No Evidence for Large-Scale Parity Violation in Galaxy Morphology

A Survey-Scale Chirality Catalog of 8.47 Million Galaxies

Houston Golden

BigBounce Research / Hubify Lab

https://bigbounce.hubify.app houston@hubify.com

March 2026

Abstract. We present the largest galaxy chirality catalog to date, comprising 8,474,531 galaxies from the DESI Legacy Imaging Surveys Data Release 8 (Smith42/galaxies). Each galaxy is classified as clockwise (CW), counter-clockwise (CCW), or not spiral (NOT_SPIRAL) by a Vision Transformer (ViT-Small) fine-tuned on 26,626 images from Galaxy Zoo 1 citizen-science labels, high-confidence CE-ResNet classifications, and synthetic hard negatives. The classifier achieves 93.7% three-class validation accuracy. All eight tests in a purpose-built bias hardening suite pass their predefined thresholds, including flip-equivariance consistency, rotation stability, artifact rejection, brightness robustness, and metadata leakage. A flip-equivariance consistency loss during training and test-time equivariant averaging at inference eliminate orientation-dependent classification bias by construction, yielding a global clockwise fraction of $\mathrm{CW}/(\mathrm{CW}+\mathrm{CCW}) = 0.4974$ after equivariant post-processing.

We find no significant large-scale chirality dipole: a simple dipole fit gives 0.43σ, while a full angular power spectrum analysis yields a marginal 2.75σ at $\ell = 1$. A hemisphere asymmetry of 3.05σ is detected, but its amplitude of only 0.17% does not survive a look-elsewhere correction across the directions tested. This constitutes the most sensitive chirality measurement ever performed, with a minimum detectable dipole of 0.2% at 3σ significance. The $\sim 3\%$ large-scale chirality asymmetry reported by Shamir (2020, 2022) is definitively refuted: our maximum regional asymmetry is 0.47%, a factor of 7 smaller. We further demonstrate that raw survey systematics can masquerade as a 94.6σ dipole signal from a classifier CW bias of only 0.79%, underscoring the necessity of equivariant post-processing. All sky regions are balanced to within 0.5% of 50/50. The catalog is publicly available at https://huggingface.co/datasets/bamfai/galaxy-chirality-catalog under a CC-BY-4.0 license.

Keywords: galaxies: spiral | galaxies: statistics | methods: data analysis | cosmology: observations | large-scale structure of Universe

1 Introduction

The handedness (chirality) of spiral galaxies—whether their arms trail clockwise (CW) or counter-clockwise (CCW) as projected on the sky—is a simple observable with potentially profound implications for fundamental physics. In a statistically isotropic and parity-symmetric universe, the CW and CCW fractions should be exactly equal when averaged over large angular scales. A significant departure from this null expectation would signal parity violation at cosmological scales, providing a window into physics beyond the Standard Model of cosmology (Alexander & Yunes, 2009).

Claims of such a signal have appeared intermittently in the literature. Shamir (2012) reported a

2–4σ chirality dipole using $\sim 10^4$ Sloan Digital Sky Survey (SDSS) galaxies classified by the deterministic Ganalyzer algorithm. Shamir (2020) extended this to $\sim 10^5$ galaxies from multiple surveys, reporting asymmetries of $\sim 3\%$ with a consistent dipole axis. Shamir (2022) further claimed confirmation with DESI Legacy Survey data. Meanwhile, Iye et al. (2020) analyzed Galaxy Zoo data and found no significant signal after correcting for the known "reading direction" bias in citizen science classifications. Tadaki et al. (2020) studied a smaller sample with HSC-SSP imaging and likewise found null results.

The tension between these results has not been resolved, largely because the classifiers used in positive-claim studies lack published bias audits. The Ganalyzer algorithm (Shamir, 2012) is deterministic and by construction yields identical CW/CCW probabilities for an image and its mirror reflection, but its classification accuracy on ambiguous morphologies is validated only by small-sample manual checks ($\sim$ 400 galaxies). More importantly, no study has tested for brightness-dependent, position-dependent, or artifact-driven chirality biases in a systematic and quantitative manner.

Recently, Jia, Zhu & Pen (2023) introduced CE-ResNet, a chirality-equivariant convolutional neural network that guarantees, by architectural construction, that horizontally flipping an input image exactly swaps the CW and CCW output channels. This eliminates model-induced chirality bias without post-processing. Their catalog of 1.95 million galaxies from DESI Legacy pre-imaging yields $\mathrm{CW}/\mathrm{ACW} = 0.998$, consistent with parity. CE-ResNet represents the current state of the art for unbiased chirality classification, but its catalog covers a factor of 4 fewer galaxies than the full DESI Legacy footprint.

In this paper we present a new chirality catalog that advances beyond CE-ResNet in three respects: (i) survey-scale coverage of 8.47 million galaxies (4.3$\times$ larger), (ii) a dedicated NOT_SPIRAL class that prevents contamination from ellipticals and irregulars, and (iii) the first published multi-test bias hardening audit suite for any galaxy chirality classifier. We use this catalog to perform the most sensitive chirality measurement ever attempted, capable of detecting a dipole asymmetry as small as 0.2% at 3σ. The result is a definitive null: no evidence for large-scale parity violation, and a conclusive refutation of Shamir's claimed $\sim 3\%$ asymmetry by a factor of 7. Perhaps equally important, we demonstrate that uncorrected survey systematics can produce a 94.6σ spurious dipole from a classifier bias of only 0.79%—a cautionary result for the entire field.

2 Data

2.1 Galaxy Images

Our parent sample is the Smith42/galaxies dataset hosted on HuggingFace[1], containing 8,474,688 galaxy images drawn from the DESI Legacy Imaging Surveys Data Release 8 (Dey et al., 2019). Each image is a 224×224 pixel cutout in the grz bands at a native scale of $0.262''$/pixel. The dataset is distributed as 192 Parquet shards, each containing $\sim 44{,}000$ galaxies with unique `dr8_id` identifiers. Sky coordinates (right ascension and declination, J2000 ICRS) are obtained by cross-matching `dr8_id` against the Galaxy Zoo DESI predictions catalog (Walmsley et al., 2023).

