

# DIN SPEC 91527:2025-12 (E)

## Goals, Methods and Metrics for Automated/Semi-Automated Runtime Monitoring of AI Systems for Non-Adversarial Performance Degradations; Text in English

---

| Contents  | Page |
|---|------|
| Foreword . . . . .  | 5    |
| Introduction . . . . .  | 6    |
| 1 Scope . . . . .   | 7    |
| 2 Normative references . . . . .  | 7    |
| 3 Terms and definitions . . . . .   | 7    |
| 4 General . . . . .   | 9    |
| 5 List of Metrics . . . . .   | 9    |
| 5.1 Introduction . . . . .  | 9    |
| 5.2 How to use the list of metrics . . . . .                                  | 9    |
| 5.3 How to extend the list of metrics . . . . .                               | 10   |
| 5.4 Example of metric selection and application . . . . .                     | 11   |
| 6 Formal framework: Metric Cards and Monitoring Cards . . . . .               | 12   |
| 6.1 Introduction . . . . .  | 12   |
| 6.2 Stages . . . . .  | 14   |
| 6.2.1 Introductory remarks . . . . .  | 14   |
| 6.2.2 Template Metric Card definition stage . . . . .                         | 14   |
| 6.2.3 Monitoring Card definition stage . . . . .                              | 14   |
| 6.2.4 Runtime evaluation stage . . . . .                                      | 14   |
| 6.3 Metric Cards . . . . .  | 14   |
| 6.3.1 Introductory remarks . . . . .  | 14   |
| 6.3.2 Concept of a metric . . . . .   | 15   |
| 6.3.3 Structure of a Metric Card . . . . .                                    | 15   |
| 6.4 Monitoring Card . . . . .   | 19   |
| 6.4.1 Introductory remarks . . . . .  | 19   |
| 6.4.2 Structure of a Monitoring Card . . . . .                                | 19   |
| Annex A (normative) List of metrics . . . . .                                 | 20   |
| A.1 Metrics . . . . .   | 20   |
| A.1.1 General . . . . .   | 20   |
| A.1.2 AM Score . . . . .  | 20   |
| A.1.3 Absolute Open-Set Error (A-OSE) . . . . .                               | 20   |
| A.1.4 Adaptive Inversion . . . . .  | 21   |
| A.1.5 Adjusted R2 . . . . .   | 21   |
| A.1.6 Area Under Receiver Operating Characteristics Curve (AUC-ROC) . . . . . | 22   |
| A.1.7 Average Odds Difference . . . . .                                       | 22   |
| A.1.8 Balanced Accuracy . . . . .   | 23   |
| A.1.9 Batch Norm Statistics . . . . .   | 23   |
| A.1.10 Bilingual Evaluation Understudy . . . . .                              | 24   |
| A.1.11 Bray Curtis Dissimilarity . . . . .                                    | 24   |
| A.1.12 Brier Score . . . . .  | 25   |
| A.1.13 CLIP-MMD . . . . .   | 25   |
| A.1.14 Categorical Distribution . . . . .                                     | 26   |
| A.1.15 Chi-Squared Test . . . . .   | 26   |
| A.1.16 Class Likelihood Ratios . . . . .                                      | 26   |
| A.1.17 Class-Aware Frechet Inception Distance (CAFD) . . . . .                | 27   |
| A.1.18 Cosine Similarity . . . . .  | 27   |
| A.1.19 Coverage Error . . . . .   | 28   |
| A.1.20 Critical Error Rate (CER) . . . . .                                    | 28   |
| A.1.21 D2 Score . . . . .   | 29   |
| A.1.22 Demographic Parity Difference . . . . .                                | 29   |
| A.1.23 Dice Coefficient . . . . .   | 30   |
| A.1.24 Difference of Means . . . . .  | 30   |

