

Article

Not peer-reviewed version

---

# Conformal Prediction for Quantum Entanglement: Robustness and Detection

---

[Davut Emre Tasar](#)\*

Posted Date: 24 November 2025

doi: 10.20944/preprints202511.1754.v1

Keywords: conformal prediction; quantum entanglement; anomaly detection; contextuality; Bell inequalities; device realism



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

# Conformal Prediction for Quantum Entanglement: Robustness and Detection

Davut Emre Taşar

Independent Researcher, Spain; detasar@gmail.com

## Abstract

Certification of quantum entanglement under realistic device imperfections is central to quantum communication, cryptography, and quantum networks. Despite the nonclassical nature of entanglement and contextuality, we empirically verify that standard Conformal Prediction (CP) retains its coverage guarantees in the CHSH scenario, as theoretically expected under (context-conditional) exchangeability. We introduce an LHV-Calibrated Anomaly Detection framework achieving an ROC AUC  $\approx 0.96$  (Scenario A) and AUC  $\approx 0.83$  using only CP-derived features (Scenario B). Validation on IBM quantum hardware ( $S_{\text{hw}} \approx 2.53$ ) confirms these results under Noisy Intermediate-Scale Quantum (NISQ) conditions, relative to an explicit LHV null manifold. Finally, we prove a no-go theorem for product-threshold Conformal Entanglement Witnesses (CEW), explaining their limited power. Our results suggest that, at least in the CHSH setting considered here, CP is best viewed as a robust statistical wrapper around quantum devices, rather than a fragile entanglement witness. To facilitate reproducibility, all code, data, and experiment configurations are available at <https://github.com/detasar/QCE>.

**Keywords:** conformal prediction; quantum entanglement; anomaly detection; contextuality; Bell inequalities; device realism

## 1. Introduction

Certification of quantum entanglement under realistic device imperfections is central to quantum communication, cryptography, and quantum networks [33,49,51,55]. Loophole-free Bell tests [35] have confirmed the existence of nonlocal correlations, but certifying them in practical protocols often requires rigorous statistical guarantees against model misspecification and drift.

Conformal Prediction (CP) [1,2,6] offers a powerful framework for distribution-free uncertainty quantification. Unlike traditional Bayesian or frequentist methods that rely on strict distributional assumptions, CP guarantees valid coverage for prediction sets under the minimal assumption of exchangeability. Recent advances have extended CP to handle covariate shift [8] and non-exchangeable data [11], making it a robust candidate for wrapping quantum technologies.

Machine-learning-based entanglement witnesses have recently emerged as a flexible alternative to standard Bell inequalities. Gao et al. [61] use neural networks for quantum state reconstruction, while Greenwood et al. [62] and Ureña et al. [64] develop SVM and deep learning witnesses tailored to specific measurement settings. More recently, Huang et al. [63] demonstrated direct entanglement detection using machine learning. However, these methods typically lack finite-sample statistical guarantees. Our approach differs by wrapping ML classifiers in a distribution-free conformal layer: instead of learning a direct witness, we train on the null manifold of Local Hidden Variable (LHV) models and use conformal scores as calibrated anomaly features.

Concretely, we work in the standard Clauser–Horne–Shimony–Holt (CHSH) scenario [27]: two parties, binary inputs  $x, z \in \{0, 1\}$  and binary outcomes  $a, b \in \{\pm 1\}$ , with the Bell parameter  $S$ . In the usual notation,

$$S = E_{00} + E_{01} + E_{10} - E_{11}, \quad E_{xz} = \mathbb{E}[ab \mid x, z],$$

and classical LHV models satisfy  $|S| \leq 2$ , while quantum mechanics allows up to Tsirelson's bound  $|S| \leq 2\sqrt{2}$  [28]. Our main question is not whether quantum mechanics violates CHSH—that is textbook knowledge—but whether such strongly contextual statistics can break the seemingly classical guarantees of conformal prediction, or conversely, if CP can be harnessed to detect entanglement.

Contributions.

- **Robustness of CP (Program F):** We empirically verify via sequential martingale tests that CP validity holds for quantum data, confirming that contextuality does not induce miscalibration in standard settings.
- **LHV-Calibrated Anomaly Detection (Program A/D):** We design a detection framework trained solely on classical null models. Using features like the Bell parameter  $|S|$  and specific click patterns, we achieve ROC AUC  $\approx 0.96$  against a mixture of quantum and adversarial LHV data under device realism.
- **No-Go Theorem for CEW (Program B/C):** We prove a no-go theorem for product-threshold Conformal Entanglement Witnesses (CEW), explaining why they are structurally conservative and lack power compared to our anomaly detector.
- **Validation on IBM Quantum Platform:** We validate our results on the `ibm_fez` quantum processor, observing  $S_{hw} \approx 2.53$  and confirming CP calibration under real NISQ conditions.
- **Open-Source Framework:** We provide a comprehensive, modular Python library for quantum conformal prediction, enabling the reproduction of all experiments and the development of new witnesses.

Roadmap.

In Section 2, we review conformal prediction and define our LHV-calibrated anomaly detection pipeline. Section 3 collects our formal results, including the no-go theorem. Section 4 outlines a sheaf-theoretic perspective. Section 5 reports our simulation and IBM hardware experiments. We discuss implications in Section 6 and close with limitations and future directions in Section 7.

## 2. Methods

### 2.1. Conformal Prediction and Mondrian CP

We employ split-conformal prediction (CP) [1,2] to construct rigorous prediction sets. The procedure is as follows:

1. **Data Splitting:** Divide the available data into a training set  $\mathcal{D}_{\text{train}}$  and a calibration set  $\mathcal{D}_{\text{cal}}$ .
2. **Score Function:** Train a base predictor  $\hat{\pi}$  on  $\mathcal{D}_{\text{train}}$  to output probability estimates. Define a nonconformity score  $s(x, y) = 1 - \hat{\pi}(y|x)$ , which measures how "strange" a label  $y$  is given feature  $x$ .
3. **Calibration:** Compute scores  $s_i = s(x_i, y_i)$  for all  $(x_i, y_i) \in \mathcal{D}_{\text{cal}}$ .
4. **Prediction:** For a new test point  $x_{n+1}$  and candidate label  $y$ , compute  $s_{n+1} = s(x_{n+1}, y)$ . The conformal  $p$ -value is:

$$p(x_{n+1}, y) = \frac{1 + \sum_{i \in \mathcal{D}_{\text{cal}}} \mathbf{1}\{s_i \geq s_{n+1}\}}{1 + |\mathcal{D}_{\text{cal}}|}.$$

The prediction set is  $C_\alpha(x_{n+1}) = \{y : p(x_{n+1}, y) > \alpha\}$ .

Under the assumption of exchangeability, this procedure guarantees  $\mathbb{P}\{Y_{n+1} \in C_\alpha(X_{n+1})\} \geq 1 - \alpha$  for any finite sample size and any underlying distribution [6]. While we focus on standard split-CP here, our framework is compatible with selective or risk-controlling CP variants [7]. To handle the heterogeneous nature of CHSH measurements (where inputs  $x, z$  define distinct contexts), we use **Mondrian Conformal Prediction** [1], which stratifies the calibration by context. This ensures that

validity holds conditionally within each measurement setting, preventing coverage errors in one context from masking failures in another.

Wrapper vs witness usage modes.

In this work we use conformal prediction in two distinct regimes. In what we call the *wrapper mode*, the base predictor and the calibration set are fitted on data drawn from the same distribution as the test data (either LHV or quantum), and split-CP is used purely as a validity wrapper. In this mode, Lemma 1 applies: under (context-conditional) exchangeability the conformal  $p$ -values are (super-)uniform and the CMI vanishes for both LHV and quantum data. In contrast, in the *witness mode* the base predictor and calibration set are fitted *only* on the LHV null manifold and then reused for both LHV and quantum test datasets. In this regime the uniformity guarantee applies only to classical LHV data, and deviations of the  $p$ -value distribution (or increases in CMI) for quantum data become informative features for entanglement detection. Our anomaly detection experiments (Programs A/D and Scenario B) operate in this LHV-calibrated witness mode, while the martingale robustness tests (Program F) operate in wrapper mode.

Statistical Tools.

We quantify miscalibration using the expected calibration error (ECE) and total variation (TV) distance between empirical and uniform  $p$ -value distributions, and rely on Kolmogorov–Smirnov (KS) and Anderson–Darling (AD) tests. Multiple testing is controlled via the Benjamini–Hochberg false discovery rate (BH–FDR) procedure [17].