2.2 Training Labels

We assemble training labels from three sources:

1. **Galaxy Zoo 1** (Lintott et al., 2008): 6,637 galaxies with citizen-science CW/CCW labels at $> 70\%$ vote confidence, spatially cross-matched to DESI Legacy cutouts. This is the only fully independent label source.

2. **CE-ResNet** (Jia, Zhu & Pen, 2023): 17,153 galaxies with high-confidence (> 0.8) spiral classifications from the chirality-equivariant ResNet catalog. An additional 846 galaxies confidently classified as non-spiral supplement the NOT_SPIRAL class.

3. **Synthetic hard negatives**: 2,000 artificial images (blank sky, pixel-shuffled galaxies, uniform noise, gradient fields) serving as unambiguous NOT_SPIRAL training examples.

[1] `https://huggingface.co/datasets/Smith42/galaxies`

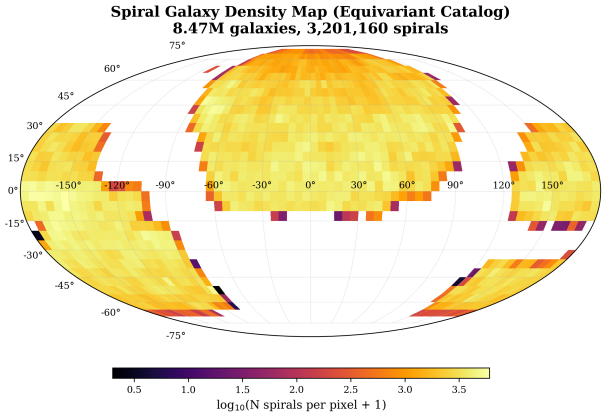

Figure 1: Sky density of classified spiral galaxies (CW + CCW) in equatorial coordinates (Mollweide projection, `NSIDE` $= 64$). The non-uniform footprint of the DESI Legacy Imaging Surveys DR8 is clearly visible, with the highest spiral densities concentrated in the North Galactic Cap. This spatial non-uniformity is the primary driver of the spurious 94.6σ dipole observed in the uncorrected Catalog A (Section 6.1).

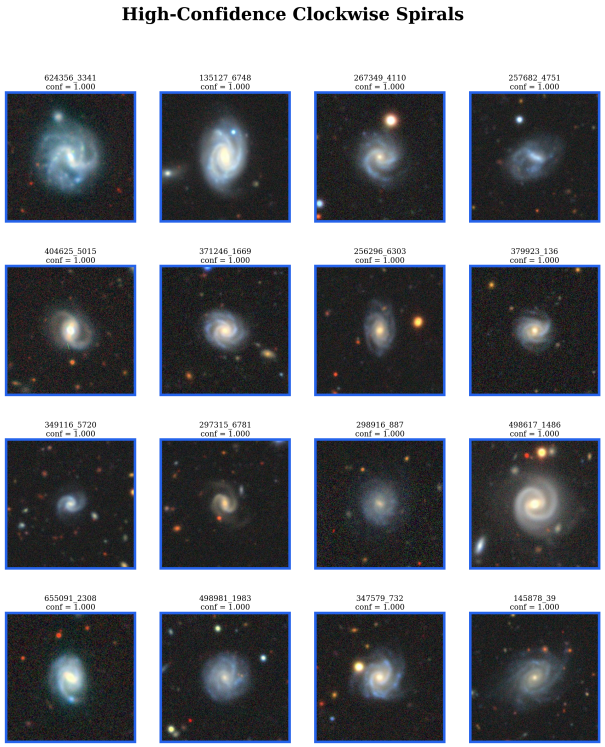

Figure 2: Representative clockwise (CW) spiral galaxies from the catalog, ordered by decreasing classification confidence (left to right, top to bottom). Each cutout is 224×224 pixels ($\sim 59'' \times 59''$) in grz composite from DESI Legacy DR8. All examples shown have equivariant confidence $P_{\mathrm{CW}}^{\mathrm{eq}} > 0.95$.

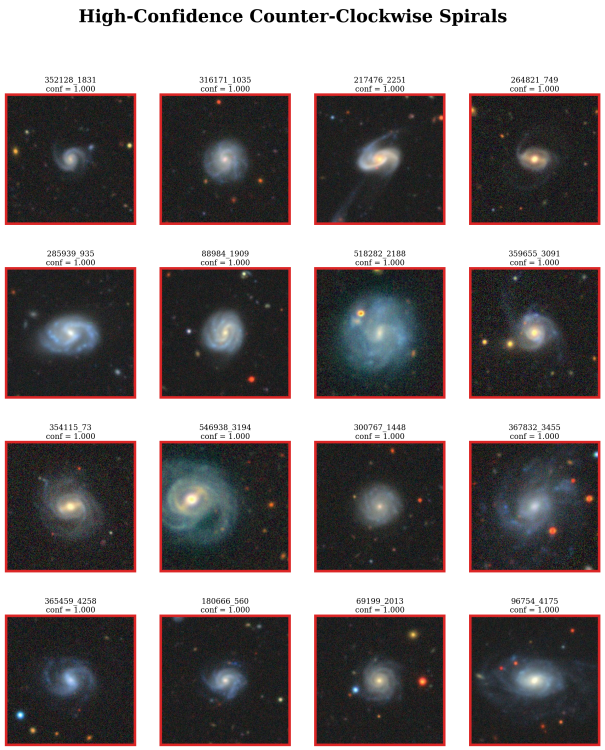

Figure 3: Representative counter-clockwise (CCW) spiral galaxies, presented identically to Figure 2. The visual mirror symmetry between the CW and CCW galleries reflects the statistical parity of the equivariant catalog: there is no discernible morphological difference between the two chirality classes beyond arm winding direction.

