Strategic Plan

**AI KR Strategists**

For: **Artificial Intelligence Knowledge Representation Community Group (AIKR CG)**

Submitted By: **Carl Mattocks**

⮹ **CarlMattocks@WellnessIntelligence.Institute**

**Overview**

This plan defines the roles AI KR Strategists.

Contents

**No table of contents entries found.**

# Articulation

## Vision

For all AI systems to have clearly and transparently documented goals and performance data showing that they are being achieved.

## Mission

The mission of an AI Strategist is to define the purpose and goals of AI systems, as well as the KPIs by which we can determine if the system is meeting its goals.

## Scorecard

|  | **Goals** | **Objectives** | **Performance Indicators** | **Commentary** |
| --- | --- | --- | --- | --- |
| **Goals with no perspective** | Strategic Plan (see pp1) | Algorithms (see pp1) |  |  |  |  |
| Ontology (see pp1) |  |  |  |  |
| Applications (see pp1) |  |  |  |  |  |
| Requirements (see pp1) |  |  |  |  |  |
| Glossaries (see pp1) |  |  |  |  |  |
| Risks (see pp1) | Bias (see pp1) |  |  |  |  |
| Consequences (see pp1) |  |  |  |  |
| Control (see pp1) |  |  |  |  |
| Data (see pp1) |  |  |  |  |
| Governance (see pp1) |  |  |  |  |
| Privacy  (see pp1) |  |  |  |  |
| Security  (see pp1) |  |  |  |  |
| Compliance (see pp1) |  |  |  |  |  |
| Ethics (see pp1) | Accountability (see pp1) |  |  |  |  |
| Autonomy (see pp1) |  |  |  |  |
| Confidentiality (see pp1) |  |  |  |  |
| Veracity (see pp1) |  |  |  |  |
| Robustness (see pp1) |  |  |  |  |  |
| Outcomes (see pp1) |  |  |  |  |  |
| Algorithm Evaluation (see pp1) | Classification (see pp1) | Precision Recall |  |  |  |
| Accuracy |  |  |  |
| Confusion Matrix |  |  |  |
| Per-class accuracy |  |  |  |
| Log-Loss |  |  |  |
| AUC-ROC Curve |  |  |  |
| F-measure |  |  |  |
| NDCG |  |  |  |
| Regression Analysis |  |  |  |
| Quantiles of Errors |  |  |  |
| "Almost correct" predictions |  |  |  |
| Trustworthiness (see pp1) |  |  |  |  |
| KR Objects (see pp1) |  |  |  |  |  |

## Goals

### Goals with no perspective

#### Goal: Strategic Plan

**Document the vision, values, goals, objectives for one or more AIKR objects**

An AI KR Object may be :

* an algorithm (example - enable an entity to determine consequences; a set of instructions that provide the ability to monitor and/or move the environment; the rules that are used to change/manipulate/interpret data)
* an ontology (which has a set of ontological commitments) See Goal - Ontological Statements (provides sufficient definition to allow measurement to be performed) (see pp1)
* an Intelligent Reasoning (fragmentary) Theory, such as,
	+ deduction,
	+ induction,
	+ abduction,
	+ by analogy,
	+ probabilistic,
	+ case-based
* a Reasoning Mechanism (computational environment), such as,
	+ natural language processer,
	+ rules engine,
	+ machine learning
* a Vocabulary (medium of human expression)

##### Objective: Algorithms

Algorithms' capabilities and limitations should be explained in a manner adapted to stakeholder concerns.

##### Objective: Ontology

The AI Strategist will work, with the AI KR Strategist and / or other experts, to ensure that ontology content mitigates bias by employing a complete glossary of all the data used and utilizing an accurate representation of the (data) relationship rules identified in processing instructions

#### Goal: Applications

**Understand the potential applications of AI to business strategies.**

#### Goal: Requirements

**Identify which areas of the requirements warrant AI solutions versus which can be achieved with other types of solutions**

#### Goal: Glossaries

**Employ definitions from one or more glossaries when explaining AIKR object audit data, veracity facts and (human, social and technology) risk mitigation factors**

So that (business) people more readily understand the value that the glossaries bring.

#### Goal: Risks

**Identify and mitigate risks and known threats**

A guiding principle is that AIKR systems must mitigate risks.

**Organisations:**

DARPA:

This goal arose in reference to the DARPA initiative.

##### Objective: Bias

Bias is disproportionate weight in favour of or against an idea or thing, usually in a way that is closed-minded, prejudicial or unfair.

A bias is a systematic error.

##### Objective: Consequences

"Environment" includes the natural environment, as well as socio-economic and societal environments.

* Minimise the risk of unintended consequences.
* AI shall do no harm
* When you're testing something, you should not alter the environmental conditions.
* The social and societal impact should be carefully considered

##### Objective: Control

Control:

* control of the algorithms: To stop them from learning beyond our ability to control them.
* control of the people who develop (strategists and developers) and use algorithms: they can be used for good or for evil.

##### Objective: Data

Data quality: the data is fit for its intended purpose/use.

Data integrity: is the maintenance and assurance of the accuracy and consistency of data over its entire life-cycle.

##### Objective: Governance

Data and algorithm governance.

##### Objective: Privacy

Ensuring data is processed with the permission of the people to whom it pertains.

E.g. GDPR, intellectual property, etc.

##### Objective: Security

Security means protection as well as the measures taken to be safe or protected.

