2023-09

Designing Education Programs based on Competencies using Advanced Analytical Methods

Alves, Miriam C. Bergue

Monterey, California. Naval Postgraduate School

https://hdl.handle.net/10945/72450

NPS-VPAA-23-001

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DESIGNING EDUCATION PROGRAMS BASED ON COMPETENCIES USING ADVANCED ANALYTICAL METHODS

by

Miriam C. Bergue Alves, Ph.D.

September 2023

Approved for public release. Distribution is unlimited.

Prepared for: OPNAV N711, Navy Education Strategy and Policy Branch. This research is supported by funding from the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098). NRP Project NPS-23-N186-A

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	3. DATES COVERED END DATE
09/22/2022	Technical Report	10/23/2022	10/21/2023

4. TITLE AND SUBTITLE

Designing Education Programs based on Competencies using Advanced Analytical Methods

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
		0605853N/2098
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
NRP-23-N186-A		

6. AUTHOR(S)

Bergue Alves, Miriam C.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Naval Postgraduate School, Vice Provost for Academic Affairs, Monterey, CA 93943-5000	NPS-VPAA-23-001

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Naval Postgraduate School, Naval Research Program; OPNAV N711, Navy Education Strategy and Policy Branch	NRP; OPNAV N711	NRP-23-N186-A

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views expressed in this document are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT

We propose an innovative framework leveraging artificial intelligence (AI) for the creation and assessment of outcome-based education (OBE) programs, particularly those with interdisciplinary and multidisciplinary aspects, emphasizing the importance of students' learning outcomes (LOs) as it interrelates with competencies. The primary aim is to empower educational institutions to swiftly adapt to competency demands. The framework enables timely strategic adjustments within educational programs, aligning them with the dynamic higher education landscape driven by emerging technologies and the need for upskilling. It comprises two core components: the Structured Data Model and the AI-Assisted component. The Structured Data Model systematically organizes educational program elements, creating a database that supports advanced queries, facilitating the identification and incorporation of changes within programs. The AI-Assisted component uses natural language processing (NLP) techniques to classify competencies within existing educational offering with measurable accuracy. We defined the framework's strategic objectives considering four different, although interrelated, perspectives: Data, AI-Model, Classification, and Recommendation, which will serve as a reference for future implementation of this framework as an operative system. We also conducted a comprehensive case study using the NPS educational model, applying the framework to assess its value and effectiveness. Four AI-based classifiers were examined, set to classify intended LOs into existing NPS curricula. The classification results were promising, with one of the AI models reaching approximately 70% accuracy on test dataset predictions, demonstrating the feasibility and potential benefits of this type of AI application for DOD education and training institutions. Our intention is to offer a systematic and reliable process for addressing new competency needs while adapting to evolving education and institutional requirements, combining human decision-making with AI-driven method.

15. SUBJECT TERMS

academic programs, AI, artificial intelligence, assessment, competency, outcome-based education, continuum of learning, Core Skill Requirement, (CSR) curricula, education skill requirement (ESR), education alignment, decision-making, interdisciplinary, multidisciplinary, Naval Education Enterprise, P-code, science of teaching and learning, warfighting advantage

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
U	U	U	UU	64

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Dr. Miriam C. Bergue Alves	831-656-7547

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL

Monterey, California 93943-5000

Ann E. Rondeau
President

Scott Gartner
Provost

The report entitled "Designing Education Programs based on Competencies using Advanced Analytical Methods" was prepared for OPNAV N711, Navy Education Strategy and Policy Branch and funded by Naval Postgraduate School, Naval Research Program (PE 0605853N/2098).

Further distribution of all or part of this report is authorized.

This report was prepared by:

BERGUE ALVES.MIRIAM.CELIA.1558057919
Digitally signed by BERGUE ALVES.MIRIAM.CELIA.1558057919
Date: 2023.10.13 09:25:56 -07'00'

Miriam C. Bergue Alves, Ph.D.
Faculty Associate - Research

Reviewed by:

LESTER.DENNIS.LEE.1156800852
Digitally signed by LESTER.DENNIS.LEE.1156800852
Date: 2023.10.13 09:43:02 -07'00'

Dennis Lester, Ph.D.
Associate Provost

Reviewed by:

HOOPER.JOSEPH.P.1298307274
Digitally signed by HOOPER.JOSEPH.P.1298307274
Date: 2023.10.13 15:33:58 -07'00'

Joseph Hooper, Ph.D.
Vice Provost, Academic Affairs

Released by:

SMITH.KEVIN.B.1230026366
Digitally signed by SMITH.KEVIN.B.1230026366
Date: 2023.10.16 10:01:39 -07'00'

Kevin B. Smith, Ph.D.
Vice Provost for Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

We propose an innovative framework leveraging artificial intelligence (AI) for the creation and assessment of outcome-based education (OBE) programs, particularly those with interdisciplinary and multidisciplinary aspects, emphasizing the importance of students' learning outcomes (LOs) as they interrelates with competencies. The primary aim is to empower educational institutions to swiftly adapt to competency demands. The framework enables timely strategic adjustments within educational programs, aligning them with the dynamic higher education landscape driven by emerging technologies and the need for upskilling. It comprises two core components: the Structured Data Model (SDM) and the AI-Assisted component. The SDM systematically organizes educational program elements, creating a database that supports advanced queries, facilitating the identification and incorporation of changes within programs. The AI-Assisted component uses natural language processing (NLP) techniques to classify competencies within an existing educational offering with measurable accuracy. We defined the framework's strategic objectives considering four different, although interrelated, perspectives: Data, AI-Model, Classification, and Recommendation, which will serve as a reference for future implementation of this framework as an operative system. We also conducted a comprehensive case study using the NPS educational model, applying the framework to assess its value and effectiveness. Four AI-based classifiers were examined, set to classify intended LOs into existing NPS curricula. The classification results were promising, with one of the AI models reaching approximately 70% accuracy on test dataset predictions, demonstrating the feasibility and potential benefits of this type of AI application for DOD education and training institutions. Our intention is to offer a systematic and reliable process for addressing new competency needs while adapting to evolving education and institutional requirements, combining human decision-making with AI-driven methods.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

Evolving warfighters' needs to upskill and employ new capabilities in highly contested environments have created challenges in aligning higher education institutions' programs and curricula to meet these needs and the expectations of military students and other stakeholders. Accordingly, education institutions must responsively address the gaps between their program learning outcomes (LOs) and the actual competencies required to maintain warfighters' readiness. The Department of the Navy (DON) recognized the need for realignment of their education programs and curricula as stated in the Chief of Naval Operations (CNO) NAVPLAN 2022 [1]– "*The Navy's education enterprise must align its curriculum and research to deliver warfighting advantage. Students and faculty research will focus on warfighting concepts and capabilities our fleet needs to compete and win*". The Commandant of the Marine Corps Force Design 2030 also reinforces the fact that educating and training Marines and sailors to use high-end warfighting by providing ready, relevant education is imperative to prepare them to joint and combined naval operations [2, 3].The Naval Education Strategy 2023 [4] also reinforces the need to invest in the right competencies as a mean for force readiness and competitive advantage, which requires that the Naval Education Enterprise (NEE) continuously creates and updates its education programs.

Traditionally programs and curriculum's development and reviews are long processes, with frequencies not necessary matching the rapid changes either suggested by evaluation feedback or emerging disciplines [5]. At N Naval Postgraduate School (NPS), each curriculum leading to an academic degree is formally reviewed every two years by the curriculum sponsors and related department faculty and staff. Innovative approaches are necessary to respond to the evolving nature of current education and to help sponsors, faculty and staff analyze the potential redesign alternatives, dependencies, and benefits of modifying or introducing new inter and multidisciplinary programs and curricula, while complying with education skill requitements (ESRs) and institutional accreditation standards.

Frequently, research studies address the potential use of intelligent augmentation (IA) [6] to blend the results of artificial intelligence (AI) algorithms to provide strategic insights and support for decision-making process in several areas [7], including project management [8]. Examples of using AI methods in support to decision making process are frequently related to healthcare, retail, manufacturing, energy, and financial services. In education settings, research studies have been focused on the traditional benefits that AI analytical methods can bring to smart learning, tutoring systems [9, 10], social robots and other intelligent technologies, such as virtual facilitator and learning analytics [11]. Other studies have synthetized areas of AI education applications as profiling and prediction, assessment, evaluation, adaptive systems, and personalization. Science teachers' professional development also shows a prospective use of AI methods support [12]. Examples of the use of AI methods to support curriculum design [13] and continuous curriculum engineering [14] are less frequent in the literature and are usually focused on specific case scenarios, not addressing the more comprehensive focus of this research.