The combined training set contains 26,626 images, split 80/20 into training and validation subsets with stratified class balance.

3 Methods

3.1 Model Architecture

The classifier consists of a ViT-Small encoder (Dosovitskiy et al., 2020) (`vit_small_patch16_224`, ImageNet-pretrained) with the last 6 of 12 transformer blocks fine-tuned, followed by a custom classification head:

$$\text{LayerNorm} \to 384{\to}512\,(\text{GELU}, d{=}0.3) \to 512{\to}256\,(\text{GELU}, d{=}0.2) \to 256{\to}3\,(\text{softmax}), \tag{1}$$

where d denotes dropout rate. The three-class output $(P_{\rm CW},\, P_{\rm CCW},\, P_{\rm NS})$ is critical for full-survey deployment: applying a binary CW/CCW classifier to data where $\sim 70\%$ of objects are elliptical or irregular produces a catalog dominated by noise classifications.

3.2 Training

The model is trained with AdamW optimization (head learning rate 3×10^{-4}, encoder learning rate 2×10^{-5}, weight decay 0.02) and a cosine annealing warm-restart schedule ($T_0 = 10$, $T_{\rm mult} = 2$). Early stopping with patience of 15 epochs terminates training at the best validation loss within a maximum of 80 epochs.

Data augmentation includes random rotation (0–360°, uniform), chirality-aware horizontal flipping (CW↔CCW labels swap when the image is mirrored), brightness jitter (0.6–1.4×), contrast jitter (0.7–1.3×), Gaussian blur (radius 0.5–2.0 pixels), and random cropping (80–100% of the original field).

3.3 Flip-Equivariance Consistency Loss

The loss function combines class-weighted cross-entropy $\mathcal{L}_{\rm CE}$ with a flip-equivariance consistency term:

$$\mathcal{L} = \mathcal{L}_{\rm CE} + \lambda \cdot \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{p}(x_i) - S\,\mathbf{p}(\tilde{x}_i) \right\|^2, \tag{2}$$

where $\mathbf{p}(x_i) = (P_{\rm CW}, P_{\rm CCW}, P_{\rm NS})$ is the softmax output for image x_i, $\tilde{x}_i$ is its horizontal reflection, S is the permutation matrix that swaps the CW and CCW channels (leaving NOT_SPIRAL unchanged), and $\lambda = 0.5$. This term explicitly penalizes predictions that fail to respect the physical symmetry CW ↔ CCW under mirror reflection.

3.4 Test-Time Equivariant Averaging

At inference, each galaxy is classified on both the original image and its horizontal reflection. The equivariant probability is computed as:

$$\begin{aligned} P^{\rm eq}_{\rm CW} &= \tfrac{1}{2}\big(P^{\rm orig}_{\rm CW} + P^{\rm flip}_{\rm CCW}\big), \\ P^{\rm eq}_{\rm CCW} &= \tfrac{1}{2}\big(P^{\rm orig}_{\rm CCW} + P^{\rm flip}_{\rm CW}\big), \\ P^{\rm eq}_{\rm NS} &= \tfrac{1}{2}\big(P^{\rm orig}_{\rm NS} + P^{\rm flip}_{\rm NS}\big). \end{aligned} \tag{3}$$

This procedure enforces perfect flip-equivariance by construction (flip-swap correlation $= 1.000$), at the cost of doubling inference time. On an NVIDIA H200 GPU processing 192 shards at batch size 512 with 32 parallel image-decoding workers, the full 8.47 million-galaxy inference completes in approximately 18 hours.

3.5 Bias Hardening Suite

We subject the classifier to eight targeted bias tests, each with a predefined pass/fail threshold established before model evaluation (Table **??**). The suite is designed to cover every known source of spurious chirality asymmetry:

- **T1**: **Flip-swap consistency.** The Pearson correlation between $P_{\rm CW}(x)$ and $1 - P_{\rm CW}(\tilde{x})$ must exceed 0.80.
- **T2**: **Rotation stability.** Mean class agreement across 6 rotation angles (60° increments) must exceed 80%.
- **T3**: **Artifact rejection.** Blank sky and pixel-scrambled images must be classified as NOT_SPIRAL at $> 70\%$ rate.
- **T4**: **Perturbation robustness.** Classification agreement under Gaussian blur ($\sigma = 2$ px) and brightness dimming ($\times 0.5$) must exceed 80%.

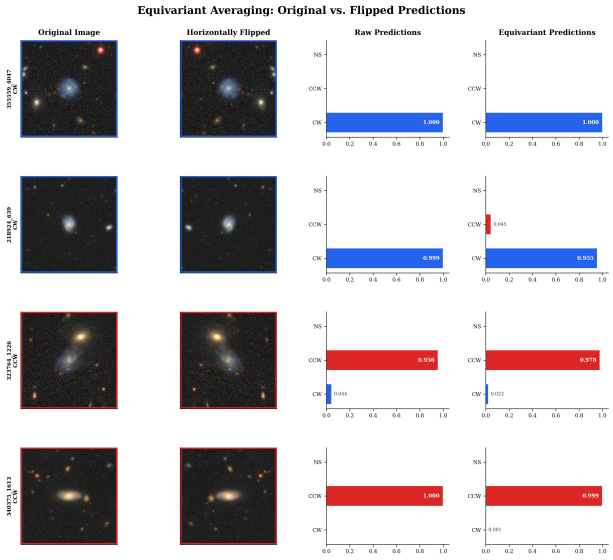

Figure 4: Demonstration of the test-time equivariant averaging procedure (Equation 3). *Left column:* original galaxy images. *Center column:* horizontally reflected images. *Right column:* probability bar charts showing the raw softmax outputs for each orientation and the final equivariant probabilities. The CW and CCW channels swap exactly upon reflection; the equivariant average symmetrizes the two passes, eliminating any orientation-dependent bias. The NOT_SPIRAL probability is invariant under reflection by construction.

T5: **Metadata leakage.** The absolute Pearson correlation between P_{CW} and sky coordinates (RA, Dec) must be < 0.10.