|        |   |    |
|--------|---|----|
| A.1.25 | Difference of Variance                                    | 31 |
| A.1.26 | Equal Opportunity   | 31 |
| A.1.27 | Equalized Odds  | 31 |
| A.1.28 | Explained Variance Score                                  | 32 |
| A.1.29 | F-score   | 32 |
| A.1.30 | Fréchet AutoEncoder Distance (FAED)                       | 33 |
| A.1.31 | Fréchet Inception Distance (FID)                          | 33 |
| A.1.32 | Generalized Entropy Index                                 | 34 |
| A.1.33 | Generative Adversarial Metric (GAM)                       | 34 |
| A.1.34 | Gini Coefficient  | 35 |
| A.1.35 | Hamming Distance  | 35 |
| A.1.36 | Hinge Loss  | 36 |
| A.1.37 | Hit Rate  | 36 |
| A.1.38 | Inception Score (IS)                                      | 37 |
| A.1.39 | Interquartile Range (IQR)                                 | 37 |
| A.1.40 | Jaccard Index/Intersection over Union (IoU)               | 38 |
| A.1.41 | Jensen-Shannon Divergence                                 | 38 |
| A.1.42 | Kernel Inception Distance (KID)                           | 38 |
| A.1.43 | Kolmogorov-Smirnov Test                                   | 39 |
| A.1.44 | Kullback-Leibler Divergence                               | 39 |
| A.1.45 | Label Ranking Average Precision                           | 40 |
| A.1.46 | Label Ranking Loss  | 40 |
| A.1.47 | Learned Perceptual Image Patch Similarity (LPIPS)         | 41 |
| A.1.48 | Length  | 41 |
| A.1.49 | Linear Separability                                       | 42 |
| A.1.50 | Log Loss  | 42 |
| A.1.51 | Mahalanobis Distance                                      | 43 |
| A.1.52 | Matthews Correlation Coefficient                          | 43 |
| A.1.53 | Max Error   | 44 |
| A.1.54 | Mean Absolute Error (MAE)                                 | 44 |
| A.1.55 | Mean Absolute Percentage Error (MAPE)                     | 45 |
| A.1.56 | Mean Reciprocal Rank                                      | 45 |
| A.1.57 | Mean Squared Error (MSE)                                  | 46 |
| A.1.58 | Mean Squared Logarithmic Error (MSLE)                     | 46 |
| A.1.59 | Mean Tweedie Deviance                                     | 47 |
| A.1.60 | Median Absolute Error (MedAE)                             | 47 |
| A.1.61 | Memorization-Informed Frechet Inception Distance (MiFID)  | 47 |
| A.1.62 | Mode Score  | 48 |
| A.1.63 | Modified Inception Score (m-IS)                           | 48 |
| A.1.64 | Multi-Scale Structural Similarity Index Measure (MS-SSIM) | 49 |
| A.1.65 | Norm Distance   | 49 |
| A.1.66 | Normalized Discounted Cumulative Gain                     | 50 |
| A.1.67 | Panoptic Quality (PQ)                                     | 50 |
| A.1.68 | Parsing Covering (PC)                                     | 51 |
| A.1.69 | Peak Signal-to-Noise Ratio (PSNR)                         | 51 |
| A.1.70 | Percentile  | 52 |
| A.1.71 | Perceptual Path Length (PPL)                              | 52 |
| A.1.72 | Perplexity  | 53 |
| A.1.73 | Pinball Loss  | 53 |
| A.1.74 | Pixel Accuracy  | 54 |
| A.1.75 | Population Stability Index                                | 54 |
| A.1.76 | Precision   | 55 |
| A.1.77 | Predictive Parity   | 55 |
| A.1.78 | R2 Score  | 56 |
| A.1.79 | Recall  | 56 |
| A.1.80 | Recall-Oriented Understudy for Gisting Evaluation         | 56 |
| A.1.81 | Root Mean Squared Error (RMSE)                            | 57 |

|         |  |    |
|---------|--|----|
| A.1.82  | Smoothed empirical differential fairness           | 57 |
| A.1.83  | Spatial FID (sFID)                                 | 58 |
| A.1.84  | Structural Similarity Index Measure (SSIM)         | 58 |
| A.1.85  | Theil Index  | 59 |
| A.1.86  | Top-k Accuracy                                     | 59 |
| A.1.87  | Total Variation Distance                           | 60 |
| A.1.88  | Toxicity   | 60 |
| A.1.89  | WaM  | 61 |
| A.1.90  | Wasserstein Distance                               | 61 |
| A.1.91  | Wilderness Impact (WI)                             | 62 |
| A.1.92  | Z-Score  | 62 |
| A.1.93  | Zero One Loss                                      | 63 |
| Annex B | (informative) Monitoring Card Examples             | 64 |
| B.1     | Fairness Monitoring Example: Credit Classification | 64 |
| B.2     | Regression Monitoring Example: Energy Prediction   | 66 |
|         | Bibliography                                       | 68 |

## Figures

|          |   |    |
|----------|---|----|
| Figure 1 | — Example overview of the interplay between Metric Cards and Monitoring Card in a practical example | 13 |
| Figure 2 | — Main stages of runtime monitoring   | 14 |

## Tables

|            |   |    |
|------------|---|----|
| Table 1    | — Metric Card structure                             | 16 |
| Table 2    | — Input Card structure                              | 17 |
| Table 3    | — Input classification values used in Input Cards   | 18 |
| Table 4    | — Input volatility values used in Input Cards       | 18 |
| Table 5    | — Output Card Structure                             | 18 |
| Table 6    | — Output parameter goals to be used in Output Cards | 19 |
| Table 7    | — Monitoring Card structure                         | 19 |
| Table B.1  | — Metric Card — Demographic Parity Difference       | 64 |
| Table B.2  | — Input Card — Sensitive Feature                    | 64 |
| Table B.3  | — Input Card — Positive Outcome                     | 65 |
| Table B.4  | — Input Card — Model Input Data                     | 65 |
| Table B.5  | — Input Card — Model Predictions                    | 65 |
| Table B.6  | — Output Card — Demographic Parity Difference       | 66 |
| Table B.7  | — Metric Card — Root Mean Squared Error             | 66 |
| Table B.8  | — Input Card — Model Prediction                     | 67 |
| Table B.9  | — Input Card — Reference Data                       | 67 |
| Table B.10 | — Output Card — Root Mean Squared Error             | 67 |