This research proposes a framework that leverages artificial intelligence (AI) to assist in both the creation and assessment of outcome-based education (OBE) and competency-based education (CBE) programs with interdisciplinary and multidisciplinary facets. The primary objective is to enable educational institutions to align their offerings with contemporary and emerging competency demands. By implementing a functional system based on the proposed framework, the intention is to facilitate well-timed strategic adjustments within educational programs, curricula, and courses demanded by new competencies needs while addressing the dynamic landscape of higher education.

B. MEASURABLE, MEANINGFUL, AND INTEGRATED COMPETENCIES

In light of preparing officers capable of addressing complex scenarios [3, 16], the paradigm of outcome-based military education (OBME) has surfaced as a strategic response to effectively address this challenge. Evidencing this shift, the OBME manual from the U.S. Joint Chiefs of Staff [17] underscores the transition of joint professional military education (JPME) from an input-driven to an output-centric educational model. This transformation emphasizes outcomes and underlines the attainment of targeted learning objectives by the students. This also requires that advances in training, education and supporting technology enable the analysis, design and measurement of competencies.

Department of Defense (DOD) defines competency as "*an observable measurable pattern of knowledge, abilities, skills, and other characteristics that individuals need to successfully perform their work*." [18]. Competencies are the necessary skills and knowledge a learner needs to acquire, closely linked to the education and training program's learning outcomes [19]. The DOD Advanced Distributed Learning (ADL) [20] frames competency [21] as this: "*Competencies are represented as networks of knowledge, skills, abilities, or other behaviors (KSAOs) that collectively define a set of related performance elements, along with any collected evidence that the learner has achieved mastery of these elements at some gradated level of performance*"

Competencies are meant to be assessed and are dependable of a specific context. In the context of learning, competencies are often associated with learning objectives or outcomes.

According to [22], competency frameworks can be represented as a network of knowledge, skills, abilities, and other characteristics (KSAOs), where the KSAOs are grouped and characterized by different relationships. These competencies may be associated with different LOs resulting from different educational or training activities. Examples of competency framework definition and management systems include CaSS [23], CASE [24], ASN [25] and O*Net [26].

In an OBE, competencies are associated to a prescribed curriculum, and competency frameworks are important sources for curriculum development or review, as well as courses sequence definition. When applied to curriculum development, top-level competencies will not necessary match one-to-one with LOs since in many instances, a competency may be composed of several LOs.

Another two important components of OBE programs are data collection and analysis, ensuring that such programs incorporate improvements, monitor gaps and keep the high quality required by accreditation bodies and curriculum sponsors.

C. AI ASSISTED CURRICULAR REVIEW

Military education institutions must responsively address the gaps between their LOs and the required competencies to keep the warfighters' readiness. Traditionally, development and reviews of educational programs and curricula are long processes, with frequencies not necessary matching the rapid changes either suggested by evaluation feedback or emerging disciplines [27]. At NPS, for instance, curricula leading to an academic degree is reviewed every two years [28, 29, 30] by its respective curriculum sponsors and associated department faculty, as established in [31].

The Department of Defense Instruction (DODI) 1322.10 [30] defines the policy on graduate education for military officers. According to this policy, graduate education programs should be periodically audited and assessed by their respective Military Department regarding the "*disciplines that fulfill a present need, anticipated requirement, or future capability and that contribute to the effectiveness of the Military Departments and the Department of Defense*." This policy recommendation, in conjunction with the evolving military operational and functional environment, demands a more responsive education system that supports upskilling needs of the force.

Figure 1 illustrates the traditional versus our recommended AI-assisted curriculum design and review. Based on this new approach, and on the strength of AI prescriptive analysis, the intention of the proposed framework is to accelerate the design of a new program or curriculum, as well as the review process, to effectively support changes, audits and assessments carried out by different stakeholders. The AI-assisted approach considers an integrated and common data model to base the review, in contrast to a siloed review process.

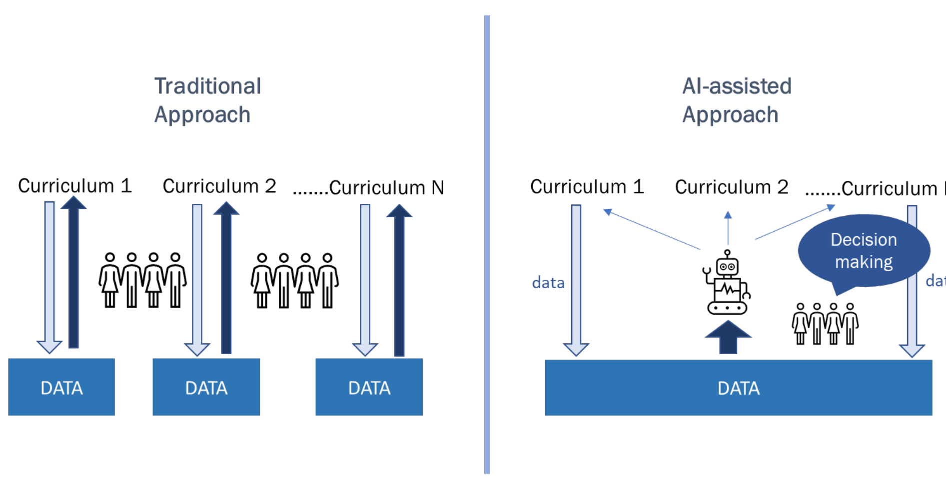

Figure 1. Traditional vs. AI-assisted approach for education program review.

D. COMPETENCIES AND LOs ALIGNMENT

In an educational setting, the emergence of new competencies highlights the need to examine the relation between these competencies and the existing curricula and courses LOs. By understanding this relation, educators can identify any gaps or misalignments that may exist. They can also assess whether the current curriculum adequately addresses the required KSAOs associated with the new competencies. If there are discrepancies, adjustments can be made to the curriculum to better prepare students for the evolving demands of the operational environment.

Discovering these alignments in time allows educational programs to proactively address the evolving needs of the Navy operational scenarios and ensures that graduates are equipped with the necessary KSAOs to succeed in their chosen fields. Ultimately, the discovery and understanding of the alignment (or lack of) between new competencies and existing LOs is vital for educational programs to remain relevant and responsive to sponsor needs.

According to the Credential Transparency Definition language [32] "*An alignment is an assertion of some degree of equivalency between the subject and the object of the assertion.*" The language defines nine different types of alignment between two competencies or competency frameworks.

We propose the adoption of a subset of these definitions as a reference to examine the alignment of a required competency with existing LOs' descriptions. Based on the accuracy results using optimal AI approaches, we expect to identify the competency as belonging to one of following five types of alignment:

a. **Broad** alignment**:** LOs covers all the relevant concepts in the required competency as well as relevant concepts not found in the competency's description.

b. **Exact** alignment: LOs and the required competency are coextensive.

c. **Major** alignment: Major overlap of relevant concepts between the LOs and the required competency.

d. **Minor** alignment: Minor overlap of relevant concepts between LOs and the required competency.

e. **Narrow** alignment: The required competency covers all the relevant concepts in the related LOs as well as relevant concepts not found in the current educational programs.

E. RESEARCH OBJECTIVES

Considering the vision established in the Naval Education Strategy 2023 [15] and its three lines of efforts, our research addressed the following questions:

1. How can NEE strategically adapt its programs/curricula to the evolving warfighter educational requirements leveraging new technology/upskilling needs and required competencies?

2. How can gaps between education programs, LOs and required competencies be identified in inter, multidisciplinary programs?

3. How can NEE explore and assess the benefits, drawbacks, impacts, costs, risks, and effectiveness of different educational solutions?

The remaining of this report is structured as follows: Section II summarizes the AI concepts and methods involved in the proposed framework conceptualization. The framework is presented in Section III. To validate the proposed framework, Section IV dives in a case study, based on NPS's education structure, including data understanding and preparation, as well as experimentation with AI models. Finally, Section VI summarizes our main findings, conclusions and future research.

II. AI KEY CONCEPTS AND METHODS

A. AI MAIN CATEGORIES

AI can be divided into three main categories [33]:

1. **Narrow AI**: Also called weak AI or artificial narrow intelligence, which include almost all the current applications. Some examples include image classification, object detection, speech recognition, translation, and Natural language processing (NLP).
2. General AI: Also called strong AI or artificial general intelligence, which would be able to learn, think, invent and solve complicated problems.
3. **Super AI**: Also called superintelligence, and it is referred as AI after the singularity point[1].