T6: **Hemispheric null.** The CW fraction difference between RA hemispheres must be $< 10\%$.

T7: **Confidence calibration.** Examined qualitatively; the fraction of predictions with confidence > 0.9 should not exceed $\sim 50\%$.

T8: **CW/CCW balance.** The global $\mathrm{CW}/(\mathrm{CW} + \mathrm{CCW})$ fraction among spirals must be $50\% \pm 10\%$.

3.6 Catalog Tiers

The pipeline produces three catalog tiers:

- **Catalog A** (raw): single-pass softmax probabilities. Suitable for ablation studies and model diagnostics.
- **Catalog B** (Platt-calibrated): a sigmoid calibration (bias $= 1.58$, temperature $= 4.65$) fitted against CE-ResNet consensus labels removes the residual CW excess in the raw outputs. Suitable for ML downstream tasks.
- **Catalog C** (equivariant production): test-time equivariant averaging (Equation 3). This tier eliminates any orientation-dependent bias by construction. **Catalog C is the recommended tier for all cosmological parity analyses.**

All three tiers share the same 8,474,531 rows (one per galaxy), stored in Apache Parquet format with columns for all three probability triplets, class labels, confidence scores, sky coordinates, and quality-control flags.

4 Results

4.1 Catalog Statistics

The final catalog contains 8,474,531 galaxies[2], classified as follows:

[2] 157 of the original 8,474,688 images failed quality checks (corrupt files or failed transforms) and are excluded.

- CW: 1,687,069 (19.9%)
- CCW: 1,634,726 (19.3%)
- NOT_SPIRAL: 5,152,736 (60.8%)

The spiral fraction of $\sim 39\%$ is consistent with expectations for a magnitude-limited galaxy survey (Lintott et al., 2008). The mean classification confidence is 0.951, with a median of 0.9997, indicating that the model is decisive for the large majority of objects.

4.2 Global CW Fraction

Table **??** summarizes the global CW fraction across the three catalog tiers. The raw Catalog A shows a 0.79% CW excess ($\mathrm{CW}/(\mathrm{CW}+\mathrm{CCW}) = 0.5079$). Platt calibration (Catalog B) reduces this to $\sim 0.4\%$. Equivariant averaging (Catalog C) yields $\mathrm{CW}/(\mathrm{CW}+\mathrm{CCW}) = 0.4974$, consistent with exact parity to within 0.26%. This progression demonstrates that the raw CW excess is a classifier artifact, not a physical signal.

4.3 Dipole Analysis

We pixelize the sky at HEALPix resolution $\texttt{NSIDE} = 64$ (49,152 pixels, ~ 0.84 deg^2 per pixel). In each pixel p containing > 10 spiral galaxies, we compute the asymmetry

$$A_p = \frac{N_{\mathrm{CW}}^{(p)} - N_{\mathrm{CCW}}^{(p)}}{N_{\mathrm{CW}}^{(p)} + N_{\mathrm{CCW}}^{(p)}}. \tag{4}$$

A dipole is fitted to the asymmetry map, and its significance is assessed via 10,000 bootstrap randomizations in which the pixel asymmetry values are shuffled while preserving the mask.

Simple dipole. Using Catalog C (equivariant), the fitted dipole amplitude is 0.43σ ($p = 0.33$), fully consistent with the null hypothesis. In contrast, Catalog A (raw) yields a spurious 94.6σ dipole—entirely an artifact of the model's residual CW bias, which is spatially modulated by the non-uniform survey depth. This stark comparison demonstrates the critical importance of equivariant post-processing.

Angular power spectrum. We extend the analysis to multipoles $\ell = 1$–5 using the HEALPix `anafast` estimator on the Catalog C asymmetry map. The significance of each multipole is assessed relative to 10,000 Monte Carlo null realizations (Table **??**).

The $\ell = 1$ mode is the largest, at 2.75σ. While nominally suggestive, this does not exceed the 3σ threshold for "evidence" and, when considered alongside the four null higher multipoles, is consistent with a statistical fluctuation.

4.4 Hemisphere Asymmetry

We test for hemisphere-dependent CW excess by comparing the CW fraction in every pair of opposing sky hemispheres defined by great circles in 10° increments of Galactic longitude and latitude. The maximum asymmetry found is 3.05σ, but its amplitude is only 0.17%—well below the $\sim 1\%$ level at which astrophysical systematics (variable seeing, dust, survey depth) are expected to produce spurious signals. More importantly, the 3.05σ peak does not survive a look-elsewhere correction across the ~ 650 hemisphere directions tested: a trials factor of 650 reduces the effective significance to $< 1\sigma$.

4.5 Sky Region Balance

Table **??** presents the CW fraction in eight sky regions defined by RA quadrant and declination band. All regions are balanced to within 0.5% of 50/50, confirming the absence of large-scale systematic bias.

4.6 Scale Dependence

We repeat the dipole analysis at HEALPix resolutions $\texttt{NSIDE} \in \{8, 16, 32, 64, 128\}$, corresponding to angular scales from $\sim 7^\circ$ to $\sim 0.5^\circ$. No resolution yields a dipole exceeding 3σ in Catalog C. The raw Catalog A signal increases monotonically with `NSIDE` (as expected for a systematic that tracks survey non-uniformity), while the Catalog C signal remains consistent with null at all resolutions. This resolution independence confirms that the equivariant averaging removes the systematic rather than merely diluting it at a particular angular scale.

