\partial g/\partial x = (\lambda\, u / x)\, h^{-(1+\nu)/\nu} , \quad \partial g/\partial \kappa = -(\lambda\, u / \kappa)\, h^{-(1+\nu)/\nu} ,$$

$$\partial g/\partial \lambda = (u \ln(x/\kappa))\, h^{-(1+\nu)/\nu} , \quad \partial g/\partial \nu = h^{-1/\nu} \left[\nu^{-2} \ln h - (u)/(\nu\, h) \right] ,$$

together with ∂μ/∂φ = 1 − g and ∂μ/∂α = g. These expressions are continuous in the interior of the parameter space, and the Jacobian has full column rank when the design includes points on both sides of the half-saturation volume, which is the practical content of the identifiability conditions of Section 4.2.

References

1. Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.
2. Bates, D. M., & Watts, D. G. (1988). Nonlinear Regression Analysis and Its Applications. New York: Wiley.
3. Burnham, K. P., & Anderson, D. R. (2002). Model Selection and Multimodel Inference (2nd ed.). New York: Springer.
4. Davidian, M., & Giltinan, D. M. (1995). Nonlinear Models for Repeated Measurement Data. London: Chapman & Hall.
5. Draper, N. R., & Smith, H. (1998). Applied Regression Analysis (3rd ed.). New York: Wiley.
6. Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. New York: Chapman & Hall.
7. Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., … Zhou, Y. (2017). Deep learning scaling is predictable, empirically. arXiv:1712.00409.
8. Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. Journal of Physiology, 40(Suppl.), iv–vii.
9. Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., … Sifre, L. (2022). Training compute-optimal large language models. arXiv:2203.15556.
10. Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. Annals of Mathematical Statistics, 40(2), 633–643.
11. Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., … Amodei, D. (2020). Scaling laws for neural language models. arXiv:2001.08361.
12. Michaelis, L., & Menten, M. L. (1913). Die Kinetik der Invertinwirkung. Biochemische Zeitschrift, 49, 333–369.
13. Pinheiro, J. C., & Bates, D. M. (2000). Mixed-Effects Models in S and S-PLUS. New York: Springer.
14. Ratkowsky, D. A. (1983). Nonlinear Regression Modeling: A Unified Practical Approach. New York: Marcel Dekker.
15. Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), 461–464.
16. Seber, G. A. F., & Wild, C. J. (1989). Nonlinear Regression. New York: Wiley.
17. Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 843–852.
18. Wu, C. F. J. (1981). Asymptotic theory of nonlinear least squares estimation. Annals of Statistics, 9(3), 501–513.