Narrow AI is what researchers and developers have been achieved so far and general AI is what is expected in the future. The next subsections offer AI central ideas and techniques, providing foundational knowledge to support the understanding of the concepts underpinning this study..

B. MACHINE LEARNING

In the realm of machine learning (ML), success hinges on a methodical process that revolves around identifying patterns and constructing models through systematic steps. The overarching goal is to teach algorithms how to learn a target function, which essentially maps input variables to corresponding output variables (mapping function). As the field is diverse and dynamic, a selection of distinct approaches is needed to perform experiments, each aiming to ascertain the optimal strategy for a given context [34]. In order to navigate this multifaceted landscape, three questions should be asked [35]:

1. Does a ML approach align with the requirements of the project (fitness)?
2. Is the chosen approach suitable for the desired predictions and the descriptive features utilized (compatibility)?
3. What results should be considered reliable and valid (accuracy)?

The systematic nature of the ML process involves pattern identification, model creation, and training, all underscored by rigorous experimentation for the most fitting approach.

When considering the suitability of a ML model, besides the accuracy, specific criteria also come into play. These encompass factors like prediction speed, especially if real-time responses are essential for the context. Computational loading times could potentially influence the model selection, though in our specific case, high prediction speed isn't imperative.

---

[1] According to Wikipedia, "In mathematics, a **singularity** is a point at which a given mathematical object is not defined, or a point where the mathematical object ceases to be well-behaved in some particular way, such as by lacking differentiability or analyticity.[1][2][3]" (https://en.wikipedia.org/wiki/Singularity_(mathematics))

Additionally, the model's retraining capability is paramount, ensuring adaptability in the face of obsolescence, updates, or deviations from expectations. This involves assessing the ease of modifying the existing model or transitioning to a new one. Moreover, interpretability of the results should be comprehensible and justifiable. While certain models like decision trees [35] offer greater interpretability compared to complex counterparts such as deep neural networks [36], this set of criteria remain integral to our model selection process.

1. ML for Predictive Analysis

ML's effectiveness thrives when handling extensive datasets with multiple features, consequently it becomes imperative to explore diverse models suitable for our specific context. When evaluating model reliability, a crucial factor lies in its capacity to establish mappings across the entire spectrum of descriptive values and target feature predictions. In our case, this translates to the question: Given a new description of a required competency in natural language (descriptive values), what education offering (curriculum or set of courses) would address the knowledge and skills needed to fulfill it (target feature)?

Inductive learning refers to the process in which a model extracts a general rule from a finite set of examples. This is why ML is often referred to as inductive learning. The set of assumptions employed by a ML algorithm that influences its selection of a single model is known as the algorithm's inductive bias [35]. This set of assumptions encompasses the criteria for model selection, the search space explored by the algorithm, and the search process employed.

It is important to recognize that the use of inductive bias is necessary for learning to take place. Without inductive bias, a ML algorithm would not be able to make inferences beyond what is explicitly present in the data. Additionally, it is crucial to minimize sampling bias, which occurs when the data sample used in a data-driven process is not representative of the population it aims to represent. While inductive bias is necessary for model selection in ML, it is important to note that all other choices we make, such as the selection of data, descriptive features, and the deployment of a model, introduce their own biases to the overall process. It is crucial to be highly aware of these biases and their potential impact on the outcomes.

2. Performance Measures

The F1 score is an evaluation metric for measuring the performance of a classification model in ML. The F1 score is a very common metric used to evaluate the model's accuracy and it is especially useful when working with imbalanced datasets, with varying number of entries for each of the labels in the datasets.

In imbalanced datasets, we need to consider other metrics besides accuracy, as a model can achieve high accuracy by solely predicting the majority class. Important performance measures are described as follows.

Accuracy is the ratio of correct predictions to the total number of predictions made by the model, It is calculated by the total sum of true positive (TP) and true negative (TN) predictions divided by the sum of all predictions (TP, TN, false positives (FP), and false negatives (FN). In simpler terms, accuracy provides an understanding of how often the model's predictions are correct in relation to all predictions made, considering both positive and negative outcomes.

1. **Accuracy**: Correct predictions made by the model, calculated as (TP + TN) / (TP + TN + FP + FN).
2. **Precision**: Accuracy of positive predictions, calculated as TP / (TP + FP).
3. **Recall (Sensitivity or True Positive Rate)**: Sensitivity in predicting positive instances, calculated as TP / (TP + FN).
4. **Specificity (True Negative Rate)**: Sensitivity in predicting negative instances, given by TN / (TN + FP).
5. **F1-Score**: Calculated as 2 * (**Precision** * **Recall)** / (**Precision** + **Recall)**.
6. **False Positive Rate (FPR)**: Ratio of false positive predictions to all actual negative instances, calculated as FP / (FP + TN).

Precision measures the accuracy of the positive predictions made by the model. It is calculated as the ratio of TP predictions to the sum of TP and FP predictions. Precision indicates how well the model identifies relevant instances among the predicted positive instances.

Recall, also known as sensitivity or true positive rate, measures the model's ability to identify all positive instances correctly. It is calculated as the ratio of TP predictions to the sum of TP and FN predictions. Recall indicates how well the model captures all relevant instances.

The F1 score is the harmonic mean of precision and recall. It provides a balanced measure that considers both precision and recall. The F1 score ranges between 0 and 1, where 1 represents the best possible value, indicating perfect precision and recall. A higher F1 score implies better overall performance of the classification model.

Confusion matrix-based performance measures [33, 35] provide a comprehensive evaluation of the model's predictive accuracy by breaking down the outcomes of the model's predictions into different categories. The confusion matrix is a tabular representation that presents the TP, TN, FP, and FN results generated by the model's predictions. The TN indicates the model's capability to correctly identify negative instances.

3. Underfitting and overfitting

When selecting a prediction model, there are two types of errors to be cautious of: underfitting and overfitting. Underfitting transpires when the chosen model is overly simplistic and fails to capture the underlying relationships between the descriptive features and the target feature within the dataset. On the other hand, overfitting emerges when the selected model is excessively complex, leading it to closely conform to the dataset and become overly sensitive to noise present in the data.

4. Supervised, Unsupervised and Reinforcement Learning

Supervised learning [33, 34] is one of the paradigms in ML where the algorithm learns to map input data to corresponding output labels based on a labeled dataset. In other words, it involves training a model on a dataset where both the input and the desired output (label) are provided. The goal is for the model to learn the underlying patterns and relationships between the input features and the output labels so that it can accurately predict the labels for new, unseen data.

Supervised learning problems can be broadly categorized into two types: classification (predict a categorical label or class from a set of predefined classes) and regression (predict a continuous numerical value).

Typical classification techniques include:

a. **Logistic Regression:** A simple linear model used for binary classification problems, such as spam detection, sentiment analysis, and medical diagnosis (disease present or not) [37].

b. **Support Vector Machines (SVM):** A powerful algorithm that separates data points with a hyperplane and is used for both binary and multi-class classification tasks, such as image classification, text categorization, and handwriting recognition [38].

c. **Random Forest:** An ensemble method that combines multiple decision trees to improve accuracy. Used for tasks like image recognition, credit risk assessment, and species classification [39].

d. **K-Nearest Neighbors (KNN):** A method that predicts the label of a data point based on the labels of its nearest neighbors. It's used for recommendation systems, image classification, and anomaly detection [40].

e. **Naive Bayes:** A probabilistic classifier based on Bayes' theorem that's particularly useful for text classification, spam filtering, and document categorization [41].

Regression techniques include:

a. **Linear Regression:** A basic technique for predicting a continuous numerical value, such as predicting house prices, stock prices, or temperature [42].

b. **Decision Trees:** Used for both regression and classification, decision trees are tree-like structures that partition data into subsets based on feature values. They're used in predicting house prices, predicting sales, and more [37].

c. **Gradient Boosting:** An ensemble technique that combines multiple weak learners (often decision trees) to make more accurate predictions. It's used in tasks like predicting customer churn, demand forecasting, and financial modeling [43].

d. **Neural Networks:** Deep learning models consisting of interconnected nodes or neurons. They can handle complex data and are used in various regression tasks like predicting sales, disease progression, and energy consumption [36].

e. **Support Vector Regression (SVR):** An extension of SVM for regression problems, used when the relationship between input and output variables is not linear, such as stock price prediction and medical diagnosis [44].