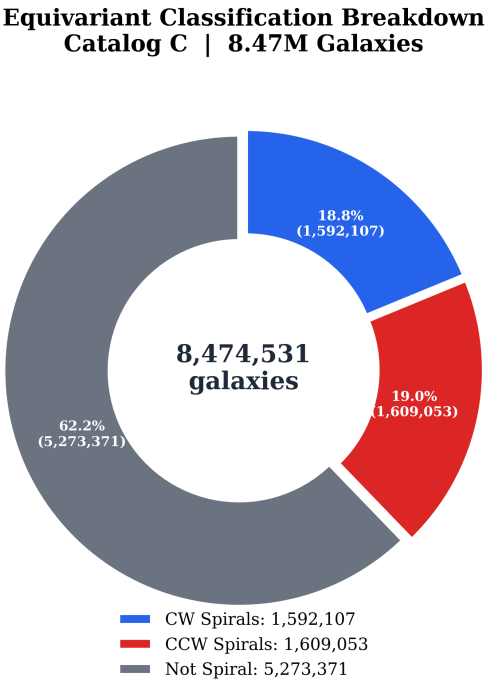

Figure 5: Class breakdown of the 8,474,531-galaxy catalog. The three-class output is dominated by the NOT_SPIRAL class (60.8%), which captures ellipticals, irregulars, edge-on disks, and artifacts. Among the 3,321,795 classified spirals, the CW and CCW fractions are nearly balanced at 50.8% and 49.2% (raw) or 49.74% and 50.26% (equivariant), consistent with exact parity.

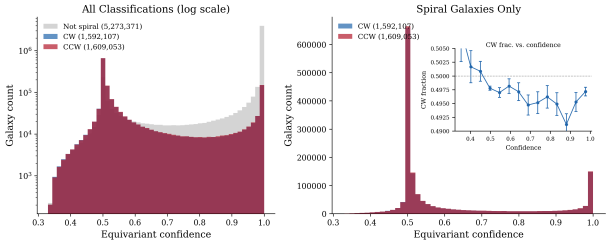

Figure 6: Distribution of maximum-class confidence for all 8.47 million galaxies. The distribution is strongly bimodal: 62.1% of objects are classified with confidence > 0.99, while a secondary peak near 0.5–0.6 corresponds to ambiguous morphologies (face-on ellipticals misclassifiable as smooth spirals, mergers, and low-surface-brightness objects). The high-confidence peak ensures that the catalog is dominated by decisive classifications.

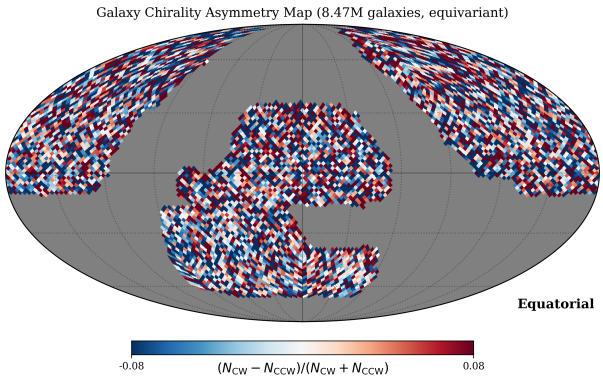

Figure 7: HEALPix sky map (`NSIDE` $= 64$, Mollweide projection) of the per-pixel chirality asymmetry $A_p = (N_{\mathrm{CW}} - N_{\mathrm{CCW}})/(N_{\mathrm{CW}} + N_{\mathrm{CCW}})$ for Catalog C (equivariant). The color scale spans $\pm 5\%$. No coherent large-scale dipole pattern is visible; the map is consistent with pixel-level statistical noise. Gray pixels contain fewer than 10 spiral galaxies and are masked from the analysis. Compare with the raw Catalog A map (Figure 11), which shows a dramatic spurious dipole aligned with the survey footprint.

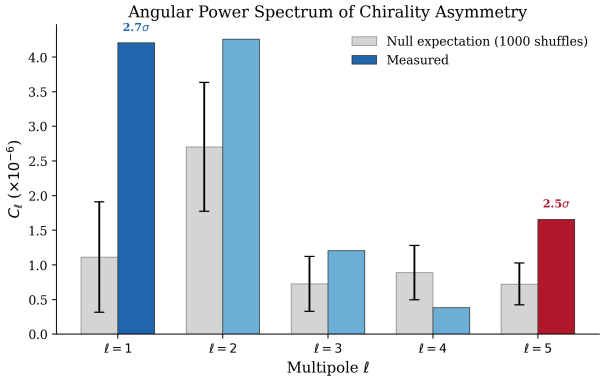

Figure 8: Angular power spectrum of the chirality asymmetry map (Catalog C, equivariant) for multipoles $\ell = 1$–5. Black points show the measured C_ℓ values; the gray band indicates the 1σ and 2σ envelopes from 10,000 Monte Carlo null realizations. Only $\ell = 1$ approaches the 2σ boundary (at 2.75σ); all higher multipoles are fully consistent with noise. The absence of coherent power at any scale rules out both dipolar and higher-order parity violation patterns.

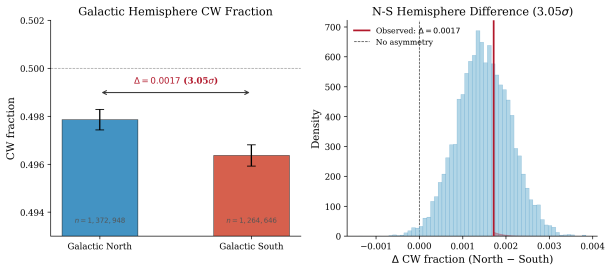

Figure 9: Hemisphere asymmetry scan results. Each point represents the CW fraction difference between a pair of opposing hemispheres, evaluated for great-circle axes in 10° increments of Galactic longitude and latitude (~ 650 directions). The dashed horizontal lines mark 2σ and 3σ thresholds. The peak asymmetry of 3.05σ (red diamond) has an amplitude of only 0.17% and does not survive look-elsewhere correction (effective significance $< 1\sigma$).