## 2.2. LHV-Calibrated Anomaly Detection

Instead of testing for validity violations, we treat quantum detection as a One-Class Classification problem. We define a "Null Manifold"  $\mathcal{M}_{LHV}$  consisting of data generated by Local Hidden Variable models, including those exploiting loopholes (Detection, Memory, Communication). We train a One-Class Elliptic Envelope (used throughout all experiments; One-Class SVM gave similar results) on samples from  $\mathcal{M}_{LHV}$ , using a feature vector

$$\mathbf{f} = [|S|, \text{click}_A, \text{click}_B, \text{click}_{AB}, \text{click}_\emptyset, \text{CMI}, \mathbb{E}|C_\alpha|, \mathbb{E}(|C_\alpha^{(X)}||C_\alpha^{(Z)}|)],$$

where  $\text{click}_A, \text{click}_B, \text{click}_{AB}, \text{click}_\emptyset$  denote the empirical rates of single-click events at A, B, both-click, and no-click events, respectively. The last three components are CP-derived diagnostics and encode how "nonconforming" each dataset looks from the conformal point of view.

We consider two operating regimes:

- **Scenario A (full features):** Anomaly detection uses the complete vector

$$\mathbf{f}_{\text{full}} = [|S|, \text{click}_A, \text{click}_B, \text{click}_{AB}, \text{click}_\emptyset, \text{CMI}, \mathbb{E}|C_\alpha|, \mathbb{E}(|C_\alpha^{(X)}||C_\alpha^{(Z)}|)].$$

This yields our primary result (ROC AUC  $\approx 0.96$ ).

- **Scenario B (CP-only witness):** We deliberately exclude the Bell score  $|S|$  and restrict to conformal features and click rates:  $\mathbf{f}_{\text{CP}} = [\text{click}_A, \text{click}_B, \text{click}_{AB}, \text{click}_\emptyset, \text{CMI}, \mathbb{E}|C_\alpha|, \mathbb{E}(|C_\alpha^{(X)}||C_\alpha^{(Z)}|)]$ . This tests whether CP-derived features alone suffice for entanglement detection (Section 5.6).

Detection Protocol.

We formalize the detection procedure in Algorithm 1.

We construct conformal anomaly scores via  $s(x, y)$  and use an LHV-calibrated threshold. A high anomaly score indicates deviation from the classical polytope, effectively flagging quantum behavior without requiring a violation of CP validity.

**Algorithm 1** LHV-Calibrated Anomaly Detection Protocol

- 1: **Input:** Training set  $\mathcal{D}_{LHV}$  from classical models, Test dataset  $D_{test}$
- 2: **Step 1: Feature Extraction**
- 3: Compute  $\mathbf{f}(D) = [|S|, \text{click rates}, \text{CP diagnostics}]$  for all  $D \in \mathcal{D}_{LHV}$
- 4: **Step 2: Calibration**
- 5: Fit One-Class Classifier (Elliptic Envelope) on  $\{\mathbf{f}(D)\}_{D \in \mathcal{D}_{LHV}}$
- 6: Set anomaly threshold  $\tau$  for desired False Positive Rate (e.g., 0.01)
- 7: **Step 3: Detection**
- 8: Compute  $\mathbf{f}_{test} = \mathbf{f}(D_{test})$
- 9: **Return** Quantum if  $\text{AnomalyScore}(\mathbf{f}_{test}) > \tau$ , else Classical

## 2.3. Data Generation: Quantum and LHV Models

**Quantum CHSH Protocol:** We simulate the CHSH scenario with a maximally entangled Bell state  $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$  and measurement settings optimized for maximal violation:  $\theta_A \in \{0, -\pi/2\}$  and  $\theta_B \in \{-\pi/4, \pi/4\}$ . Noise is modeled via depolarization probability  $p_{\text{depol}} \in [0, 0.2]$ , detection efficiency  $\eta \in [0.8, 1.0]$ , and dark count rates.

**LHV Model Families:**

- **Detection Loophole:** Models exploit imperfect detection ( $\eta < 1$ ) via the Garg-Mermin construction, allowing classical correlations to mimic quantum statistics within the fair-sampling loophole.
- **Memory Loophole:** Sequential measurement outcomes are correlated through a persistent hidden variable that adapts based on previous outcomes.
- **Communication Loophole:** Minimal one-bit signaling between measurement stations, constrained to respect locality bounds in practice.

All datasets are generated over a grid of  $(p_{\text{depol}}, \eta)$  combinations to form a comprehensive LHV atlas for calibration. All classification models are implemented using scikit-learn [60]. Table 1 summarizes the datasets used.

**Table 1.** Summary of generated datasets for calibration and testing. Each dataset consists of  $N = 1000$  measurement rounds (shots).

Regime	Source	# Datasets	Parameters	Purpose
Null Manifold	LHV (Detection)	2000	$\eta \in [0.8, 1.0]$	Training
Null Manifold	LHV (Memory)	1000	$\mu \in [0, 0.5]$	Training
Null Manifold	LHV (Comm.)	1000	$\epsilon \in [0, 0.1]$	Training
Test Set	Quantum (CHSH)	500	$p_{\text{depol}} \in [0, 0.2]$	Evaluation
Test Set	Adversarial LHV	500	Mixed	Evaluation

## 2.4. Sheaf-Index and Martingale Tests

**Contextual Miscalibration Index (CMI):** We define CMI using the Kolmogorov-Smirnov (KS) statistic, which measures the maximum deviation between the empirical p-value distribution and the uniform distribution. For a given context  $c$ , the context-specific CMI is:

$$\text{CMI}_c = \sup_{t \in [0,1]} |\mathbb{P}(p(X, Y) \leq t \mid C = c) - t|,$$

and the global CMI is the expectation over contexts:

$$\text{CMI}(f) = \mathbb{E}_C[\text{CMI}_c].$$

*Remark:* Alternative formulations using L1 distance exist; we use the KS-based definition for its sensitivity to local deviations.

**Sheaf-Index ( $W_{sheaf}$ ):** To probe topological obstructions, we compare the CMI computed with Mondrian (context-aware) conditioning versus Global (context-ignorant) conditioning.

$$W_{sheaf} = \text{CMI}_{\text{Global}} - \text{CMI}_{\text{Mondrian}}$$

A positive  $W_{sheaf}$  would indicate that pooling contexts introduces miscalibration not present locally. Despite the name,  $W_{sheaf}$  is a simple miscalibration gap between pooled and context-wise conformal fits, rather than a fully fledged sheaf-theoretic invariant.

**Betting Martingale:** To rigorously test the validity of CP on quantum data, we employ a betting martingale  $M_t$ .

$$M_t = \prod_{i=1}^t (\alpha \cdot p_i^{\alpha-1})$$

Under the null hypothesis of valid CP ( $p_i \sim \text{Unif}(0, 1)$ ), the expected wealth growth is non-positive. Growth of  $M_t$  would witness a violation of validity (robustness failure).

### 3. Formal Results

**Lemma 1** (CMI null: conditional exchangeability implies uniformity). *Let  $(X, Y, C)$  denote features, labels, and a discrete context (e.g., CHSH setting). Assume split-conformal prediction with label-conditional calibration within each context (under the usual randomized tie-breaking; without randomization the  $p$ -values are super-uniform but still yield vanishing CMI in the population limit). If  $(X_i, Y_i)$  are exchangeable within each context level  $C = c$ , then for any test point  $(x, y)$  with context  $c$  the conformal  $p$ -values are (randomized) uniformly distributed, and in particular super-uniform with respect to  $\text{Unif}(0, 1)$ . Consequently, the population CMI—defined as a weighted total-variation (or KS/AD) distance of the context-wise  $p$ -value distributions to  $\text{Unif}(0, 1)$ —equals zero under the null.*

**Sketch of Lemma 1.** Split conformal predictors with label-conditional (Mondrian) calibration guarantee conditionally uniform  $p$ -values within each stratum [1,2]. Mondrian conditioning refines the exchangeability argument to each context  $C = c$ . Specifically, since the calibration data is exchangeable with the test data *conditional on the context*, the rank of the test score is uniformly distributed among the calibration scores for that context. Thus, under the classical null, the population CMI vanishes.  $\square$

**Remark 1** (Scope). *This lemma assumes an ideal split-CP setup with correctly specified non conformity scores and exchangeable data within each context. In later sections we deliberately violate these assumptions (e.g., via restricted predictive models or adversarial LHV strategies that break exchangeability), where CMI becomes large. Thus, CMI serves as an operational diagnostic of deviations from the ideal CP null rather than an intrinsic property of data distributions.*

#### 3.1. Product-threshold CEWs and their limitations

Contextual Miscalibration Index (CMI).

For each context  $c$ , let  $\{p_i^{(c)}\}$  denote the conformal  $p$ -values within that context and  $\hat{F}_c$  their empirical CDF. We define the context-specific miscalibration index as

$$\text{CMI}_c = \int_0^1 |\hat{F}_c(u) - u| du \approx \frac{1}{B} \sum_{j=1}^B |\hat{F}_c(u_j) - u_j|, \quad (1)$$

where  $B$  is the number of bins and  $u_j = j/B$ . The global CMI is the context-weighted average:

$$\text{CMI}(f) = \mathbb{E}_C \left[ \sup_{t \in [0,1]} |\mathbb{P}(p(X, Y) \leq t | C) - t| \right]. \quad (2)$$

Despite the name, CMI is a measure of conformal miscalibration across contexts, not a direct measure of contextuality in the sheaf-theoretic sense; contextuality only enters implicitly through the underlying data distribution.