Unsupervised learning [36, 42] and reinforcement learning are the other two fundamental paradigms in the field of ML. Unsupervised learning involves training models on data without explicit supervision, aiming to discover hidden patterns or structures within the dataset. Unlike supervised learning, where the algorithm is provided with labeled examples, unsupervised learning works with unlabeled data, relying on techniques like clustering and dimensionality reduction. Clustering algorithms group similar data points together, aiding in data exploration and segmentation. Dimensionality reduction techniques, on the other hand, help streamline the dataset by representing it in a lower-dimensional space, preserving crucial information while eliminating noise and redundancy.

Reinforcement learning [45], on the other hand, involves an agent learning how to make sequential decisions to maximize a reward signal in a dynamic environment. Inspired by behavioral psychology, reinforcement learning employs a trial-and-error approach, where the agent takes actions to interact with the environment and receives feedback in the form of rewards or penalties. Over time, the agent learns optimal strategies to achieve long-term objectives. Key components of reinforcement learning include the agent's policy (strategy for selecting actions), the environment's dynamics, and the reward function that guides the agent's learning process. This paradigm is particularly relevant in scenarios where explicit training data is scarce, and the agent must learn by interacting with the environment, as seen in applications like robotics, game playing, and autonomous systems.

C. NLP AND TEXT CLASSSIFICATION

NLP is an interdisciplinary domain including computer science, artificial intelligence, and linguistics and focuses on the development of computer systems able to analyze and comprehend human language. Text classification plays a crucial role in NLP and finds numerous applications across various domains [46], such as sentiment analysis, spam detection, topic categorization, and more. Its significance lies in its ability to automatically categorize and organize vast amounts of textual data by assigning predefined categories or labels to them based on their content. The key concepts and tasks involved in NLP are described in the next subsections.

1. Tokenization

Tokenization is the process of breaking down text into individual words or tokens, which can be words, subwords, or characters. This transformation enables the model to work with discrete elements rather than raw text and is essential for further analysis, as it forms the basis for understanding the structure of the text [47].

Tokenization is a fundamental task in NLP, and it can be achieved using various techniques such as word tokenization, character tokenization, and subword tokenization, the latter being beneficial for handling out-of-vocabulary words or reducing the size of the vocabulary.

2. Feature Extraction

Text data needs to be transformed into numerical features that ML algorithms can understand. Techniques like Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) are common methods for representing text as numerical vectors [48]. Figure 2 illustrates this process.

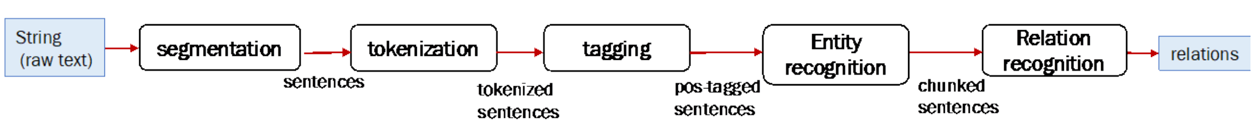

Figure 2. NLP of LO corpus: feature extractor.

3. Feature Selection and Engineering

Choosing the right features and engineering new ones can significantly impact the performance of a text classifier. Techniques like n-grams, word embeddings (e.g., Word2Vec [50], GloVe [51]), and contextual embeddings, like Bidirectional Encoder Representations from Transformers (BERT) [52] and Generative Pretrained Transformer (GPT) [53] help capture semantic meaning.

4. ML Algorithms

Various ML algorithms are used for text classification, including Naive Bayes [41], SVMs [38], and more recently, deep learning methods like Convolutional Neural Networks (CNN) [36] and Transformers [47].

5. Training Data

The construction of an effective text classifier involves a substantial dataset, containing text samples that are properly labeled with their corresponding categories or classes. This dataset serves as the foundation for training the ML model to learn the patterns, relationships, and distinctions present within the text data. Through this training process, the model becomes capable of making predictions and classifications on new, unseen text data [49].

Following model training, key performance metrics are used to assess the model's efficacy. Guarding against overfitting, which can occur due to noise in training data, is achieved via methods like cross-validation and regularization [36]. Addressing imbalanced class scenarios involves techniques like oversampling and under sampling. In addition, pretrained language models (e.g., BERT, GPT) exhibit remarkable text classification performance by learning contextual nuances from vast text data, with fine-tuning yielding impressive outcomes using less training data. Compared to BERT, DistilBERT [54] is significantly smaller and more efficient, enabling to train a classifier faster.

D. CORPUS DATA

In a classification task, a "data corpus" refers to a collection of data, often text-based, that has been assembled and organized for the purpose of training, validating and evaluating machine learning models and classifiers. After evaluating the machine learning models, the aim is to categorize new data into one or more predefined classes or categories.

The Natural Language Toolkit (NLTK) [55] provides algorithms for text classification and it was used in the NPS Case, when coding the AI models.

E. DEEP LEARNING

Deep learning [36] is a subfield of ML that focuses on training artificial neural networks to learn and make intelligent decisions from data. It's inspired by the structure and functioning of the human brain's neural networks. Deep learning algorithms learn to perform tasks by analyzing and learning patterns in vast amounts of data. Unlike traditional ML, where feature engineering is crucial, deep learning aims to automatically learn relevant features from the data itself.

Deep learning has been particularly successful in tasks that involve complex patterns and large datasets, such as image and speech recognition, NLP, and playing strategic games. There are a couple of different types of networks related to deep learning, as follows [36]:

a. **Feedforward Neural Networks (FNNs):** Also known as multilayer perceptrons (MLPs), generate predictions or classifications.

b. **Convolutional Neural Networks (CNNs):** CNNs are designed for image and video analysis, image classification and object detection.

c. **Recurrent Neural Networks (RNNs):** RNNs are specialized for sequential data, like time series or natural language.

d. **Generative Adversarial Networks (GANs):** GANs consist of two networks—a generator and a discriminator—competing in a game. GANs generate realistic data, making them potent for image generation, style transfer, and more.

e. **Autoencoders:** Autoencoders are used for unsupervised learning and data compression.

f. **Transformer-based Models:** Transformers, like BERT and GPT use attention mechanisms to process input data in parallel and capture long-range dependencies. BERT excels in understanding context, while GPT generates coherent text.

F. THE TRANSFORMER ARCHITECTURE

The Transformer architecture, introduced by Vaswani et al. [56] has revolutionized NLP and various other tasks. It relies heavily on the concept of attention mechanisms, allowing it to capture relationships between words regardless of their position in a sequence. There are two widely recognized transformers today: GPT and BERT. Both of them reached substantial improvements by merging the Transformer architecture with unsupervised learning and can be used to train NLP models in different NLP contexts.

The main components of the Transformer architecture are summarized below:

a. **Self-Attention Mechanism:** It allows the model to weigh the importance of different words in a sequence when predicting a specific word. This attention mechanism is computed in parallel for all words in the sequence and captures contextual relationships effectively.

b. **Multi-Head Attention:** It captures different types of information and relationships by employing multiple self-attention mechanisms in parallel.

c. **Positional Encoding:** It provides information about the position of each word in the sequence, ensuring the model can differentiate between words solely based on their position.

d. **Encoder and Decoder Stacks:** The Transformer consists of an encoder stack and a decoder stack. The encoder processes input data, while the decoder generates output sequences.

e. **Position-wise Feedforward Networks:** This network involves fully connected layers and non-linear activation functions, allowing the model to learn complex transformations.

f. **Residual Connections and Layer Normalization:** Each sub-layer in a Transformer layer uses residual connections, helping to mitigate the vanishing gradient problem and accelerating convergence. Layer normalization is also applied to stabilize training.

g. **Encoder-Decoder Attention:** It enables the decoder to focus on relevant parts of the input sequence when generating an output.

1. Benefits and Challenges

Transformers have brought about a significant advancement in NLP and classification tasks. Here are some benefits and challenges associated with the use of transformers in NLP and classification:

Benefits:

a. **State-of-the-Art Performance:** Transformers, particularly large models like BERT, GPT-3, and their variants, have achieved state-of-the-art performance across a wide range of NLP tasks.

b. **Contextual Understanding:** Transformers capture contextual relationships between words by considering the entire sequence of words in a sentence, helping in understanding the meaning of words based on their context.

c. **Pre-trained Representations:** Transformers have the capability to undergo pre-training on extensive textual datasets to acquire broad language understanding with posterior fine-tuning to specific contexts. This approach has proved to be effective in terms of training time and use of computational resources.

d. **Transfer Learning:** Pre-trained transformers allow for transfer learning. Models pre-trained on a large corpus can be fine-tuned for specific tasks, even when you have limited labeled data for that task.

e. **Attention Mechanism:** Transformers use self-attention mechanisms that enable them to weigh the importance of different words in a sequence when making prediction, capturing long-range dependencies and understanding the relationships between distant words.