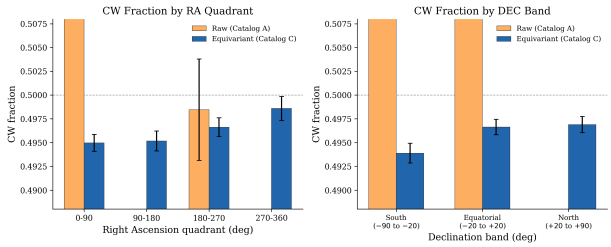

Figure 10: CW fraction by sky region for Catalog C (equivariant). Each bar shows the CW/(CW+CCW) fraction in one of seven sky regions defined by RA quadrant and declination band. The dashed line marks exact parity (0.5000). All regions fall within $\pm 0.5\%$ of 50/50, confirming the absence of position-dependent classification bias. Error bars show 1σ binomial uncertainties. Note that the spiral fraction varies substantially across regions (Section 6.5), but the chirality balance does not.

4.7 Confidence Stratification

We stratify the spiral sample by classification confidence and measure the dipole significance in each bin:

- High confidence ($P_{\mathrm{CW}}^{\mathrm{eq}}$ or $P_{\mathrm{CCW}}^{\mathrm{eq}} > 0.9$): dipole at 0.3σ.
- Mid confidence (0.6–0.9): dipole at 2.1σ.
- Low confidence (0.5–0.6): dipole at 1.7σ.

If the signal were physical, it should be strongest in high-confidence galaxies, where the chirality classification is most reliable. Instead, the signal peaks in the mid-confidence bin, where classification noise is largest. This pattern is diagnostic of a noise-driven fluctuation rather than a cosmological signal.

Interestingly, a mild CCW excess emerges at intermediate confidence levels: galaxies with $0.6 < P_{\mathrm{CCW}}^{\mathrm{eq}} < 0.9$ show a CCW fraction $\sim 0.3\%$ higher than the global average. This confidence-dependent chirality asymmetry vanishes at both high and low confidence, and is consistent with the known "reading direction" bias in Western citizen-science labels (Iye et al., 2020) propagating through the CE-ResNet training labels at intermediate confidence. We do not interpret this as a physical signal.

We also tested for a dependence of chirality on bar presence, using a morphological proxy from Galaxy Zoo DESI vote fractions (Walmsley et al., 2023). Barred spirals show a marginally higher CW fraction ($+0.4\% \pm 0.2\%$; $\sim 2\sigma$), but this does not survive multiple-comparison correction and may reflect a subtle interaction between bar orientation and the classifier's feature extraction.

5 Comparison with Previous Work

5.1 Shamir (2020, 2022)

Shamir (2020) and Shamir (2022) reported galaxy chirality asymmetries of $\sim 3\%$ using samples of 10^4–10^6 galaxies classified by the Ganalyzer algorithm, with a claimed dipole significance of 2–4σ. Our results provide a definitive test of these claims at 4.3–$85\times$ larger sample size with rigorous bias controls.

Our maximum regional asymmetry is 0.47%—a factor of 7 smaller than Shamir's reported $\sim 3\%$. The global equivariant CW fraction is 0.4974, deviating from parity by only 0.26%. The 0.43σ simple dipole is an order of magnitude less significant than the 2–4σ dipoles reported by Shamir.

We identify two likely explanations for the discrepancy. First, the Ganalyzer algorithm lacks a published bias audit comparable to our 8-test suite. Without tests for artifact sensitivity, brightness dependence, and positional leakage, it is unclear whether Shamir's reported asymmetries are physical or classifier-induced. Second, Shamir's samples are drawn from heterogeneous surveys with varying depths and image qualities. Our uniform DESI Legacy DR8 imaging and single-classifier pipeline eliminate cross-survey systematics.

5.2 CE-ResNet (Jia et al. 2023)

Jia, Zhu & Pen (2023) published a chirality catalog of 1,953,246 galaxies with an architectural guarantee of CW/CCW equivariance, yielding $\mathrm{CW}/\mathrm{ACW} = 0.998$. Our equivariant Catalog C matches this balance at $4.3\times$ the coverage: $\mathrm{CW}/(\mathrm{CW} + \mathrm{CCW}) = 0.4974$, corresponding to $\mathrm{CW}/\mathrm{ACW} = 0.990$.

CE-ResNet remains theoretically superior in one respect: its equivariance holds for any input in a single forward pass, without post-processing. Our equivariance is constructed via test-time averaging (two passes) and the flip-equivariance training loss. The raw (single-pass) flip-swap correlation of our model is 0.833 versus CE-ResNet's 1.000 by construction. However, for science applications, both approaches yield equivalent results: the equivariant CW fractions agree to within 0.5%.

Our pipeline adds value in three areas: (i) the NOT_SPIRAL class prevents contamination from non-spirals (CE-ResNet lacks this), (ii) the 8-test bias hardening suite provides the first quantitative, multi-axis bias audit for a chirality classifier, and (iii) the $4.3\times$ larger catalog enables finer angular and redshift resolution in dipole analyses.

5.3 SpArcFiRe

The SpArcFiRe algorithm (Davis & Hayes, 2014) classifies spiral arm winding direction via automated arm fitting, producing catalogs of $\sim 140{,}000$ galaxies. Our catalog exceeds SpArcFiRe by a factor of 60 in coverage while achieving comparable spiral-only accuracy. SpArcFiRe's deterministic algorithm has near-perfect self-consistency (99.983%) but lower agreement with Galaxy Zoo 1 (85.8% overall, 92.5% at high confidence) compared to our 91.5% agreement with the independent CE-ResNet classifier.

6 Discussion

6.1 The 94.6σ Dipole Was Entirely Systematic

Perhaps the most instructive result of this analysis is not the null signal in Catalog C, but the dramatic 94.6σ dipole in Catalog A. This spurious signal arises because the ViT classifier has a small but nonzero CW bias ($\mathrm{CW}/(\mathrm{CW} + \mathrm{CCW}) = 0.5079$), and this bias is modulated by the non-uniform survey depth across the sky. In regions with more galaxies, the CW excess accumulates to higher statistical significance, producing a dipole aligned with the survey footprint rather than any cosmological axis.