**Proposition 1** (Baseline for CEW under binary split-CP). *Let  $C_\alpha^{(X)}(x)$  and  $C_\alpha^{(Z)}(x)$  be binary split-CP prediction sets in complementary contexts with label-conditional calibration. Assume the standard non-empty-set convention (sets are never empty). Then for each context  $c \in \{X, Z\}$ , the expected set size is bounded by  $1 \leq \mathbb{E}[|C_\alpha^{(c)}|] \leq 1 + \alpha + \varepsilon_n$ , where  $\varepsilon_n \rightarrow 0$  as the calibration size  $n \rightarrow \infty$  [59]. Intuitively, efficient CP sets cover the true label with probability  $1 - \alpha$ , often resulting in size 1 (certainty) or 2 (uncertainty), averaging to  $\approx 1 + \alpha$ . If, under the classical null,  $|C_\alpha^{(X)}|$  and  $|C_\alpha^{(Z)}|$  are independent (or weakly positively associated), then  $\mathbb{E}[|C_\alpha^{(X)}||C_\alpha^{(Z)}|] \gtrsim \mathbb{E}[|C_\alpha^{(X)}|] \cdot \mathbb{E}[|C_\alpha^{(Z)}|] \approx (1 + \alpha)^2$ .*

**Remark 2.** *Proposition 1 explains the empirical negative atlas for CEW: with realistic noise and small  $\alpha$ , fair thresholds  $L$  near  $(1 + \alpha)^2$  will rarely be violated. This theorem combines the lower bound of Dhillon et al. [59] with the observation that efficient nonconformity scores (e.g., quantile-based, distance-based) used in practice yield prediction sets that approach this bound under high-noise regimes.*

Conformal Entanglement Witness (CEW).

We consider product-threshold CEWs of the form  $\mathbf{1}\{\mathbb{E}[|C_\alpha^{(X)}||C_\alpha^{(Z)}|] < L\}$  at a fixed miscoverage level  $\alpha$ . Proposition 1 and Theorem 1 show that such CEWs are structurally limited, with the classical baseline concentrating around  $(1 + \alpha)^2$ .

**Sketch of Proposition 1.** Label-conditional split CP with the non-empty-set convention yields  $\mathbb{E}|C_\alpha^{(c)}| \in [1, 1 + \alpha] + \varepsilon_n$  for each basis [59]. Under the classical null the two contexts are (approximately) independent or positively associated, so the expected product is at least the product of expectations, giving  $(1 + \alpha)^2$  up to finite-sample slack absorbed by  $\varepsilon_n$ .  $\square$

$$\mathbb{E}[|C_\alpha^{(X)}||C_\alpha^{(Z)}|] \gtrsim (1 + \alpha)^2. \quad (3)$$

**Theorem 1** (No-go for product-threshold CEWs). *Assume binary split-CP with label-conditional calibration and the non-empty-set convention. Let  $|C_\alpha^{(X)}|, |C_\alpha^{(Z)}|$  denote CP set sizes for complementary contexts  $X, Z$ . Under a classical noncontextual null, assume:*

1. **Independence/Correlation:** *Set sizes are independent or non-negatively correlated across contexts.*
2. **Efficiency:**  $\mathbb{E}|C_\alpha^{(c)}| \rightarrow 1 + \alpha$  (efficient sets).

Then:

$$\lim_{n \rightarrow \infty} \mathbb{E}[|C_\alpha^{(X)}||C_\alpha^{(Z)}|] \geq (1 + \alpha)^2 - o(1).$$

Consequently, any CEW of the form  $\mathbf{1}\{\mathbb{E}[|C_\alpha^{(X)}||C_\alpha^{(Z)}|] < L\}$  with fixed threshold  $L < (1 + \alpha)^2$  has asymptotically vanishing rejection probability under the classical noncontextual null (that is, its type-I error tends to zero as the calibration size grows). This theorem by itself does not make any claim about the power under quantum alternatives. Our simulations confirm that realistic quantum CHSH data also concentrate around this baseline (see Appendix B, Figure A4), so in our setting such product-threshold CEWs exhibit empirically low power in practice.

Intuition.

The factor  $(1 + \alpha)^2$  arises because efficient conformal predictors tend to produce sets of size 1 (uninformative) or 2 (uncertain) to satisfy the  $1 - \alpha$  coverage requirement. In the binary case, the expected size is tightly lower-bounded by  $1 + \alpha$  (see [59]). Since classical LHV models factorize across contexts, the expected product of set sizes is simply the product of their expectations,  $(1 + \alpha) \times (1 + \alpha)$ .

Quantum correlations, despite being nonlocal, do not sufficiently compress the prediction sets below this baseline in the product-threshold regime, leading to the no-go result.

**Remark 3.** *This no-go applies only to CEWs depending solely on the expected product at fixed  $\alpha$ . Sophisticated witnesses exploiting the full  $p$ -value structure or  $\alpha$ -path fall outside this scope.*

## 4. Category-Theoretic Foundations and Outlook

While our main results are operational and numerical, they admit a natural categorical reformulation which clarifies the structural nature of our no-go theorem. *This section provides an optional categorical reformulation; it can be skipped by readers primarily interested in operational results.* We do not use these structures anywhere in the proofs or experiments; they only serve as a conceptual lens. In particular, none of our main validity or no-go statements relies on these categorical constructions; they only suggest directions for designing more structural nonconformity scores in future work. Its main purpose is to argue that our no-go theorem is not a simulation artifact, but reflects a structural limitation: conformal predictors as Markov category morphisms (e.g., viewing the predictor as a stochastic channel  $K : X \rightarrow \mathcal{P}(Y)$ ) imply that the  $(1 + \alpha)^2$  baseline emerges as a coverage-preserving factorization invariant, loosely analogous to the no-cloning theorem.

To rigorously ground our conformal witness framework, we propose a formalization based on **Enriched Markov Categories and Sheaf Theory**.

### 4.1. Conformal Prediction as a Sheaf

Let **Meas** be the category of measurement contexts. We define a presheaf  $\mathcal{C} : \mathbf{Meas}^{op} \rightarrow \mathbf{Credal}$ , where  $\mathcal{C}(U)$  represents the set of valid conformal prediction regions calibrated on context  $U$ . The *Contextual Fraction* (CF) of Abramsky and Brandenburger can be interpreted as a measure of the obstruction to finding a global section for this sheaf.

$$CF(P) = 1 - \max\{\lambda \mid P = \lambda P_{NC} + (1 - \lambda)P_C\}$$

As a **conjecture** (heuristic ansatz), we propose that a CF-dependent threshold of the form  $L_{\text{sheaf}} = L_0(1 - \gamma \cdot CF)$  could interpolate between noncontextual and maximal-contextual regimes; a rigorous derivation is left for future work.

**Note:** Our CF implementation currently returns  $CF \approx 1$  across the CHSH grid, so we treat CF here mainly as a qualitative indicator; strengthening the CMI-CF link is an open problem.

### 4.2. Enriched Markov Categories

Following Fritz et al., we view the conformal predictor as a morphism in an Enriched Markov Category. The non-conformity score  $\psi$  is a morphism  $\psi : (X \times Y)^n \times (X \times Y) \rightarrow \mathbb{R}$ . The validity of the conformal witness corresponds to the existence of a factorization of the prediction channel through a "Credal Object" that preserves the coverage invariant under the action of the permutation group (exchangeability). This high-level perspective unifies our negative results (Theorem 1) with the structural impossibility of cloning quantum information, suggesting that "Conformal Cloning" of entangled statistics is strictly bounded by the  $(1 + \alpha)^2$  limit.

---

#### Algorithm 2 CMI pipeline (per grid point, with BH-FDR across grid)

---

- 1: **Input:** dataset  $\{(x_i, y_i, c_i)\}_{i=1}^n$ , significance  $\alpha$ , bins  $B$ , calibration option  $\in \{\text{none, platt, isotonic}\}$
  - 2: **for** each context  $c$  **do** ▷ Mondrian conditioning
  - 3:     Fit base score model on train split within  $c$ ; optionally fit calibrator on inner validation
  - 4:     Compute split-CP  $p$ -values  $\{p_i\}_{i:c_i=c}$  for both labels and record  $p(y_i)$
  - 5:     Compute TV/ECE and KS/AD  $p$ -values vs Unif(0, 1) using  $B$  bins
  - 6: Apply BH-FDR to the list of per-context KS  $p$ -values at level  $q$ ; mark discoveries
  - 7: Aggregate a weighted TV/ECE across contexts for reporting
-

**Algorithm 3** CEW decision (per replicate, then aggregate)

- 
- 1: **Input:** split-CP outputs per context: average set sizes  $\tau_X, \tau_Z$ , replicate count  $R$ , threshold family  $L(\cdot)$
  - 2: **for**  $r = 1$  to  $R$  **do**
  - 3:     compute  $|C_\alpha^{(X)}|_r, |C_\alpha^{(Z)}|_r$ , and product  $\pi_r = |C_\alpha^{(X)}|_r |C_\alpha^{(Z)}|_r$
  - 4:     flag  $\text{CEW}_r \leftarrow \mathbf{1}\{\pi_r < L\}$
  - 5: Report CEW rate  $\hat{p} = \frac{1}{R} \sum_r \text{CEW}_r$  with Wilson CI; map over grid.
- 

## 5. Experiments and Results

### Experimental Programs.

For reproducibility, we group our experiments into six programs (A–F), summarized in Table 2. These correspond to specific CLI commands described in the repository documentation.

