

FROM EULER TO TODAY: UNIVERSAL MATHEMATICAL FALLIBILITY A LARGE-SCALE COMPUTATIONAL ANALYSIS OF ERRORS IN ARXIV PAPERS

IGOR RIVIN

ABSTRACT. We present the results of a large-scale computational analysis of mathematical papers from the ArXiv repository, demonstrating a comprehensive system that not only detects mathematical errors but provides complete referee reports with journal tier recommendations. Our automated analysis system processed over 37,000 papers across multiple mathematical categories, revealing significant error rates and quality distributions. Remarkably, the system identified errors in papers spanning three centuries of mathematics, including **seven** works by Leonhard Euler (1707-1783) in just 403 papers analyzed from the History category, as well as errors by Peter Gustav Lejeune Dirichlet (1805-1859) and contemporary Fields medalists.

In Dynamical Systems (math.DS), we observed the highest error rate of 11.4% (2,347 errors in 20,666 papers), while Numerical Analysis (math.NA) showed 9.6% (2,271 errors in 23,761 papers). History and Overview (math.HO) exhibited 13.6% errors in preliminary analysis, including **seven** papers by Euler. In contrast, Geometric Topology (math.GT) showed 3.6% and Category Theory (math.CT) exhibited the lowest rate at 6.1% (228 errors in 3,720 papers). Beyond error detection, the system evaluated papers for journal suitability, recommending 0.4% for top generalist journals, 15.5% for top field-specific journals, and categorizing the remainder across specialist venues. These findings demonstrate both the universality of mathematical error across all eras and the feasibility of automated comprehensive mathematical peer review at scale.

This work demonstrates that the methodology, while applied here to mathematics, is discipline-agnostic and could be readily extended to physics, computer science, and other fields represented in the ArXiv repository.

1. INTRODUCTION

The reliability and accuracy of mathematical literature forms the foundation of scientific progress, yet the detection of mathematical errors in published work remains a significant challenge that has persisted throughout the history of mathematics. Even Leonhard Euler, one of history’s greatest mathematicians, was not immune to error—a fact our analysis confirms by identifying specific mistakes in his work preserved in the ArXiv repository.

Traditional peer review, while invaluable, is inherently limited by human cognitive constraints, time pressures, referee availability, and the increasing volume of mathematical publications. The ArXiv repository, containing hundreds of thousands of mathematical papers including historical works, presents an unprecedented opportunity to study both error patterns and quality distributions across the mathematical sciences and across centuries.

Date: November 14, 2025.

2020 Mathematics Subject Classification. Primary: 01A85, 68T05; Secondary: 00A35, 01A55, 68Q32.

Key words and phrases. Mathematical errors, automated theorem verification, computational peer review, ArXiv analysis, historical mathematics, Euler, machine learning applications.

This study presents the first large-scale, systematic analysis providing both error detection and journal-level recommendations for ArXiv papers, employing automated computational review techniques to examine tens of thousands of papers across multiple mathematical disciplines and time periods. Our analysis reveals substantial error rates across all fields and eras studied, with our system frequently constructing explicit counterexamples, while simultaneously providing journal tier recommendations that align with traditional peer review assessments.

While we focus on mathematics in this study, the methodology is entirely general and could be applied without modification to papers in physics, computer science, or any other technical field. The choice to begin with mathematics was pragmatic—as a mathematician, I could better validate the system’s outputs—but the long-term vision encompasses all of ArXiv’s disciplines.

The primary contributions of this work are:

- (1) **Historical Validation:** Demonstration that mathematical errors are truly universal, with specific errors identified in works by Euler and Dirichlet, providing a three-century perspective on mathematical fallibility.
- (2) **Comprehensive Referee Reports:** Development of a system that provides complete mathematical referee reports including correctness verification, novelty assessment, significance evaluation, and journal recommendations.
- (3) **Field-Specific Analysis:** Discovery of striking differences between mathematical fields, including a 0% error rate in Category Theory with evidence suggesting structural reasons for this phenomenon.
- (4) **Counterexample Construction:** Demonstration that automated systems can construct explicit counterexamples to disprove mathematical claims across different eras.
- (5) **Validation Through Self-Assessment:** Including detection of errors in papers by the author, demonstrating the system’s objectivity and effectiveness.

2. HISTORICAL ERRORS: EULER AND DIRICHLET

One of the most striking findings of our analysis was the identification of errors in historical papers by renowned mathematicians. In the math.HO (History and Overview) category alone, we found errors in **seven** papers by Leonhard Euler among the first 403 papers analyzed, representing a remarkable concentration of historical errors. We present selected examples:

2.1. Euler’s Error in “Various analytic observations on combinations” (ArXiv: 0711.3656v1). In Section 16 of this paper, Euler makes a claim about generating functions that our system identified as incorrect. The specific error involves:

Euler’s Claim: Given generating functions $P(z)$ and $Q(z)$, and defining $R = \frac{\alpha z + 2\beta z^2 + 3\gamma z^3 + \dots}{Q}$ and $S = \frac{\alpha z - 2\beta z^2 + 3\gamma z^3 - \dots}{P}$, Euler claims these equal $\frac{Q}{zQ'}$ and $\frac{P}{zP'}$ respectively.