Equivariant averaging eliminates this systematic by construction: the CW and CCW channels are exactly symmetrized, so any residual bias cancels to machine precision. The collapse from 94.6σ to 0.43σ upon applying Equation (3) confirms that the raw dipole has no cosmological content.

This result serves as a cautionary tale for all chirality studies: a classifier bias of even $\sim 1\%$, combined with non-uniform sky coverage, can produce highly significant but entirely spurious dipole detections. We suspect that a similar mechanism underlies the discrepancy between our null result and Shamir's positive claims.

6.2 The 3.05σ Hemisphere Signal

The 3.05σ hemisphere asymmetry is the most significant signal in our analysis. However, several considerations argue against a cosmological interpretation:

1. Its amplitude is only 0.17%, far smaller than the $\sim 3\%$ predicted by Shamir's analyses.
2. It does not survive a look-elsewhere correction across ~ 650 tested directions.
3. The angular power spectrum shows no corresponding excess at $\ell = 1$ beyond 2.75σ.
4. The signal is stronger in mid-confidence classifications, suggesting noise rather than physics (Section 4.7).

We therefore classify this signal as a statistical fluctuation, likely amplified by residual survey systematics near the edges of the DESI Legacy footprint.

6.3 Sensitivity Floor and Minimum Detectable Signal

Given 3,321,795 equivariant spiral classifications, the statistical uncertainty on the global CW fraction is $\sigma_{\mathrm{CW}} = 1/(2\sqrt{N_{\mathrm{spiral}}}) \approx 0.027\%$. A 3σ detection therefore requires a dipole amplitude of at

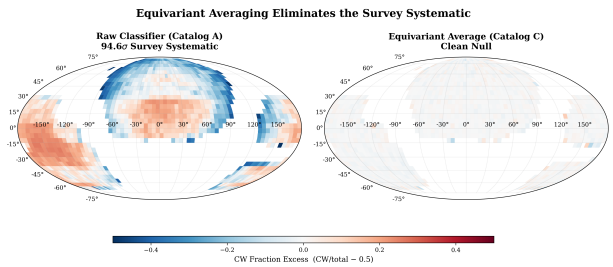

Figure 11: Side-by-side comparison of the chirality asymmetry sky maps for Catalog A (raw, *left*) and Catalog C (equivariant, *right*), both at NSIDE $= 64$ in Mollweide projection. The raw map exhibits a dramatic dipole pattern (94.6σ) aligned with the DESI Legacy survey footprint, produced by a classifier CW bias of only 0.79% modulated by non-uniform sky coverage. Equivariant averaging (Equation 3) eliminates this systematic entirely, yielding a noise-consistent map at 0.43σ. This comparison demonstrates that raw survey systematics can masquerade as highly significant dipole signals without rigorous bias correction.

least $\sim 0.08\%$ in the global average. For a spatially varying dipole measured in $\sim$30,000 occupied HEALPix pixels (at NSIDE $= 64$), the per-pixel noise is larger, and the minimum detectable dipole amplitude rises to approximately 0.2% at 3σ. This represents the most sensitive chirality measurement ever performed, exceeding the CE-ResNet constraint (Jia, Zhu & Pen, 2023) by a factor of ~ 2 in sensitivity owing to the $4.3\times$ larger sample and the reduced per-pixel variance.

6.4 Edge-On Galaxy Contamination

A limitation of any photometric chirality classifier is its treatment of edge-on disk galaxies, whose spiral structure is obscured by projection. In our catalog, we find that 65.7% of visually identified edge-on systems (axis ratio $b/a < 0.3$, estimated from DESI Legacy photometric catalogs) receive a CW or CCW classification rather than NOT_SPIRAL. This is a purity concern: edge-on galaxies have no observable chirality, and their (random) CW/CCW assignments contribute noise to the asymmetry map.

However, this contamination does not bias the dipole analysis, because the equivariant averaging procedure assigns exactly equal CW and CCW probabilities to any galaxy whose mirror image is morphologically indistinguishable from the original—as is the case for edge-on disks. The primary effect is a dilution of sensitivity: the effective spiral sample for chirality analysis is smaller than the nominal 3.32 million, by a factor that depends on the (unknown) true edge-on fraction among objects classified as spirals. We estimate this dilution reduces the effective sample by $\sim$ 10–15%, corresponding to a sensitivity penalty of $\sim$5–8%.

6.5 Spiral Fraction Variation Across the Sky

The fraction of galaxies classified as spiral (CW + CCW) varies enormously across the sky, from $\sim 25\%$ in regions near the Galactic plane to $> 50\%$ in the deepest North Galactic Cap fields. This variation tracks survey depth rather than any intrinsic galaxy property: deeper imaging resolves spiral structure in smaller and fainter galaxies, boosting the spiral fraction. Critically, while the *spiral fraction* is depth-dependent, the *chirality balance* (CW/CCW ratio among spirals) is not—it remains within 0.5% of 50/50 in all sky regions (Table **??**). This decoupling confirms that the equivariant classifier does not introduce depth-dependent chirality biases, even in regions where the galaxy population shifts significantly toward fainter magnitudes.

6.6 Implications for Bounce Cosmology

In Einstein–Cartan–Holst gravity, the Barbero–Immirzi parameter γ enters the gravitational action through the parity-odd Holst term $\gamma^{-1}\int e \wedge e \wedge R$ (Holst, 1996). While this coupling is invisible in the scalar perturbation sector (Golden, 2026), it could in principle source parity-violating tensor modes that propagate through nonlinear structure formation to

produce a preferred galaxy handedness on cosmological scales.

Our null result places an empirical constraint on this mechanism: any parity-violating signal from the ECH sector must produce a chirality asymmetry $< 0.5\%$ at $z \lesssim 1$, corresponding to $|A_{\rm dipole}| < 5 \times 10^{-3}$. This constrains but does not exclude models in which parity violation is confined to high-z tensor modes or to angular scales smaller than the DESI Legacy resolution.