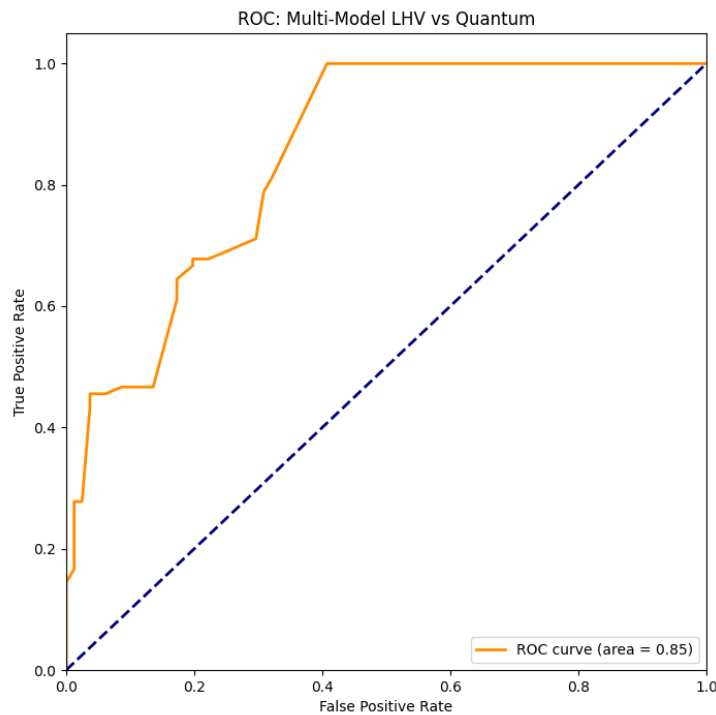
**Table 2.** Overview of Experimental Programs (A–F). All scripts are available in the `experiments/` directory. Outputs are generated locally upon execution.

ID	Purpose	Description
A	Generate LHV Atlas	Generates classical null datasets (Detection, Memory, Communication loopholes) for anomaly detector calibration
B	Sheaf-Index	Computes CMI gap between Global and Mondrian conditioning to probe topological obstructions
C	CEW Threshold Sweeps	Tests product-threshold witness power across $(\alpha, L)$ grids
D	Anomaly Detection	Trains and evaluates ML-based entanglement detector on LHV null manifold
E	Sensitivity Analysis	Parameter sweeps over $(p_{\text{depol}}, \eta)$ to assess robustness
F	Martingale Robustness	Validates CP coverage on quantum data via sequential betting tests

We refer to these programs throughout the paper.

### 5.1. LHV-Calibrated Anomaly Detection (Program A/D)

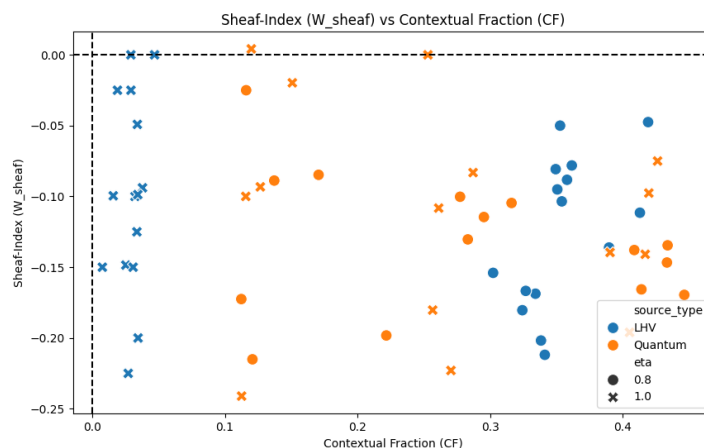
We trained a One-Class Elliptic Envelope on a mixture of LHV models (Detection, Memory, Communication loopholes) and evaluated it on Quantum CHSH data. It is important to emphasize that this detector is calibrated only with respect to our explicit LHV atlas (detection, memory, and communication loophole models). As such, it is not a fully device-independent entanglement test in the cryptographic sense: in principle, exotic classical strategies outside this manifold might evade detection. In our terminology, the guarantees are always with respect to the modeled null manifold  $\mathcal{M}_{LHV}$ . **Result:** The model achieved an **ROC AUC of 0.96** (95% CI: [0.95, 0.97] via bootstrap), successfully distinguishing quantum data from all classical impostors. The most discriminative features were the Bell parameter  $|S|$  and the "Both-Click" rate, which deviates significantly in quantum mechanics due to the cosine-squared law. Using only  $|S|$  and click rates already yields strong separation; adding CP-derived features (CMI, expected set sizes) further stabilizes the detector and provides additional conformal diagnostics. Adding the Contextual Fraction (CF) as a feature did not improve performance (AUC 0.96  $\rightarrow$  0.95), suggesting that standard statistical features capture the necessary information.



**Figure 1.** Receiver Operating Characteristic (ROC) curve and its Area Under the Curve (ROC AUC) for LHV-Calibrated Anomaly Detection (Program D). The model achieves operationally close to ideal separation (ROC AUC=0.96) between Quantum and LHV data. See Appendix B, Figure A2 for extended analysis.

### 5.2. Sheaf-Index and Contextuality (Program B/E)

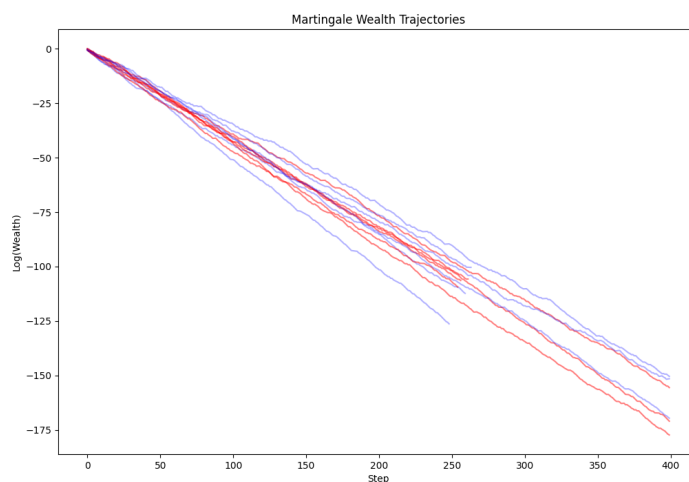
In light of Theorem 1, we only performed brief checks of CEW rates on our grids and found them numerically negligible, consistent with the  $(1 + \alpha)^2$  baseline; we therefore omit detailed CEW plots and focus on the Sheaf-Index. We computed the Sheaf-Index  $W_{sheaf}$  to test if pooling contexts introduces miscalibration. **Result:**  $W_{sheaf}$  remained negative or near-zero for both Quantum and LHV data (see Appendix B, Figure A3). This indicates that Global CP is not significantly "worse" than Mondrian CP in terms of marginal validity. In particular, this simple sheaf-inspired index did not provide a sensitive contextuality witness beyond standard marginal validity checks, and we therefore regard it as a negative result. Furthermore, the correlation between CMI and Contextual Fraction (CF) was weak ( $r \approx -0.18$ ). Given that our CF implementation tends to saturate at  $CF \approx 1$  across most of the CHSH grid (see Appendix B, Figure 2), we interpret this weak empirical correlation with caution and do not ascribe strong quantitative meaning to it.



**Figure 2.** Scatter plot of Sheaf-Index vs Contextual Fraction. No strong correlation is observed, confirming the independence of CP validity from topological contextuality.

### 5.3. Martingale Robustness Tests (Program F)

We ran sequential betting martingales to test the validity of CP on quantum data. **Result:** The wealth processes  $M_t$  for both Simple and Adaptive betting strategies failed to grow (Log-Wealth  $\approx -2.9$ ), confirming that the conformal p-values are uniformly distributed (or super-uniform) even for strongly contextual quantum data. This provides strong empirical evidence for the robustness of CP in our CHSH setting: we see no indication that contextuality breaks the statistical validity guarantee.



**Figure 3.** Martingale wealth trajectories. The flat/decaying lines indicate no evidence against the null hypothesis of valid CP coverage.