Challenges:

a. **Computational Resources:** Large transformer models require significant computational resources, including memory and processing power, for training and inference.

b. **Data Requirements:** While transfer learning helps in reducing the need for large, labeled datasets, fine-tuning still requires some amount of labeled data specific to the target task. Acquiring and annotating such data can be time-consuming and expensive.

c. **Model Size:** State-of-the-art transformer models can be extremely large, making them challenging to deploy in resource-constrained environments like mobile devices or edge devices.

d. **Opacity:** interpreting their decisions can be difficult, particularly for complex tasks. This can be problematic for scenarios where model decisions need to be explained or justified.

e. **Bias:** Transformer models can inherit biases present in the training data, leading to biased predictions. Ensuring fairness and mitigating biases in these models can be a challenge.

f. **Fine-tuning:** While transfer learning is powerful, fine-tuning on specific tasks can sometimes lead to catastrophic forgetting or instability if not performed carefully [57].

G. TRADEOFFS AND CONSIDERATIONS

We propose the following guidelines related to the AI techniques choice within the context of the proposed framework presented in the next section:

Model:

a. Looking for the right balance between overfitting and underfitting, effectively managing the trade-off between bias and variance.

b. Evaluating trade-off between model performance and interpretability/explainability, acknowledging the potential trade-off associated with using more complex "black box" models.

c. Weighing the benefits of complexity against simplicity, considering the parsimony of the model.

Data and features:

a. Ensuring that the dataset is representative and balanced.

b. Assessing the sufficiency of quantity, depth, and completeness of the data and features used, recognizing that inadequacies in these areas may impact model performance.

c. Addressing the challenges related to intractability, including difficulties with data, algorithms, and feature engineering, and finding appropriate solutions.

Performance:

Considering the trade-offs involved in minimizing errors, recognizing that reducing one type of error may result in an increase in another type of error.

Opacity:

As with other deep learning models, transformers are to a large extent opaque. It is hard or impossible to unravel “why” a model made a certain prediction. This is an especially hard challenge when these models are deployed to make critical decisions.

III. THE PROPOSED AI-ASSISTED FRAMEWORK

A. USE CASES

The initial step in establishing the requirements and components of the proposed AI-assisted framework was the delineation of utilization scenarios for stakeholders. This phase encompassed the discernment and comprehension of requisites and standpoints held by diverse individuals or collectives who would encounter, influence, or interface with an operational setup instantiating the proposed framework. The key stakeholders identified during this project phase are depicted in Figure 3. Each of these stakeholders possesses distinct vantage points concerning system utilization, these being demarcated via use case scenarios. However, it remains plausible that additional prospective stakeholders could come in place during subsequent phases of system design and implementation.

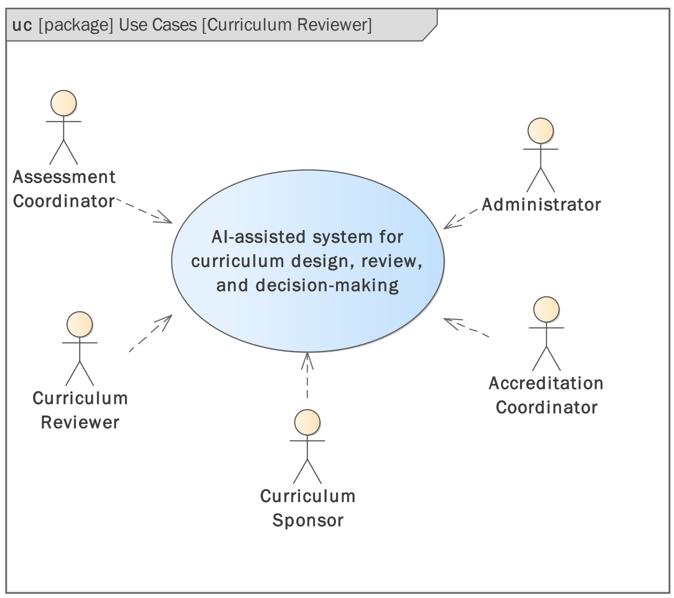

Figure 3. Main stakeholders identified for the initial framework proposal.

Consideration of stakeholders since the beginning is particularly important when an AI-assisted system provides recommendations or insights for decision-making. Stakeholders need to understand the reasoning behind AI's suggestions to trust and embrace its use. In addition, involving stakeholders in the use case definition process fosters a sense of ownership and buy-in. When individuals see that their needs and concerns have been considered, they are more likely to embrace the new technology and support its integration into their workflows.

The associated use cases for each identified stakeholder are described in the next subsections.

1. Curriculum Reviewers[2]

Use Case 1: Analyze similarities of LOs in existing curricula.

a. Query the system to identify commonalities and overlaps in LOs across different courses or programs.

b. Receive insights into areas where courses share similar educational goals, allowing for potential consolidation or cross-listing.

Use Case 2: Explore alignment of new competencies within existing curricula.

a. Input new competency requirements into the system.

b. Receive recommendations on how these new competencies can be integrated into existing curricula, identifying suitable courses and pathways.

c. Understand the impact of adding new competencies on the overall curriculum structure.

2. Curriculum Sponsors

Use Case: Assess curriculum adaptation.

a. Input proposed changes to curricula, such as new courses or updated LOs.

b. Receive feedback on how these changes align with existing competencies and LOs.

c. Gain insights into the feasibility and effectiveness of proposed curriculum modifications.

3. Accreditation Coordinator

Use Case: Ensure alignment with accreditation standards.

a. Input accreditation standards and criteria into the system.

b. Receive an analysis of how the curriculum meets these standards, highlighting areas of compliance and potential gaps.

c. Facilitate the accreditation review process by utilizing data-driven insights.

4. Assessment Coordinator

Use Case 1: Monitor classification accuracy.

a. Access reports on the accuracy of competency classification.

b. Identify trends in classification performance and take corrective actions if necessary.

Use Case 2: Evaluate LOs over time.

a. Query the system to track changes in LOs across different iterations of curricula.

---

[2] Also include Department Chairs and other interested faculty.

b. Analyze the evolution of educational goals and identify areas of improvement based on historical data.

5. Administrators (Registrar and Institutional Research)

Use Case: Monitor curriculum performance.

a. Access dashboards that display the alignment of competencies, LOs, and course offerings.

b. Make informed decisions about curriculum adjustments and resource allocation based on data-driven insights.

The use cases defined in this subsection guided the proposal of the general structure of the framework as described in the next subsection.

B. GENERAL STRUCTURE

1. Functional requirements

The functional requirements, which were derived from the use cases described in the previous subsection, are presented in Table 1 with the respective use cases that originated them. These requirements reflect the main system functionality and were used as the central guide to define the framework general structure.

Table 1. Functional requirements derived from Use Cases.

Requirement	Use case
1. Provide robust NLP and capabilities to process, analyze LOs and competencies, and optimize curricula and course offerings	Curriculum Reviewer
2. Provide AI methods for competencies classification suitable to the nature of the datasets[3].	Curriculum Reviewer and Curriculum Sponsor
3. Provide recommendation that suggest ways to integrate new competencies into existing curricula.	Curriculum Reviewer and Curriculum Sponsor
4. Allow users to input and identify changes to curricula.	Curriculum Reviewer and Curriculum Sponsor
5. Provide insights into the impact of proposed changes on existing LOs and competencies.	Curriculum Reviewer and Curriculum Sponsor
6. Demonstrate the degree to which curricula align with accreditation standards and intentions.	Accreditation Coordinator
7. Incorporate a mechanism for mapping curriculum elements to accreditation criteria and producing reports that show	Accreditation Coordinator

[3] Datasets are composed of education and/or training LOs. The selection of which AI models will be adopted has to be based on the AI model's performance measures.

compliance levels.	
8. Provide assessment mechanisms and historical data tracking features within the system[4] and support the integration of assessment at the institutional, curriculum, and course levels.	Assessment Coordinator
9. Use data visualization tools, dashboards, and analytics capabilities within the system for program/curriculum management, review, and improvement.	Administrator

2. Framework main components

The framework comprises of two primary components: the **SDMl** component and the **AI-Assisted** component. The Performance Analysis is carried out based on the results of the classification process. These components are depicted in Figure 4, along with the external **Education/Training Program Structure** entity. This external entity provides insights into the core program model. Minimally, it includes the education and/or training offerings, the expected LOs, specific requirements pertaining to the Institution's education model (for instance, educational skills requirements and core skills requirements), and their interrelationships. The Education/Training Program Structure determines the requirements for the SDM component.