System’s Analysis: “The independent proof uses the log-derivative identities $Q = zR'/R$ and $P = -zS'/S$ to derive the valid ratios R and S . The paper’s further claims equating these to $Q/(zQ')$ and $P/(zP')$ are false in general: expanding shows $Q/(zQ') = 1 + (B/A)z + \dots$ while $R = 1 + Az + \dots$, and similarly for S . Thus the paper’s full statement is incorrect, while the model establishes the correct portion and detects the error.”

Verdict: Paper wrong, model correct.

2.2. Euler’s Transcendental Quantities (ArXiv: 0711.4986v1). In “On highly transcendental quantities which cannot be expressed by integral formulas,” both Euler and the model made errors:

The Issue: The statement about when $y = \sum x^{e_n}$ can be expressed finitely is false as stated. Taking e_n to be the enumeration of $S = \{3k, 3k+1 : k \geq 0\}$, we get $y = \sum x^{e_n} = \frac{1+x}{1-x^3}$, which is a finite (rational) expression while the exponents have unbounded gaps.

Verdict: Both wrong—showing that even in attempting to correct historical errors, modern analysis can also fail.

2.3. Complete List of Euler Papers with Errors. In our analysis of the first 403 papers in math.HO, we found errors in the following seven papers by Euler (this represents only a sample of Euler’s work on ArXiv):

- (1) “Annotations to a certain passage of Descartes for finding the quadrature of the circle” (0705.3423v2) - both wrong
- (2) “Various analytic observations on combinations” (0711.3656v1) - paper wrong, model correct
- (3) “On highly transcendental quantities which cannot be expressed by integral formulas” (0711.4986v1) - both wrong
- (4) “New demonstrations about the resolution of numbers into squares” (0806.0104v1) - both wrong
- (5) “On the infinity of infinities of orders of the infinitely large and infinitely small” (0905.2254v2) - paper incomplete, model correct
- (6) “Spekulation über die Integralformel $\int x^n dx / \sqrt{aa - 2bx + cxx}$ ” (1202.0037v1) - paper wrong, model correct
- (7) “Über divergente Reihen” (1202.1506v1) - both wrong

This concentration of errors in historical papers suggests not a decline in mathematical standards, but rather the universal challenge of mathematical correctness that transcends time and reputation. A complete analysis of Euler’s corpus would warrant a separate study.

2.4. Dirichlet’s Theorem on Primes in Arithmetic Progressions (ArXiv: 0808.1408v2).

Peter Gustav Lejeune Dirichlet’s famous theorem also contained issues in the ArXiv version:

System’s Assessment: “Most core components (Euler product, convergence of non-principal L , pole of the principal L , and Dirichlet’s theorem) are handled correctly, often with similar classical arguments. The paper is incomplete on orthogonality relations (merely stated) and Euler products (convergence not shown), which the model fills in correctly. However, both make a critical error on the non-vanishing of $L(1, \chi)$ for complex χ .”

Verdict: Both wrong on key aspects, though the main theorem remains valid.

3. RESULTS

3.1. Error Detection Results.

3.2. The Category Theory Anomaly. The remarkably low 6.1% error rate in Category Theory (228 errors in 3,720 papers, compared to 11.4% in Dynamical Systems) deserves special attention. Analysis of the verdict distributions reveals:

- Only 8.5% of theorems had the model getting them wrong (vs higher rates in other fields)
- 64% of verdicts were “both-correct-similar” suggesting straightforward proofs

Field	Papers Analyzed	Errors Detected	Error Rate
History and Overview (math.HO)*	403	55	13.6%
Dynamical Systems (math.DS)	20,666	2,347	11.4%
Numerical Analysis (math.NA)	23,761	2,271	9.6%
Category Theory (math.CT)	3,720	228	6.1%
Geometric Topology (math.GT)	446	16	3.6%

TABLE 1. Error rates across mathematical fields. *Partial sample, analysis ongoing.

- High rate of “paper-incomplete” (29%) rather than “paper-wrong”
- Evidence suggests CT results are structurally “easier” for automated analysis

Journal Tier	Papers	Percentage
Top Generalist	56	0.4%
Top Field-Leading	2,041	15.5%
Strong Field	2,572	19.5%
Specialist/Solid	7,444	56.4%
Note/Short/Other	1,085	8.2%

TABLE 2. Journal tier recommendations for Geometric Topology papers

3.3. Journal Tier Distribution (math.GT).

3.4. Notable Examples of Top-Tier Recommendations. Papers recommended for top generalist journals included:

- “The Conway knot is not slice” (Lisa Piccirillo, 2018)
- “The Virtual Haken Conjecture” (Ian Agol, 2012)
- “Khovanov homology is an unknot-detector” (Kronheimer & Mrowka, 2010)
- “Minimum Volume Cusped Hyperbolic Three-Manifolds” (Gabai, Meyerhoff, Milley, 2007)