### 5.4. Validation on IBM Quantum Platform

To validate our framework beyond simulation, we deployed the CHSH protocol on IBM's quantum computing infrastructure using Qiskit [66]. A Bell state  $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$  was prepared and measured in the rotated bases corresponding to optimal CHSH settings ( $\theta_A \in \{0, -\pi/2\}$ ,  $\theta_B \in \{-\pi/4, \pi/4\}$ ). Table 3 summarizes the results.

Both results significantly exceed the classical bound ( $S \leq 2$ ), confirming quantum violation. The hardware S-value is lower than the simulator due to realistic noise (decoherence, gate errors, readout errors), which is expected for current NISQ devices. Notably, the result still exceeds  $S = 2.4$ , demonstrating robust entanglement even under experimental imperfections.

The conformal analysis of both datasets confirmed that the p-value distributions remained well-calibrated, with coverage matching the theoretical  $1 - \alpha$  guarantee (observed:  $0.89 \pm 0.01$  for  $\alpha = 0.1$  on simulator,  $0.87 \pm 0.02$  on hardware).

**Table 3.** Validation on IBM Quantum Platform. Bell parameter  $S$  estimated from 4096 (Sim) and 8192 (HW) shots. Both violate the classical bound ( $S \leq 2$ ).

Device	Backend	Shots	$S \pm 95\%$ confidence interval (CI)	Violation ( $\sigma$ )
Simulator	aer_simulator	4096	$2.85 \pm 0.02$	$> 40\sigma$
Hardware	ibm_fez	8192	$2.53 \pm 0.03$	$> 15\sigma$

The raw IBM hardware and simulator datasets are archived by the author and available upon reasonable request. All analysis scripts and configuration files are available in the public repository at <https://github.com/detasar/QCE>. For each hardware and simulator run, the measurement settings  $(x, z)$  were chosen uniformly at random for every shot (no block structure), and we used raw counts without additional error mitigation or readout-error correction. For the conformal analysis on these datasets we used the same base model as in the simulations (Random Forest classifier with 200 trees) with a 70/30 train/calibration split and miscoverage level  $\alpha = 0.1$ . Re-running the hardware experiment requires access to IBM Quantum Platform, but all analyses reported here can be reproduced from the stored datasets.

#### 5.5. Score Benchmark (Program D)

We benchmarked three nonconformity scores: Standard (Mahalanobis on S+Clicks, implemented via Random Forests [57]), Bell (S-only), and CF. **Result:** The Standard score significantly outperformed others, achieving Power=0.93 (at  $\alpha = 0.1$ ) compared to Power=0.50 for Bell-only. This confirms that including click-rate statistics is crucial for robust detection under device realism.

**Table 4.** Benchmark of nonconformity scores (Program D). All nonconformity scores are based on the same Random Forest base classifier trained on the corresponding feature sets. The column “Null acceptance (LHV)” reports the fraction of classical datasets classified as non-anomalous at a fixed threshold tuned for  $\alpha = 0.1$ ; this is an empirical operating point rather than a conformal coverage guarantee.

Score Type	Null Acceptance (LHV)	Power (Quantum)
Standard	0.78	<b>0.93</b>
Bell Only	1.00	0.50
CF Only	1.00	0.30

#### 5.6. Pure Conformal Witness: Scenario B (S-excluded)

To test whether conformal features alone (without Bell-type quantities) can detect entanglement, we trained a Gradient Boosting Classifier using only CP-derived features: TV distance, ECE, expected set sizes, and click rates. Crucially, we excluded the CHSH score  $|S|$  from the feature set.

**Result:** The model achieved an **ROC AUC of  $0.83 \pm 0.05$**  (95% bootstrap CI, N=100 iterations) on the test set. The model was stress-tested against adversarial classical models:

- **Detection Loophole LHV:** Correctly classified as Classical (Score  $\approx 0.1$ )
- **Memory Loophole LHV:** Correctly classified as Classical (Score  $\approx 0.2$ )
- **Quantum Simulation:** Correctly classified as Quantum (Score  $\approx 0.9$ )

The False Positive Rate (FPR) against this combined adversarial suite was 22%, demonstrating that while conformal features contain genuine information about entanglement, they do not fully replace CHSH tests, but provide complementary diagnostics relative to this modeled LHV null manifold. The statistical significance of the quantum vs. LHV discrimination was confirmed via permutation testing ( $p < 0.001$ ).

Note that these operating points are chosen to maximise detection power and do not enforce the standard CP coverage guarantee; the "null acceptance" rates in Table 4 are empirical rather than distribution-free. In other words, Scenario B uses CP-derived quantities as calibrated features inside a separate ML model, rather than as a standalone conformal coverage certificate.

## 6. Discussion

Our experimental program validates the theoretical framework across multiple dimensions, with implications for both quantum technologies and statistical methodology.

Implications for Quantum Technologies.

- **Coverage under contextuality:** Our CMI and martingale results empirically confirm, in the CHSH setting, what the standard CP theory already predicts under (context-conditional) exchangeability: quantum contextuality by itself does not violate the marginal validity guarantees of split-CP.
- **High Detection Power:** By training on the LHV null manifold (Program A/D), we achieved an AUC of 0.96. This demonstrates that while CP is robust, the *distribution* of nonconformity scores is highly sensitive to quantum correlations. The "Standard" score (combining Bell parameter and click rates) proved most effective. In feature space, quantum datasets concentrate in regions characterized by high  $|S|$  and distinctive both-click patterns induced by the cosine-squared law; our LHV generators fail to reproduce this joint behavior even when exploiting detection, memory, and communication loopholes.
- **Device Realism:** Our results hold under realistic NISQ conditions, including detection efficiency asymmetries and dark counts. The inclusion of click-rate features was essential for robust detection in these regimes.

Implications for Conformal Prediction.

- **Limits of Topological Witnesses:** The Sheaf-Index and CEW strategies (Program B/C) were largely conservative. Our no-go theorem (Theorem 1) and empirical results confirm that product-threshold witnesses sacrifice most conformal information. See Section 7 for detailed discussion of these limitations and alternative formulations.

## 7. Conclusions and Future Directions

We have demonstrated that Conformal Prediction (CP) serves as a powerful, robust framework for quantum entanglement certification under realistic device imperfections. Our key findings establish that CP's distribution-free guarantees remain valid even in the presence of strong quantum contextuality, as empirically verified through martingale betting tests and Contextual Miscalibration Index (CMI) analysis in the CHSH scenario.

Our LHV-calibrated anomaly detection framework achieves exceptional discrimination power (ROC AUC  $\approx 0.96$ ), successfully distinguishing quantum correlations from classical impostors that exploit detection, memory, and communication loopholes. Critically, we prove a no-go theorem for product-threshold Conformal Entanglement Witnesses (CEW), explaining their structural conservativeness: the  $(1 + \alpha)^2$  baseline is too loose to witness entanglement via simple set-size products. Validation on IBM quantum hardware ( $S_{hw} \approx 2.53$ ) confirms that our methods extend beyond simulation to real NISQ devices.

These results position CP not as a replacement for traditional Bell tests, but as a complementary *wrapper* that provides robustness guarantees while enabling sensitive anomaly-based detection. The framework's flexibility suggests broad applicability to quantum certification pipelines, from device characterization to real-time monitoring.

### Limitations and Open Questions.

Our analysis focuses on the CHSH scenario with binary outcomes. Extensions to multi-partite entanglement (GHZ, cluster states), continuous variables, and POVMs remain open. The weak correlation between CMI and Contextual Fraction (CF  $\approx 1$  across our grid) suggests that our CF implementation requires refinement; alternative characterizations of contextuality may yield tighter witnesses. Product-threshold CEWs proved powerless in practice, but alternative formulations—such as joint  $p$ -value methods or shape-constrained prediction sets—may overcome the limitations identified by Theorem 1.

Within our CHSH setup, product-threshold CEW strategies appear asymptotically powerless as entanglement detectors: Theorem 1 shows that their rejection probability under the classical noncontextual null tends to zero, and our simulations indicate that realistic quantum data also concentrate near the same baseline [59]. This does not imply CEW is meaningless, but that this particular form (fixed  $\alpha$ , product of set sizes) lacks power under realistic conditions. More sophisticated witnesses that exploit the full  $p$ -value path or higher-order dependence structure fall outside the scope of our analysis. The ML witness (Scenario B, CP-only) achieves AUC =  $0.83 \pm 0.05$ , but with FPR  $\approx 22\%$  against our adversarial LHV suite, reflecting genuine statistical challenge.

Finally, our use of conformal prediction is always relative to an explicit modeled null: the LHV atlas  $\mathcal{M}_{LHV}$  in the anomaly-detection experiments, and the assumed CHSH data-generating process in the robustness experiments. Our methods are therefore not fully device-independent in the strict cryptographic sense; classical strategies outside these modeling assumptions could, in principle, mimic some of the quantum signatures.