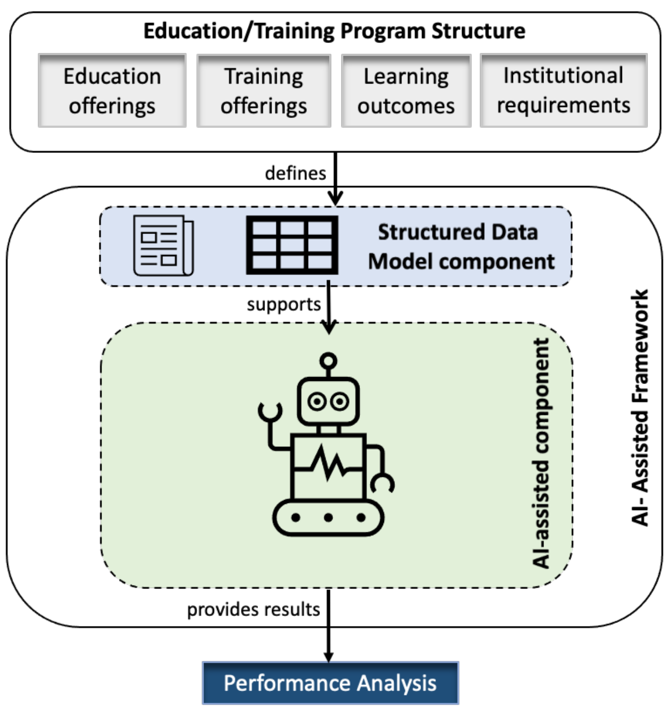

Figure 4. AI-assisted framework main components.

[4] Assessment mechanisms has to be defined in accordance with the Institution' Assessment Plan and related evidence to be collected.

3. Structured Data Model Component

Within the SDM component, educational program elements and their relationships are methodically organized, resulting in a database that will support advanced queries[5] for data exploration and analysis and incorporation of potential new competencies within an educational program.

A partial exemplar model of this nature is displayed as an Entity Relationship diagram in Figure 5.

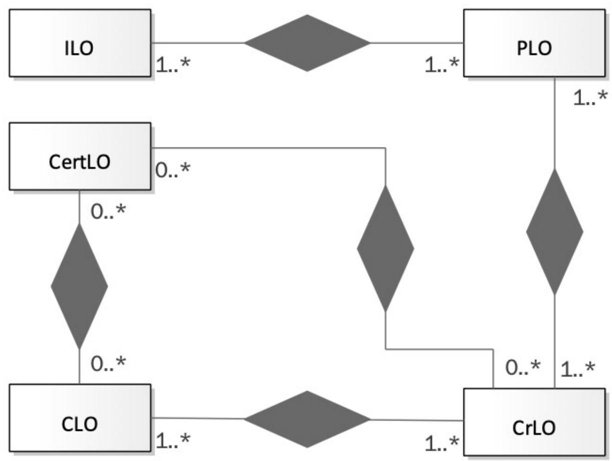

Figure 5. Example of entity-relationship diagram modeling the relationships among the different LOs.

This diagram is a visual representation adopted to model information or data and is used as a schema, which is a precursor to database modeling. The rectangles represent entities and diamonds represent connections between entities with specific cardinalities. Given a course LO (CLO), it is associated to one or more curriculum LOs (CrLOs), and given a CrLO, it is associated with one or more CLOs. The other relationships among institutional LOs (ILOs), program LOs (PLOs and certificate LOs (CertLOs) can be interpreted in a similar way

Reaching a balance is essential when preparing data for use with ML algorithms, ensuring it remains true to the underlying processes that generated it. During data preparation, analysis and organization, it is imperative to consider the following factors:

1. Data availability: Assess the availability of data required to implement the desired data features.
2. Timing: Consider the availability of data for inclusion in a feature prior to the event we aim to predict.

[5] Some of the advanced queries were described in the use cases presented in the previous Section A.

3. Longevity of features: Evaluate the potential for features to become outdated or stale over time.

4. AI-Assisted Component

The AI-assisted component's principal function is to categorize a specific competency within the existing educational or training offerings. Consequently, this component includes an AI model, trained on the LOs of these offerings. The AI model provides anticipated classifications along with a quantifiable level of accuracy. Given that LOs and competencies are expressed in natural language, the process employs NLP techniques. This transforms the data corpus of LOs into a suitable structure, subsequently divided into distinct datasets. Following these preliminary stages, the datasets are poised for training and testing the AI model. The schematic depiction in the Figure 6 illustrates these sequential steps: initial input consists of LOs with their respective classification labels (LOs data corpus), subject to NLP, succeeded by AI-model training, testing, and eventual classification. The Performance Analysis process assess the classification's validity for a given case.

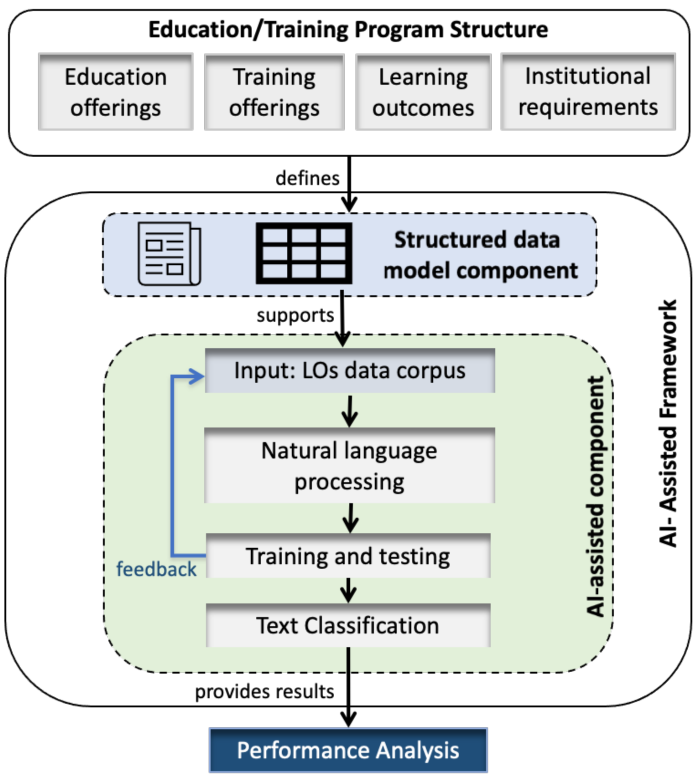

Figure 6. AI-assisted component.

After establishing an operational system embodying this framework, a series of interconnected procedures will ensure the ongoing refinement of the AI model in response to updates and novel inclusions within the educational program. The sequence of operations, as illustrated in Figure 7, encompasses a decision-making phase incorporating the AI model's capacity to evaluate and categorize competencies, followed by making refinements and integrating changes into the existing educational program. Upon these changes, the AI learning phase recommences, mandating the retraining and reevaluation of the AI model's accuracy and compatibility with the updated LOs data corpus.

Figure 8 illustrates the operational use of this framework in the context of a multi and interdisciplinary curriculum review and design, where results of advanced queries supported by the SDM, and new competencies classifications will support curricular reviews and design. Changes resulting from this process may trigger modifications in the current education or training program structure.

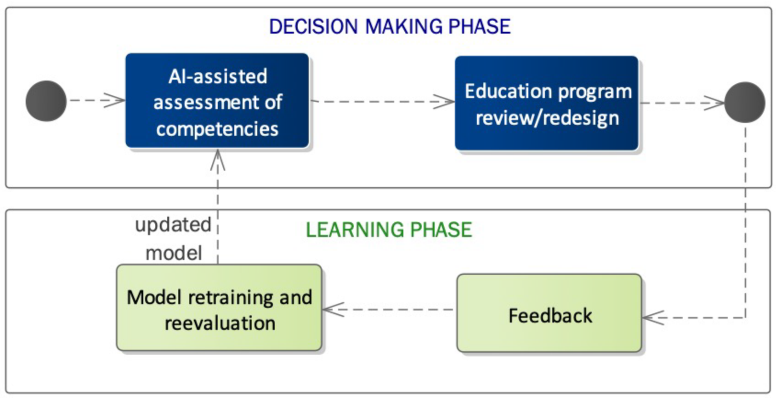

Figure 7. The decision making and learning phases of the AI model.