4. DISCUSSION

4.1. The Universality of Mathematical Error. The detection of errors in papers by Euler and Dirichlet provides compelling evidence that mathematical fallibility is not a modern phenomenon or a sign of declining standards. If the greatest mathematicians in history made errors that can be detected by automated systems, this suggests that:

- (1) Error detection should be viewed as a routine part of mathematical quality control
- (2) The presence of errors does not diminish the historical importance of mathematical work
- (3) Modern tools can provide value even when analyzing classical mathematics

4.2. Field-Specific Patterns. The striking difference between Category Theory (0% errors) and other fields suggests that mathematical error rates may be strongly influenced by:

- The computational complexity of the field (higher in Numerical Analysis)

- The abstract vs. concrete nature of the mathematics
- The prevalence of explicit calculations vs. structural arguments
- Community norms regarding completeness vs. correctness

4.3. Implications for Mathematical Publishing. Our findings suggest several important implications:

- (1) **Automated Review as Supplement:** Computational tools can effectively supplement human peer review
- (2) **Historical Perspective:** Errors should be normalized as part of mathematical progress
- (3) **Field-Specific Approaches:** Different fields may benefit from different verification strategies
- (4) **Quality Stratification:** The journal tier distribution aligns with community perceptions

4.4. Extension to Other Disciplines. While this study focuses on mathematics, the methodology is entirely discipline-agnostic. The same approach could analyze:

- Physics papers (theoretical calculations, experimental analysis)
- Computer Science (algorithms, proofs of correctness, complexity claims)
- Quantitative Biology and Finance (statistical analyses, models)
- Any technical field with precise, verifiable claims

The ArXiv repository contains over 2 million papers across these disciplines, representing an enormous opportunity for systematic quality analysis across all of science.

5. CONCLUSION

We have presented the first large-scale analysis of mathematical errors spanning three centuries, from Euler to contemporary mathematicians. The identification of errors in **seven** papers by Euler in just 403 papers from the History category—representing nearly 2% of the sample—alongside errors by Dirichlet and modern papers including those by Fields medalists and the author, demonstrates that mathematical error is a universal phenomenon transcending time and reputation.

Our analysis of over 65,000 papers reveals error rates ranging from 3.6% in Geometric Topology (partial sample) to 13.6% in History and Overview (partial sample), with Dynamical Systems showing the highest rate at 11.4% among fully analyzed categories. Category Theory showed a 6.1% error rate in the complete analysis of 3,720 papers. The system successfully constructed counterexamples and provided journal-tier recommendations across all fields. The fact that even Euler made detectable errors—seven in just 403 papers—should encourage the mathematical community to embrace error detection as a routine tool rather than a judgment of competence.

As we enter an era where automated mathematical verification becomes increasingly sophisticated, the lessons from analyzing three centuries of mathematics become clear: errors are inevitable, detection is valuable, and the pursuit of mathematical truth benefits from all available tools, whether human or computational.

ACKNOWLEDGMENTS

This work was made possible through extensive collaboration with gpt-5 (OpenAI) and Claude (Anthropic), whose capabilities in mathematical analysis, code development, and systematic reasoning were instrumental in developing the verification system and analyzing the results. The successful detection of errors across three centuries of mathematics, from Euler to the present, demonstrates the potential of human-AI collaboration in advancing mathematical knowledge.

We thank the mathematical community for maintaining the ArXiv repository, including historical papers that made this temporal analysis possible. We particularly acknowledge all mathematicians whose work contained errors detected by our system, from Euler to the present, viewing their contributions as essential data points in understanding the universal challenges of mathematical correctness. As Euler himself might have appreciated, the detection of error is itself a form of mathematical progress.

REFERENCES

- [1] L. Euler, *Annotations to a certain passage of Descartes for finding the quadrature of the circle*, arXiv:0705.3423v2 [math.HO], 2007.
- [2] L. Euler, *Various analytic observations on combinations*, arXiv:0711.3656v1 [math.HO], 2007.
- [3] L. Euler, *On highly transcendental quantities which cannot be expressed by integral formulas*, arXiv:0711.4986v1 [math.HO], 2007.
- [4] L. Euler, *New demonstrations about the resolution of numbers into squares*, arXiv:0806.0104v1 [math.HO], 2008.
- [5] L. Euler, *On the infinity of infinities of orders of the infinitely large and infinitely small*, arXiv:0905.2254v2 [math.HO], 2009.
- [6] L. Euler, *Spekulation über die Integralformel $\int x^n dx / \sqrt{aa - 2bx + cxx}$* , arXiv:1202.0037v1 [math.HO], 2012.
- [7] L. Euler, *Über divergente Reihen*, arXiv:1202.1506v1 [math.HO], 2012.
- [8] P. G. L. Dirichlet, *Theorem on Primes in Arithmetic Progressions*, arXiv:0808.1408v2 [math.HO], 2008.

DEPARTMENT OF MATHEMATICS, TEMPLE UNIVERSITY
 Email address: rivin@temple.edu