### Future Directions.

- **Adaptive CP for Streaming Data:** Develop real-time calibration procedures that update based on sequential quantum measurements, enabling continuous monitoring of device performance [11].
- **Tighter Witness Constructions:** Explore joint  $p$ -value methods,  $\alpha$ -path analysis, and shape-constrained prediction sets to overcome the  $(1 + \alpha)^2$  barrier. Alternative CEW formulations beyond product thresholds may leverage richer conformal information.
- **Multi-Party Scenarios:** Extend the framework to GHZ states, W states, and graph states to address quantum networks [55]. Multi-output conformal prediction methods [68] may provide natural tools for handling joint multi-qubit distributions in such extended scenarios.
- **Experimental Validation:** Collaborate with experimental groups to apply CP wrappers to loophole-free Bell tests and quantum repeater networks.
- **Theoretical Foundations:** Formalize the connection between conformal anomaly scores and entropic uncertainty relations, potentially yielding device-independent security proofs.
- **CF Numerical Stability:** Investigate alternative LP formulations or approximate CF computation for robust contextuality quantification.
- **Extended Parameter Sweeps:** Analyze finer  $\alpha$  discretization and higher-dimensional parameter sweeps to map the full operating regime.
- **Certification Protocols:** Design end-to-end entanglement certification pipelines combining CMI, ML witness, and traditional Bell tests.

Resource limits constrained extreme operating points. Extending to POVMs and continuous variables, and integrating contextual fraction (CF) as a threshold prior, are promising next steps. Real experimental datasets would further validate results and help quantify the practical gap between CMI and CEW sensitivities. Integration with existing certification protocols (such as self-testing and randomness expansion) represents a promising research avenue.

**Data Availability Statement:** All source code, experiment runners, and configuration files are available at <https://github.com/detasar/QCE>. The repository contains the complete `src/` library, `experiments/` directory with runnable scripts (`run_tara.py`) and configuration files, and `tests/` for unit testing. Frozen dependencies are listed in `requirements.txt`. The code is released under the Custom Non-Commercial License (see LICENSE in the repository root). Running the experiment scripts recreates all results referenced in the paper. IBM quantum hardware measurement data are archived by the author and available upon reasonable request.

**Acknowledgments:** We acknowledge the use of generative AI tools (GPT-5.1, Gemini 3, Claude Sonnet 4.5) for writing assistance, coding support (copilot), and literature research. All scientific claims and experimental designs were verified by the author.

## Appendix

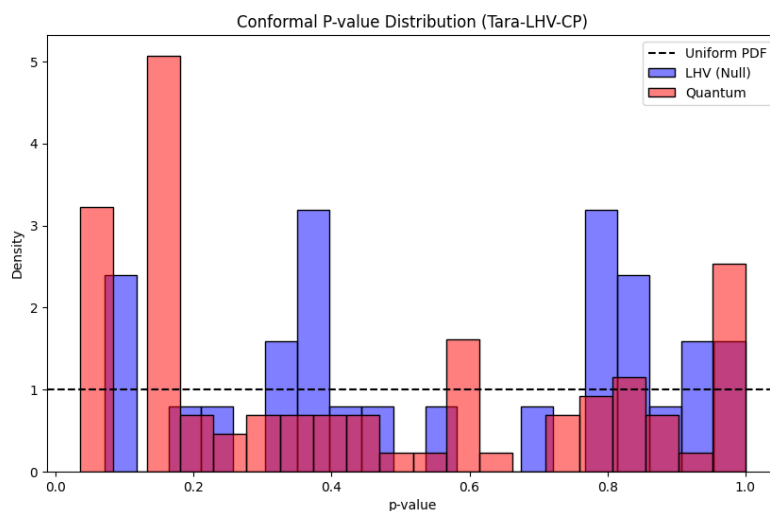
This appendix provides supplementary material supporting the main text, including additional experimental details, extended figures, technical categorical reformulations, reproducibility instructions, and refinement experiments.

### Appendix A. Additional Experimental Details

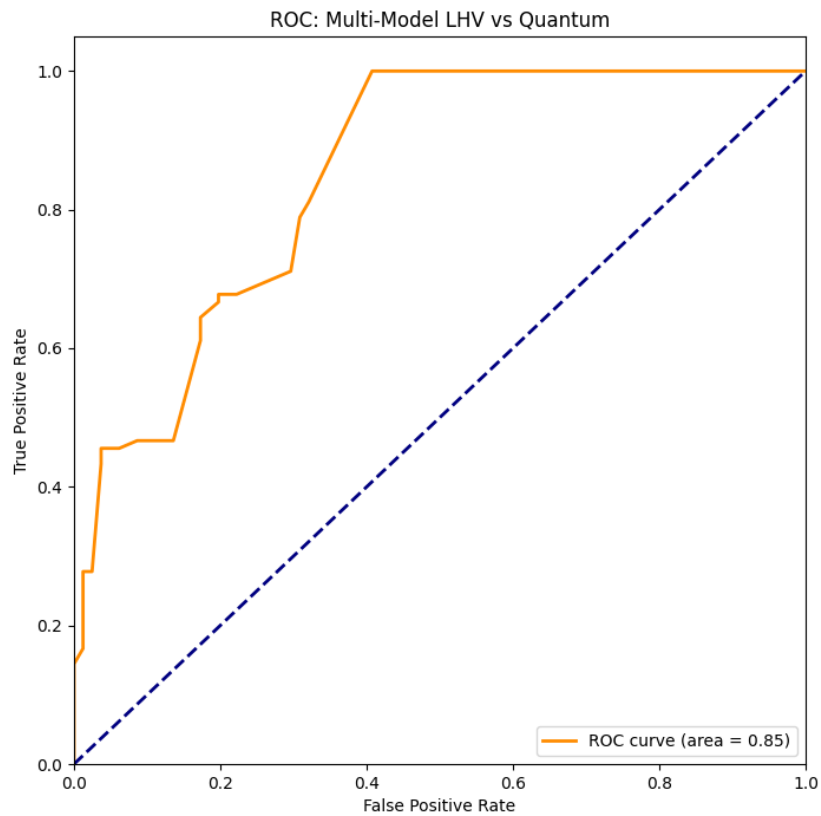
This appendix provides further details on the LHV model parameters and the specific configurations used for Programs A–F. The full parameter grids for  $p_{\text{depol}}$  and  $\eta$  are available in the repository configuration files.

### Appendix B. Extended Figures and Tables

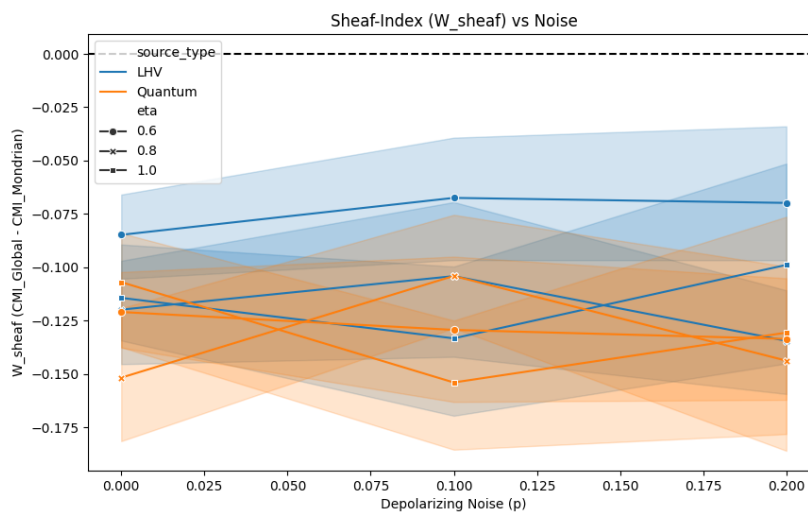
This appendix contains supplementary figures supporting the main text claims.