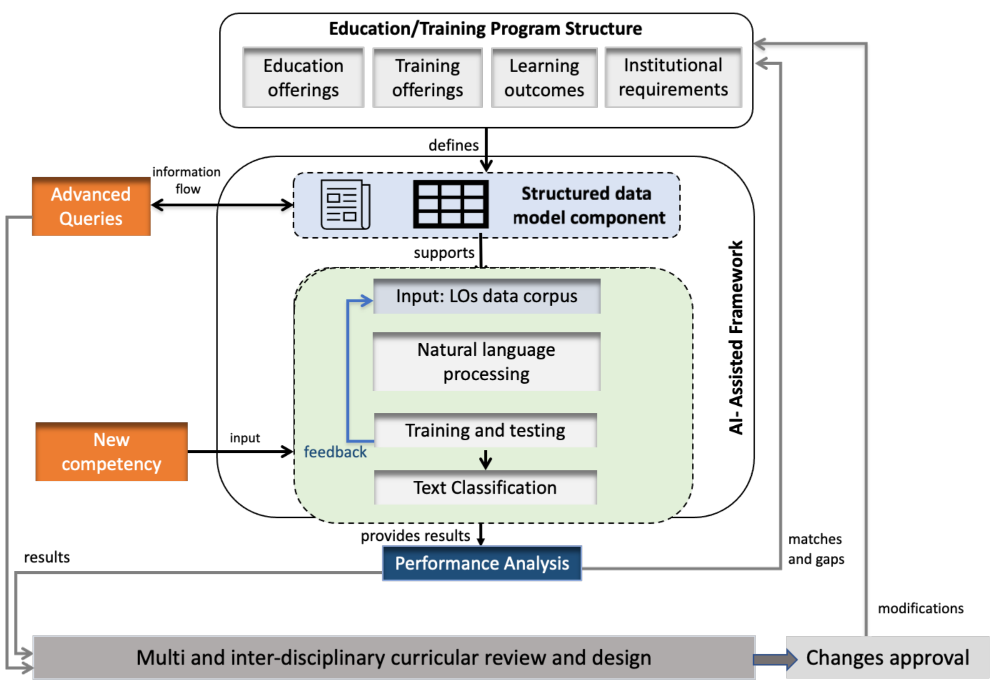

Figure 8. AI-Assisted Framework as a supporting tool for program and curricular review

C. FRAMEWORK STRATEGIC PERSPECTIVES

The framework's central strategic objectives are systematically delineated in a strategic map diagram depicted in Figure 9, offering crucial insights into its multifaceted significance. This visual representation encapsulates four pivotal perspectives, each encompassing vital objectives that articulate the framework's strategic utility: Data, AI-Model, Classification, and Recommendation. This diagram provides a tangible artifact for visual communication, promoting a shared comprehension of the framework strategy's essence and can be used effectively as an aid in evaluating the advancement of a system implementation that incorporate the framework. Within each of the four perspectives, interconnections among elements elucidate potential implementations across business, application, and technological tiers.

The Data perspective focuses on the fundamental objectives of establishing a Relational Database and crafting an Analytics Base Table (ABT). These main objectives are rooted in the underlying sub-objectives of developing the SDM and handling Unstructured Data, respectively, fed by a diverse array of data sources. The ABT serves as an input to the adjacent AI Model perspective. The presence and quality of this table directly impact the efficacy and sophistication of the adopted AI models.

The Relational Database objective supports the Recommendation perspective. The successful establishment and management of this database significantly affect the recommendations generated by the framework, underscoring the connections between the Data perspective and its ripple effects across the broader framework's strategic objectives.

The AI Model perspective focus on two core objectives: NLP of LOs and Trained Model. These objectives represent the foundational elements that drive the framework's AI capabilities. The NLP of LOs objective entails the analysis and understanding of LOs data through advanced NLP techniques. The Trained Model objective, which involves the utilization of three distinct datasets derived from the ABT (LOs data corpus), ensures that the model is trained based on a robust foundation of language-driven insights and classes.

The Trained Model objective supports the Classify Competency objective in the Classification perspective. The successful training of the model directly influences the ability to accurately classify and predict new competencies.

The Classification perspective has two central objectives: Classify Competency and Analyze Classification Performance. The Classify Competency objective holds the pivotal responsibility of categorizing and assigning competencies to appropriate classes. This classification process relies upon the Trained model developed in the AI Model perspective, which has been trained to comprehend LOs and competencies. Adjacent to this is the Analyze Classification Accuracy objective, which ensures the reliability and precision of the competency classification process.

The outcomes of the Classify Competency objective synergistically contribute to the Recommendation perspective. This connection underscores the significance of accurate competency classification as the foundation for generating pertinent recommendations. These recommendations are designed to provide a consolidated overview of competencies aligning with existing education offerings, providing comprehensive insights into potential pathways to attain the required competencies.

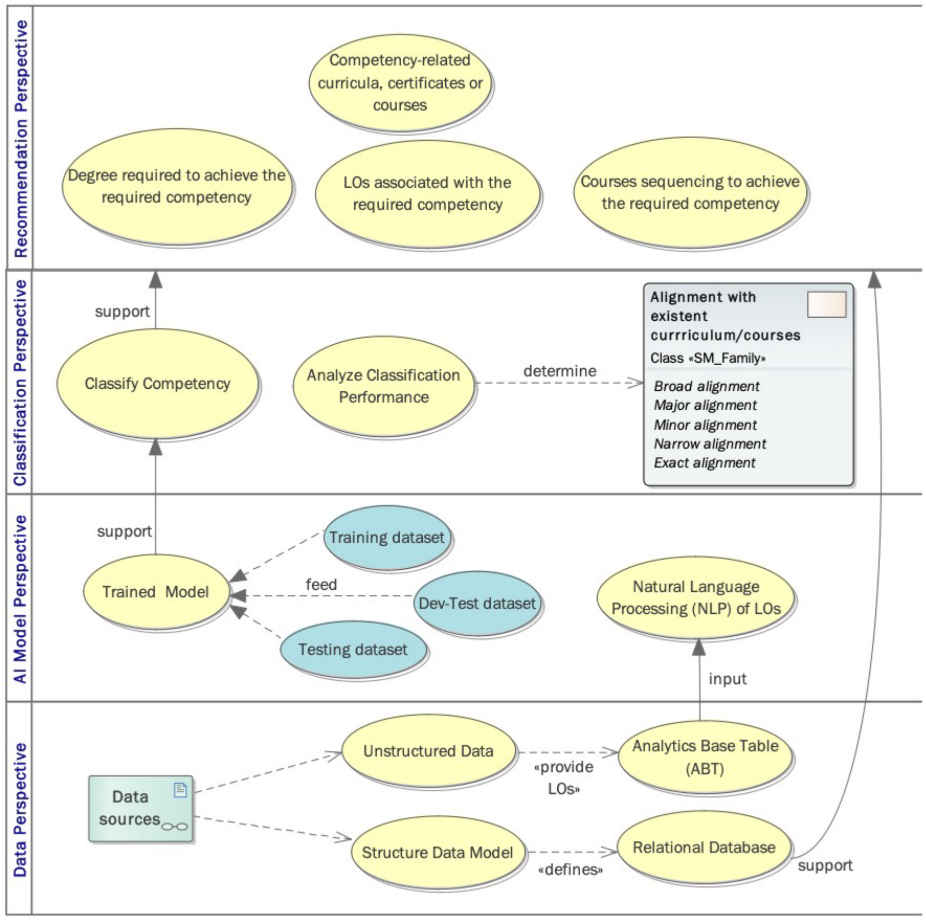

Figure 9. AI-Assisted Framework as a supporting tool for program and curricular review.

D. ADRESSING THE RESEARCH QUESTIONS

The integration of the proposed framework into an operational system addresses the research questions by providing:

a. Advancement of Curriculum Development: By conducting an exhaustive analysis of educational prerequisites, demanded warfighter competencies, and relevant variables, this framework promises the seamless incorporation of contemporary competencies into educational programs.

b. Adaptive Program Evaluation: Through the establishment of an ongoing process of continuous monitoring and evaluation, educational programs acquire the flexibility to promptly adjust to emerging competency requisites.

c. Simplified Regulatory Adherence: This framework simplifies the adherence process by correlating competencies with degree and certificate programs, thereby facilitating smoother alignment and supporting assessment and accreditation processes.

d. Informed Strategic Decision-Making: By offering valuable insights into required competencies, the framework informs strategic planning and consequent allocation of resources. This strategic approach enables the Navy and Marine Corps to prioritize educational initiatives based on criticality, further enhancing their effectiveness.

It is important for any project to follow a methodology that will ensure its successful execution. We suggested the Predictive Data Analysis Project Lifecycle methodology, described in Appendix A.