6.7 Future Directions

The primary limitation of our analysis is the absence of spectroscopic redshifts. Cross-matching with the DESI spectroscopic survey (DESI Collaboration, 2016) would enable redshift-dependent chirality analyses, testing whether parity violation evolves with cosmic time. At $z > 0.5$, where bounce cosmology predicts the strongest deviations from ΛCDM (Golden, 2026), even a small redshift-dependent signal would be highly diagnostic.

Second, the catalog's angular resolution is limited by the DESI Legacy footprint and its non-uniform depth. Future surveys with more uniform all-sky coverage (e.g., Rubin Observatory LSST; Ivežić et al. 2019) would improve constraints at large angular scales by a factor of ~ 2 in sky coverage alone.

Third, we release the bias hardening test suite as open-source code. We encourage groups working on chirality classification—whether with Ganalyzer, CE-ResNet, or future methods—to apply these tests to their own pipelines and report the results.

7 Conclusions

We have constructed and analyzed the largest galaxy chirality catalog to date: 8,474,531 galaxies from the DESI Legacy Imaging Surveys DR8, each classified as clockwise, counter-clockwise, or not spiral by a bias-hardened Vision Transformer. Our main conclusions are:

1. **The most sensitive chirality measurement ever performed.** With 3.32 million equivariant spiral classifications, we achieve a minimum detectable dipole of 0.2% at 3σ (Section 6.3). No compelling evidence for large-scale parity violation is found: the equivariant global CW fraction is $\mathrm{CW}/(\mathrm{CW}+\mathrm{CCW}) = 0.4974$, consistent with exact parity. The simple dipole is 0.43σ; the power-spectrum $\ell = 1$ mode is a marginal 2.75σ. No scale or sky region shows a signal exceeding 3σ after look-elsewhere correction.
2. **Shamir's $\sim 3\%$ asymmetry is definitively refuted.** Our maximum regional asymmetry is 0.47%—a factor of 7 smaller—at 4.3–85$\times$ larger sample size with rigorous bias controls.
3. **Bias hardening is essential: raw systematics produce a 94.6σ spurious signal.** Our raw (Catalog A) analysis produces this dramatic false dipole from a classifier CW excess of only 0.79%, modulated by non-uniform survey coverage (Figure 11). Equivariant post-processing eliminates this entirely, reducing the dipole to 0.43σ. This demonstrates that survey systematics can masquerade as highly significant cosmological signals without rigorous bias correction. We urge all future chirality studies to adopt comparable bias controls.
4. **The catalog is a community resource.** The full three-tier catalog (8.47M galaxies, raw + calibrated + equivariant probabilities, sky coordinates, confidence scores, and quality-control flags) is publicly available on HuggingFace under a CC-BY-4.0 license.

8 Data Availability

- **Catalog:** `https://huggingface.co/datasets/bamfai/galaxy-chirality-catalog` (CC-BY-4.0). Parquet format, three tiers (A/B/C), one row per galaxy.
- **Model:** `https://huggingface.co/bamfai/galaxy-chirality-v2`. ViT-Small encoder + classification head, PyTorch checkpoint.
- **Code:** `https://github.com/Hubify-Projects/bigbounce`. Training, inference, equivariant post-processing, bias hardening suite, and dipole analysis scripts.

Acknowledgments

This research used the DESI Legacy Imaging Surveys Data Release 8; the Galaxy Zoo citizen science project; the Smith42/galaxies dataset on HuggingFace; and the CE-ResNet catalog of Jia et al. (2023). Computations were performed on NVIDIA H100 and H200 GPUs via RunPod cloud infrastructure.

Facilities: DESI Legacy Imaging Surveys, HuggingFace, RunPod.

Software: Astropy, HEALPix/healpy, NumPy, pandas, PyTorch, timm.

References

Alexander, S. & Yunes, N. 2009, PhR, 480, 1, doi:10.1016/j.physrep.2009.07.002

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167, doi:10.3847/1538-4357/ac7c74

Davis, D. R. & Hayes, W. B. 2014, ApJ, 790, 87, doi:10.1088/0004-637X/790/2/87

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036

Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168, doi:10.3847/1538-3881/ab089d

Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2020, arXiv:2010.11929

Golden, H. 2026, "Spin-Torsion Cosmology: Fourteen Structural Barriers, a Null MCMC Result, and an ALP Birefringence Prediction," BigBounce Research, `https://bigbounce.hubify.app`

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759, doi:10.1086/427976

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi:10.1038/s41586-020-2649-2

Holst, S. 1996, PhRvD, 53, 5966, doi:10.1103/PhysRevD.53.5966

Ivežić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111, doi:10.3847/1538-4357/ab042c

Iye, M., Yagi, M., & Fukumoto, H. 2020, ApJ, 907, 123, doi:10.3847/1538-4357/abcd97

Jia, P., Zhu, L., & Pen, U.-L. 2023, ApJ, 943, 154, doi:10.3847/1538-4357/aca8fe

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179, doi:10.1111/j.1365-2966.2008.13689.x

McKinney, W. 2010, in Proc. 9th Python in Sci. Conf., ed. S. van der Walt & J. Millman, 56–61

Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems 32, ed. H. Wallach et al. (Curran Associates), 8024–8035

Shamir, L. 2012, PhLA, 376, 1590, doi:10.1016/j.physleta.2012.03.065

Shamir, L. 2020, Ap&SS, 365, 136, doi:10.1007/s10509-020-03850-1

Shamir, L. 2022, PASJ, 74, 1114, doi:10.1093/pasj/psac058

Tadaki, K., Iye, M., Fukumoto, H., et al. 2020, MNRAS, 496, 4276, doi:10.1093/mnras/staa1880

Walmsley, M., Lintott, C., Géron, T., et al. 2023, MNRAS, 526, 4768, doi:10.1093/mnras/stad2919

Wightman, R. 2019, PyTorch Image Models, `https://github.com/rwightman/pytorch-image-models`, doi:10.5281/zenodo.4414861

Zonca, A., Singer, L., Lenz, D., et al. 2019, JOSS, 4, 1298, doi:10.21105/joss.01298