AI KR Strategists

This plan defines the roles AI KR Strategists.

Source

https://www.stratnavapp.com/StratML/Part1/861566c8-e9be-4642-b52f-f673fa499f4e/Styled

Start Date: 2020-04-01 **End Date:**   **Publication Date:** 2022-05-05

Submitter:

Carl Mattocks

✉
CarlMattocks@WellnessIntelligence.Institute

Organisation:

Name: Artificial Intelligence Knowledge Representation Community Group

Acronym: AIKR CG

Stakeholders:

👤 Carl Mattocks

Role: CoChair

Vision

Work performed and works created for each AI value proposition is clearly and transparently documented and measured.

Mission

To be responsible and accountable for the selection, development, application and management of Knowledge Representation (KR) for Artificial Intelligence (AI).

Goal: Strategic Plan

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

Other Information: 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)
 * 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)

Stakeholders:

👥Human-in-the-Loop Controllers

Human-in-the-Loop Controllers are humans that Train/Test AI systems.
They control the inputs of the humans when humans are in the loop.
As a simple example, they tag the initial images that are fed into the algorithms; later they come back and refine the patterns identified; ultimately they may also come back and test the outcome.
See:
 * https://en.wikipedia.org/wiki/Human-in-the-loop ;
 * https://humansintheloop.org/model-training/

Role: Model, test, evaluate and implement ethical approaches for Supervised Machine Learning of curated (labeled) data sets and the Active Learning training of algorithms via adjustment of parameters

Objectives:

Objective: **Ontology**

Employ ontology content that removes ambiguity, supports performance measurement and enables buy in

Other Information: 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. That is, the ontology, Representational Adequacy is able represent all the required knowledge;Inferential Adequacy is able to manipulate the knowledge represented to produce new knowledge (inferred from the original);Acquisitional Efficiency is able to acquire new knowledge using Human In-The-Loop and /or Human Out-of-The-Loop methods

👥AI KR Strategists

👥AI Strategists

Objective: **Algorithms**

Understand the various different types of algorithms and where they can support business strategy

Other Information: Algorithms' capabilities and limitations should be explained in a manner that is adapted to stakeholder concerns and identifies how accuracy, robustness, computational cost and stability will be measured

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

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

Goal: Risks

Identify and mitigate risks and known threats

Other Information: A guiding principle is that AIKR systems must mitigate risks.

Stakeholders:

🏢 DARPA

Role: This goal arose in reference to the DARPA initiative.

Objectives:

Objective: Consequences

Identify and minimize adverse and/or unintended consequences

Other Information: "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: Data

Ensure data quality and integrity

Other Information: Data quality: the data is fit for its intended purpose/use. Is supported by a systematic method for driving agreement on the definitions of categories.
Data integrity: is the maintenance and assurance of the accuracy and consistency of data over its entire life-cycle. Is supported by a monitoring system that compares actual outcome with predicted accuracy

Objective: **Bias**

Identify and reduce bias in AI KR objects

Other Information: 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: **Security**

Guard against illegitimate access whilst ensuring legitimate access

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

Objective: **Control**

Design the criteria to control the use (and misuse) of algorithms and data

Other Information: 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.

👥AI KR Strategists

👥AI Developers

👥Users of AI systems

Objective: **Intellectual Property**

Manage Intellectual Property rights over AI KR works

Other Information: What works can be protected, and what form of protection can be used for them?

Objective: **Privacy**

Protect the rights of the individuals/corporations whose data is processed

Other Information: Ensuring data is processed with the permission of the people to whom it pertains.
E.g. GDPR, intellectual property, etc.

Objective: Governance

Design governance in line with the risk tolerance

Other Information: Data and algorithm governance.

Goal: Compliance

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

Other Information: Compliance policies and procedures ensure that a planned change to a KR Object usage will comply with applicable laws/regulations during the identification, development, documentation, testing, validation, implementation, modification, use and retirement lifecycle

Goal: Ethics

Ensure AI Systems adhere to principles of ethics

Objectives:

Objective: Autonomy

Find the balance between human control/oversight and machine autonomy

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

Objective: Veracity

Objective: Accountability

Objective: Confidentiality

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

Other Information: 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.

Stakeholders:

▥ Artificial Intelligence Knowledge Representation Community Group (AIKR CG)

Role: Community of Interest

Objectives:

Objective: **Trustworthiness**

Advance use of AI safeguards

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

Objective: **Classification**

Track Classification Performance Indicators

Other Information: 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).

Goal: **KR Objects**

Evaluate KR Object Performance

Other Information: 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