#### Goal: Compliance

**Ensure AI Systems comply with all applicable laws and regulations, such as, provision audit data defined by a governance operating model**

#### Goal: Ethics

**Ensure AI Systems adhere to principles of ethics**

##### Objective: Accountability

##### Objective: Autonomy

Oversight controls will enable the assessment of algorithms, data and design processes

##### Objective: Confidentiality

##### Objective: Veracity

#### Goal: Robustness

**Ensure AI Systems are designed to handle uncertainty and tolerate perturbation from a likely threat perspective, such as, design considerations incorporate human, social and technology risk factors**

#### Goal: Outcomes

**Track AIKR object performance outcome via KPI (Key Performance Indicator) based on supervised learning models measurements**

#### Goal: Algorithm Evaluation

**Evaluate Algorithms**

Assess how well Algorithm results match actual outcomes to determine

* how sensitive inferences made are to the parameters and
* the proportion of observations made were accurately predicted.

When needed the algorithmic impact assessments will also identify cause and effect of any biases.

**Organisations:**

Artificial Intelligence Knowledge Representation Community Group (AIKR CG) - Community of Interest

##### Objective: Classification

**Ontological Statement**: Classification Accuracy is the ratio of number of correct class label predictions to the total number of input samples data.

**Ontological Statement**: F1 Score measure the Harmonic Mean between precision and recall. The range for F1 Score is [0, 1]. It tells you how precise your classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances).

**Performance Indicator: Precision Recall**

Frequency: Monthly

Ontological Statement: Precision is the number of correct positive results divided by the number of positive results predicted by the classifier.

Ontological Statement: Recall is the number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

**Performance Indicator: Accuracy**

Frequency: Monthly

**Ontological Statement:** Classification Rate or Accuracy is given by the relation: True Positives + True Negatives / All Instances (True & False Positives + True & False Negatives)

**Performance Indicator: Confusion Matrix**

Frequency: Monthly

Ontological Statement: A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class (the types of errors being made)

Types :

* True Positives : The cases in which we predicted YES and the actual output was also YES.
* True Negatives : The cases in which we predicted NO and the actual output was NO.
* False Positives : The cases in which we predicted YES and the actual output was NO.
* False Negatives : The cases in which we predicted NO and the actual output was YES.

Accuracy for the matrix can be calculated by taking average of the values lying across the “main diagonal”

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type** | **StartDate** | **EndDate** |  | **Description** |
| Target |  |  |  | Number of True Positives |
| Target |  |  |  | Number of False Positives |
| Target |  |  |  | Number of True Negatives |
| Target |  |  |  | Number of False Negatives |
| Actual |  |  |  | [To be determined] |

**Performance Indicator: Per-class accuracy**

Frequency: Monthly

**Performance Indicator: Log-Loss**

Frequency: Monthly

**Ontological Statement:** Logarithmic loss (related to cross-entropy) measures the performance of a classification model where the prediction input is a probability value between 0 and 1 - Log loss increases as the predicted probability diverges from the actual label

Logarithmic Loss or Log Loss, works by penalising the false classifications. It works well for multi-class classification. When working with Log Loss, the classifier must assign probability to each class for all the samples. where,

y\_ij, indicates whether sample i belongs to class j or not

p\_ij, indicates the probability of sample i belonging to class j

Log Loss has no upper bound and it exists on the range [0, ∞). Log Loss nearer to 0 indicates higher accuracy, whereas if the Log Loss is away from 0 then it indicates lower accuracy.

In general, minimising Log Loss gives greater accuracy for the classifier.

**Performance Indicator: AUC-ROC Curve**

Frequency: Monthly

Ontological Statement: check performance of multi - class classification AUROC (Area Under the Receiver Operating Characteristics) curve.

Ontological Statement: Area Under Curve(AUC) is one of the most widely used metrics for evaluation. It is used for binary classification problem. AUC of a classifier is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example.

True Positive Rate (Sensitivity) : True Positive Rate is defined as TP/ (FN+TP). True Positive Rate corresponds to the proportion of positive data points that are correctly considered as positive, with respect to all positive data points.

False Positive Rate (Specificity) : False Positive Rate is defined as FP / (FP+TN). False Positive Rate corresponds to the proportion of negative data points that are mistakenly considered as positive, with respect to all negative data points.

**Performance Indicator: F-measure**

Frequency: Monthly

F1 Score is the Harmonic Mean between precision and recall.

**Ontological Statement:** F-measure represents both Precision and Recall it helps to have a measurement that represents both of them. F-measure is calculated using Harmonic Mean (in place of Arithmetic Mean).

**Ontological Statement**:  Mean Absolute Error is the average of the difference between the Original Values and the Predicted Values. It gives us the measure of how far the predictions were from the actual output.

**Ontological Statement:**  Mean Squared Error(MSE) takes the average of the square of the difference between the original values and the predicted values.

**Performance Indicator: NDCG**

Frequency: Monthly

**Ontological Statement**: Normalized discounted cumulative gain (DCG) is a measure of ranking quality. In information retrieval, DCG measures the usefulness, or gain, of a document based on its position in the result list.

**Performance Indicator: Regression Analysis**

Frequency: Monthly

**Root Mean Square Error (RMSE)**

**Ontological Statement**: Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are.

**Performance Indicator: Quantiles of Errors**

Frequency: Monthly

Quantiles (or percentiles), which is the element of a set that is larger than half of the set, and smaller than the other half.

**Performance Indicator: "Almost correct" predictions**

Frequency: Monthly

##### Objective: Trustworthiness

Advance use of AI change management, knowledge representation performance evaluation, algorithmic impact assessment and context aware safeguards for a reliable, safe and transparent outcome

#### Goal: KR Objects

**Evaluate KR Object Performance**

KR Object oversight mechanisms will define how performance measurements are used via human-in-the-loop, human-on-the-loop, and human-in-command approaches