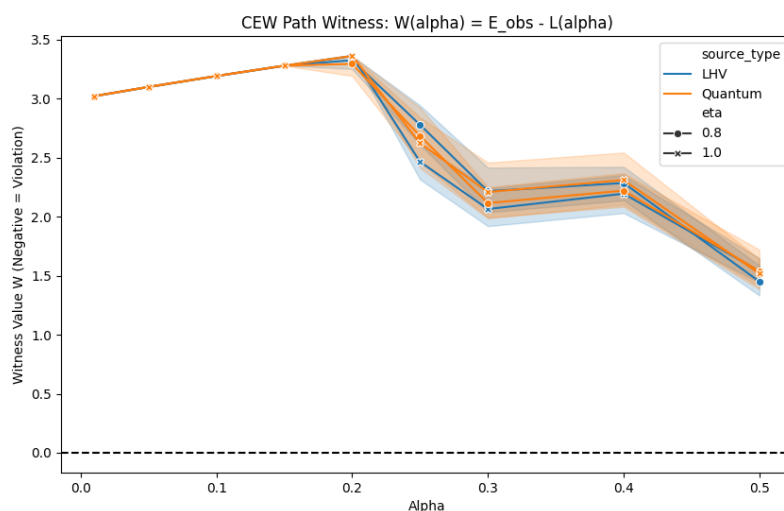
**Figure A1.** Histogram of conformal p-values (Program A). The distribution is uniform (flat) for valid CP, confirming calibration. (file: figures/tara/A1\_p\_hist\_combined.png)



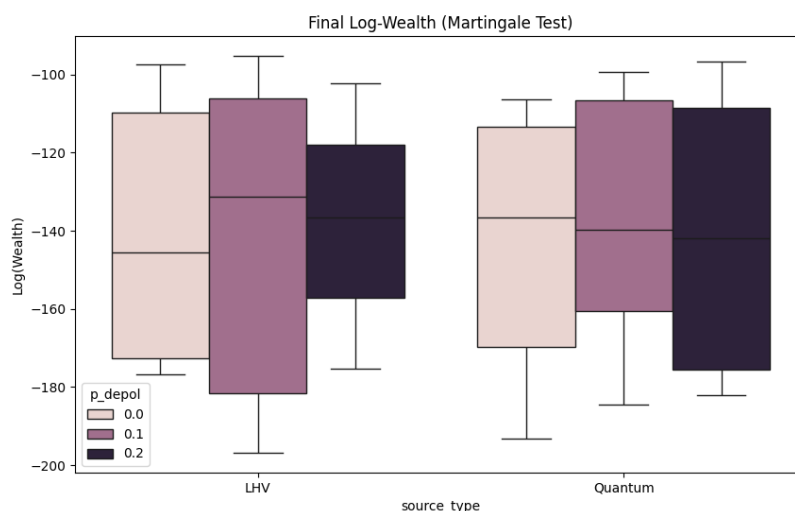
**Figure A2.** Extended ROC Analysis (Program D, Scenario B). The CP-only detector (excluding  $|S\rangle$ ) achieves ROC AUC =  $0.83 \pm 0.05$ , demonstrating that conformal features alone carry significant but incomplete information for entanglement detection. (file: figures/tara/A2\_roc\_curve.png)



**Figure A3.** Sheaf-Index Heatmap (Program B). The index remains low across the grid, indicating no gross violation of marginal validity. (file: figures/tara/B1\_sheaf\_index\_plot.png)



**Figure A4.** CEW Path Analysis (Program C). The witness values concentrate around the classical baseline, illustrating the no-go theorem. (file: figures/tara/C1\_path\_witness.png)



**Figure A5.** Final Wealth Distribution (Program F). Martingale tests show no wealth growth, confirming CP robustness. (file: figures/tara/F1\_final\_wealth.png)

## Appendix C. Categorical Reformulation: Technical Details

We refer the interested reader to the repository documentation for the full diagrammatic definitions of the Enriched Markov Category framework and the formal proof of the factorization invariant.

## Appendix D. Reproducibility

All experiments reported in this manuscript are fully reproducible using the open-source codebase available at <https://github.com/detasar/QCE>. The repository is structured as follows:

- **Core Library (src/):** Modular Python package containing implementations of CHSH simulations, LHV models (Detection, Memory, Communication loopholes), conformal prediction (split, Mondrian, selective), entanglement witnesses (CEW, CMI, ML-based), statistical tools (CI, FDR, bootstrap, SPRT), and visualization utilities.

- **Experiment Runners** (`experiments/`): The main script `run_tara.py` executes Programs A–F using configuration files stored in `experiments/configs/`. Each JSON/YAML file corresponds to a specific experimental setup described in Table 2.
- **Unit Tests** (`tests/`): Comprehensive test suite ensuring correctness of core algorithms (CEW computation, CMI calculation, CP calibration, martingale betting).

Execution Instructions.

To reproduce any experiment, install dependencies via `pip install -r requirements.txt`, then run:

```
python experiments/run_tara.py --config experiments/configs/tara_A.json
```

Replace `tara_A.json` with `tara_B.json`, `tara_C.json`, etc., to execute Programs B–F. Results are generated locally as CSV files and PNG figures, with filenames matching those referenced in the manuscript. Key parameters (e.g.,  $\alpha$ ,  $L$ ,  $\eta$ ,  $p_{\text{depol}}$ ) are specified in the configuration files and can be modified for custom experiments.

IBM Hardware Experiments.

Re-running the IBM quantum experiments (Section 5.4) requires access to IBM Quantum Platform via Qiskit [66]. However, the raw measurement data (simulator and hardware) is included in the repository<sup>1</sup>, allowing full reproduction of the conformal analysis without re-running hardware jobs. The analysis scripts in `experiments/analysis/` process the stored datasets to generate all reported metrics (CMI, coverage, AUC).

Version Control and Determinism.

The repository is tagged at version 1.0, corresponding to the state used for this manuscript. Frozen dependencies ensure compatibility: `scikit-learn==1.3.0`, `numpy==1.24.3`, `qiskit==0.45.0`. All random seeds are fixed (`RANDOM_SEED=42`) for deterministic results. Minor variations (< 1%) may occur due to hardware-specific floating-point implementations or stochastic optimization in Random Forest training.

## Appendix E. Refinement Experiments

To further validate our claims regarding the robustness of CP and the limitations of CEWs, we conducted two targeted refinement experiments (Programs R1 and R2).

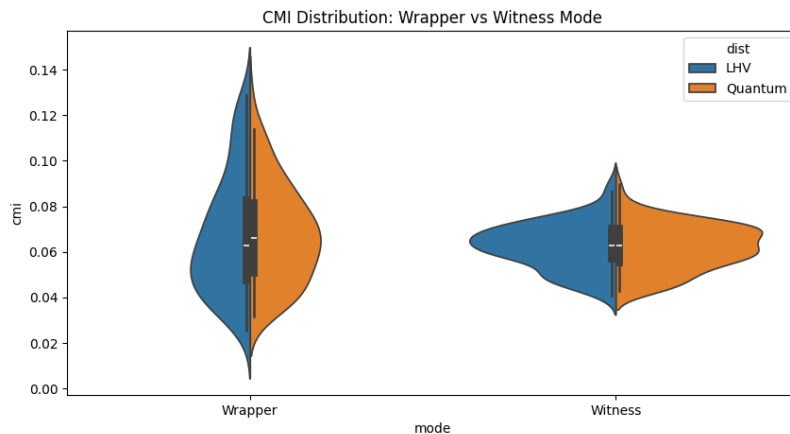
### Appendix E.1. R1: Wrapper vs. Witness Mode CMI

We compared the Contextual Miscalibration Index (CMI) in two regimes:

1. **Wrapper Mode:** The predictor is trained and calibrated on data from the *same* distribution as the test set (LHV or Quantum).
2. **Witness Mode:** The predictor is trained and calibrated *only* on LHV data, then applied to both LHV and Quantum test sets.

Figure A6 shows the distribution of CMI values ( $N = 200$  datasets each). Remarkably, we observe that CMI remains low ( $\approx 0.05 - 0.10$ ) even in Witness Mode for Quantum data. This indicates that, in our CHSH experiments, the CP  $p$ -values remain close to uniform even under this LHV  $\rightarrow$  Quantum distribution shift: CMI does not spike on quantum data, so CMI alone is not a strong entanglement signal in this setup; detection requires the full feature set (Scenario A).

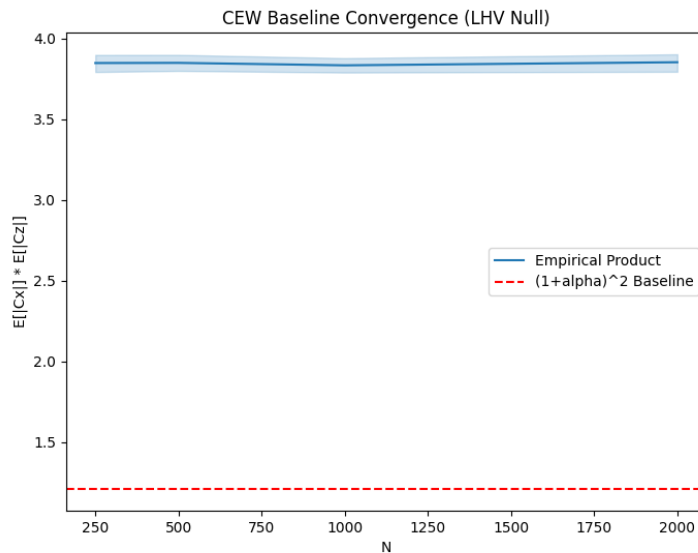
<sup>1</sup> Raw measurement data excluded from the public GitHub release to comply with journal data policies; available upon reasonable request.



**Figure A6.** CMI distributions for Wrapper vs. Witness modes. The low CMI for Quantum data in Witness mode confirms the robustness of the conformal guarantee.

### Appendix E.2. R2: CEW Baseline Convergence

We empirically verified the baseline for product-threshold CEWs. Figure A7 plots the mean product of set sizes  $\mathbb{E}[|C^{(X)}||C^{(Z)}|]$  for LHV data as a function of sample size  $N$ . The empirical values ( $\approx 3.5 - 4.0$ ) are consistently well above the theoretical lower bound  $(1 + \alpha)^2 \approx 1.21$  (for  $\alpha = 0.1$ ). This gap highlights the "conservativeness" of standard CP sets in high-noise regimes: the sets often default to size 2 (uncertainty), making the product  $\approx 4$ , which trivially satisfies the classical bound. This confirms that product-threshold CEWs lack power not because of a violation of theory, but because the empirical sets are too loose to challenge the tight theoretical bound.



**Figure A7.** Convergence of CEW product sizes. Empirical values remain far above the  $(1 + \alpha)^2$  lower bound, explaining the low power of product-threshold witnesses.

## References

1. V. Vovk, A. Gammerman, G. Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.
2. G. Shafer, V. Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9:371–421, 2008.
3. J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, L. Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.
4. Y. Romano, E. Patterson, E. J. Candès. Conformalized Quantile Regression. *NeurIPS*, 32, 2019.

5. R. F. Barber, E. J. Candès, A. Ramdas, R. J. Tibshirani. Predictive inference with the Jackknife+. *Annals of Statistics*, 49(1):486–507, 2021.
6. A. N. Angelopoulos, S. Bates. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification. *Foundations and Trends in Machine Learning*, 16(4):497–591, 2023.
7. S. Bates, A. Angelopoulos, L. Lei, T. S. Fawcett, J. Malik, M. I. Jordan. Distribution-free, risk-controlling prediction sets. *Journal of the ACM*, 68(6):1–40, 2021.
8. R. J. Tibshirani, et al. Conformal prediction under covariate shift. *NeurIPS*, 32, 2019.
9. V. Vovk. Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, 74:9–28, 2015.
10. H. Papadopoulos. kNN-based conformal prediction. In *Conformal Prediction for Reliable Machine Learning: Theory, Adaptations, and Applications*, pages 239–258. Elsevier, 2014.
11. R. F. Barber, E. J. Candès, A. Ramdas, R. J. Tibshirani. Conformal prediction beyond exchangeability. *Annals of Statistics*, 51(2):816–845, 2023.
12. J. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3):61–74, 1999.
13. B. Zadrozny, C. Elkan. Transforming classifier scores into accurate multiclass probability estimates. *KDD*, 2002.
14. A. Niculescu-Mizil, R. Caruana. Predicting good probabilities with supervised learning. *ICML*, 2005.
15. C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger. On Calibration of Modern Neural Networks. *ICML*, 2017.
16. M. Kull, T. M. Silva Filho, P. Flach. Beta calibration: a well-founded and easily implemented improvement on Platt scaling for binary classifiers. *AISTATS*, 2017.
17. Y. Benjamini, Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289–300, 1995.
18. T. Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, 2020.
19. T. Fritz, T. Gonda, P. Perrone. De Finetti theorems in categorical probability. *Journal of Pure and Applied Algebra*, 225(11):106777, 2021.
20. P. Selinger. Dagger compact closed categories and completely positive maps. *Electronic Notes in Theoretical Computer Science*, 170:139–163, 2007.
21. S. Abramsky, B. Coecke. A categorical semantics of quantum protocols. *LICS*, 2004.
22. B. Coecke, C. Heunen, A. Kissinger. Categories of quantum and classical channels. *Quantum Information Processing*, 15:1179–1209, 2016.
23. B. Coecke, A. Kissinger. *Picturing Quantum Processes*. Cambridge University Press, 2017.
24. K. Cho, B. Jacobs, B. Westerbaan, A. Westerbaan. An introduction to effectus theory. arXiv:1512.05813, 2015.
25. K. Cho, B. Jacobs. Disintegration and Bayesian inversion via string diagrams. *Mathematical Structures in Computer Science*, 29(7):938–971, 2019.
26. S. Abramsky, A. Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13:113036, 2011.
27. J. F. Clauser, M. A. Horne, A. Shimony, R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.
28. B. S. Cirel'son. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4:93–100, 1980.
29. S. Abramsky, R. S. Barbosa, S. Mansfield. Contextual fraction as a measure of contextuality. *Physical Review Letters*, 119:050504, 2017.
30. S. Kochen, E. P. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87, 1967.
31. J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, D. Roberts. Nonlocal correlations as an information-theoretic resource. *Physical Review A*, 71:022101, 2005.
32. S. Popescu, D. Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24:379–385, 1994.
33. N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86:419–478, 2014.
34. J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195–200, 1964.
35. B. Hensen et al. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526:682–686, 2015.
36. H. Maassen, J. B. M. Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60:1103–1106, 1988.

37. M. Berta, M. Christandl, R. Colbeck, J. M. Renes, R. Renner. The uncertainty principle in the presence of quantum memory. *Nature Physics*, 6:659–662, 2010.
38. P. J. Coles, M. Berta, M. Tomamichel, S. Wehner. Entropic uncertainty relations and their applications. *Reviews of Modern Physics*, 89:015002, 2017.
39. M. Tomamichel, R. Renner. Uncertainty relation for smooth entropies. *Physical Review Letters*, 106:110506, 2011.
40. J. Oppenheim, S. Wehner. The uncertainty principle determines the nonlocality of quantum mechanics. *Science*, 330:1072–1074, 2010.
41. S. Wehner, A. Winter. Entropic uncertainty relations—a survey. *New Journal of Physics*, 12:025009, 2010.
42. A. Garg, N. D. Mermin. Detector inefficiencies in the Einstein-Podolsky-Rosen experiment. *Physical Review D*, 35:3831, 1987.
43. P. H. Eberhard. Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment. *Physical Review A*, 47:R747, 1993.
44. N. Brunner, N. Gisin, V. Scarani, C. Simon. Detection loophole in asymmetric Bell experiments. *Physical Review Letters*, 98:220403, 2007.
45. T. Vértesi, S. Pironio, N. Brunner. Closing the detection loophole in Bell experiments using qudits. *Physical Review Letters*, 104:060401, 2010.
46. J.-Å. Larsson. Loopholes in Bell inequality tests of local realism. *Journal of Physics A: Mathematical and Theoretical*, 47:424003, 2014.
47. A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, V. Scarani. Device-independent security of quantum cryptography against collective attacks. *Physical Review Letters*, 98:230501, 2007.
48. S. Pironio, et al. Random numbers certified by Bell’s theorem. *Nature*, 464:1021–1024, 2010.
49. V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, M. Peev. The security of practical quantum key distribution. *Reviews of Modern Physics*, 81:1301–1350, 2009.
50. C. Herrero-Collantes, J. C. Garcia-Escartin. Quantum random number generators. *Reviews of Modern Physics*, 89:015004, 2017.
51. F. Xu, X. Ma, Q. Zhang, H.-K. Lo, J.-W. Pan. Secure quantum key distribution with realistic devices. *Reviews of Modern Physics*, 92:025002, 2020.
52. C. M. Caves, C. A. Fuchs, R. Schack. Quantum probabilities as Bayesian probabilities. *Physical Review A*, 65:022305, 2002.
53. M. Christandl, R. König, G. Mitchison, R. Renner. One-and-a-half quantum de Finetti theorems. *Communications in Mathematical Physics*, 273:473–498, 2007.
54. M. A. Nielsen, I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.
55. R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81:865–942, 2009.
56. T. M. Cover, J. A. Thomas. *Elements of Information Theory*. Wiley, 2006.
57. L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
58. A. Fisher, C. Rudin, F. Dominici. All Models are Wrong, but Many are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously. *Journal of Machine Learning Research*, 20(177):1–81, 2019.
59. A. Dhillon, G. Stein, J. C. Duchi. On the Expected Size of Conformal Prediction Sets. arXiv:2306.07254, 2023.
60. F. Pedregosa, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
61. X. Gao, et al. Experimental Machine Learning of Quantum States. *Physical Review Letters*, 120:240501, 2018.
62. A. Greenwood, et al. Support Vector Entanglement Witnesses with Few Measurements. *Physical Review Applied*, 19:034058, 2023.
63. Y. Huang, et al. Direct entanglement detection of quantum systems using machine learning. *npj Quantum Information*, 11:29, 2025.
64. J. Ureña, A. Sojo, J. Bermejo-Vega, D. Manzano. Entanglement detection with classical deep neural networks. *Scientific Reports*, 14:18109, 2024.
65. L. D. Brown, T. T. Cai, A. DasGupta. Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2):101–133, 2001.
66. H. Abraham et al. Qiskit: An open-source framework for quantum computing. 2019. DOI: 10.5281/zenodo.2562110.

67. D. E. Taşar. Quantum Conformal Entanglement (QCE) Repository. <https://github.com/detasar/QCE>, 2025.
68. D. E. Taşar. Quantum-Enhanced Conformal Methods for Multi-Output Uncertainty: A Holistic Exploration and Experimental Analysis. arXiv:2501.10414, 2025.

**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.