IV. THE NPS CASE

A. THE FRAMEWORK IN ACTION

The case study described in this section is the result of applying the proposed framework within the context of the NPS education model. This case study was executed concurrently with the development of the framework, aiming to gain deeper insights into the structuring of the framework's components and the potential AI models to be employed. Following the establishment of the framework's structure and an exhaustive analysis of viable AI techniques, we initiated a series of tests involving AI models to validate certain assumptions pertaining to the framework's AI-assisted component. The ensuing process is visually depicted in Figure 10.

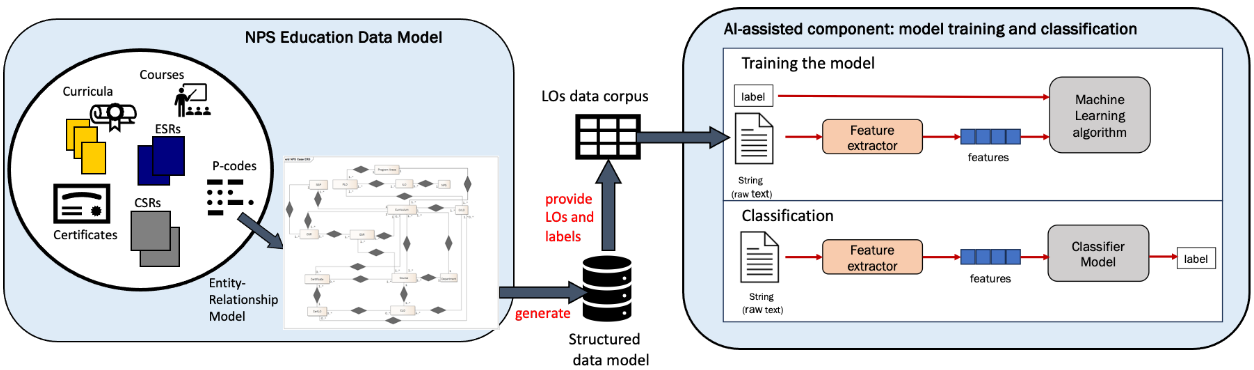

Figure 10. NPS Case: Framework in action.

B. NPS EDUCATION DATA MODEL

In this phase, we defined the elements of the NPS education model and how they interconnect. We conducted a comprehensive analysis of twenty-three diverse curricula to inform the creation of our LOs data corpus. For numerous courses, LOs weren't explicitly outlined in the Academic Catalog, prompting us to construct them based on available course descriptions to ensure completeness.

We established the key components of the data model, rooted in NPS educational offerings, to exemplify the essential relationships within an outcome-based education program. The primary objective was to create a robust data model capable of supporting advanced queries and correlating insights derived from the AI-assisted component. For additional information about the model's definition, please refer to Appendix B.

Figure 11 provides a visual representation of the principal elements within the NPS education model and their interconnections.

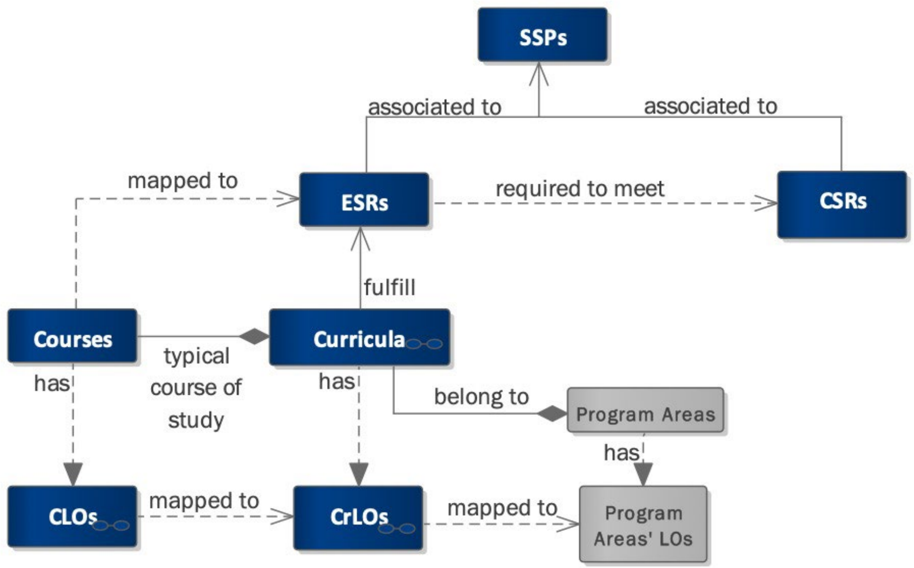

Figure 11. NPS Educational Structure: Main Elements and their Relationships.

To address these discrepancies, we constructed an Analytical Base Table (ABT) for our LOs data corpus by sampling from various curricula. Our rationale was grounded in the fact that outcomes-based education measures a student's ability to demonstrate LO achievement, which should translate into competencies upon graduation.

Initially, we experimented with different data formats (structured and non-structured) for LOs. We tried to categorize them into features such as action verbs, objects, and object modifiers based on LO descriptions from nine distinct curricula. However, this approach proved ineffective due to issues revealed in the data quality report. Data exploration uncovered problems like high standard deviation, irregular cardinality, long tails, and outliers in the distribution of features. Consequently, using such features for ML algorithms would not yield accurate classifications and predictions.

Based on the results of our experimentation, we consolidated our recommendation to employ text classification techniques using NLP to leverage textual descriptions of LOs and competencies. This approach eliminates the reliance on atomic categorical data features and aligns better with our objectives.

The uneven distribution of LOs across different curricula raised concerns about potential bias in our predictions. To mitigate this, we suggest establishing an average number of LOs per curriculum, with the goal of minimizing standard deviation. Consult Appendix B, section D, for the resulting distribution of LOs across the curricula.

C. AI- ASSISTED COMPONENT: MODEL TRAINING AND CLASSIFICATION

The primary goal of this task was not to delve deeply into the intricacies of model training or classification, nor was it to refine the LOs data corpus for enhanced performance. Instead, our focus was on assessing several candidate AI models within the scope of our scenario, essentially making a compelling case for their prospective integration into the proposed framework. It's important to note that the meticulous fine-tuning of both the models and the LOs data corpus lies beyond the purview of this project, and this aspect will be the subject of further research.

AI models can be classified into two types: generative and discriminative. Generative AI models aim to model the entire probability distribution of data and are often used for tasks such as text generation, image generation, and data augmentation. Discriminative AI models, on the other hand, focus on modeling the boundary or decision boundary between different classes or categories in the data and are typically used for classification tasks.

For the NPS case study, we used 5 different models:

1. A dummy classifier was used as a baseline model for the classification task. This model made predictions based on simple rules, such as random guessing or always predicting the majority class.

2. Decision trees are discriminative models used for classification and regression tasks. They partition the input space into regions and make decisions based on the input features to classify or predict outcomes.

3. Maximum entropy models, also known as MaxEnt models, are generative models used in various NLP and ML tasks. They model the probability distribution of data, allowing them to generate new samples that are consistent with the observed data distribution.

4. Naïve Bayes is a generative probabilistic model commonly used for classification tasks. It models the joint probability distribution of features and class labels, allowing it to generate probability distributions for different classes and make classification decisions based on these probabilities.

5. Transformers are a versatile architecture that can be used for both generative and discriminative tasks. They gained popularity for their ability to handle various NLP tasks, including text generation (e.g., GPT-3) and classification (e.g., BERT). We used DistilBERT [53] for this study.

Table 2 presents the best accuracy obtained by training these models with the LOs corpus created for this case study. For other metrics related to each model, including their confusion matrix, consult Appendix C, Table 3.

Table 2. Selected AI models for NPS Case Study

Algorithm	Type	Accuracy on the dev dataset	Accuracy on the test dataset
Dummy Classifier	Baseline model	0.22	0.20
Decision Tree	Discriminative	0.59	0.60
MaxEnt	Generative	0.70	0.72
Naïve Bayes	Generative	0.67	0.58
Transformers	Generative and Discriminative	0.64	0.60

Looking at the accuracy, all four AI models have demonstrated better performance than our simple baseline (Dummy Classifier).

Even the simple Decision Tree classifier is significantly better than our baseline model. The confusion matrix presented in Appendix C, Figure 21, visually shows the Decision Tree classifier's performance, presenting the association between the true and predicted labels.

The MaxEnt Classifier demonstrated superior performance in predicting outcomes for both the development and test datasets, with Naïve Bayes and Transformers Classifiers closely trailing on the development dataset. When it comes to the test dataset, Transformers and Decision Tree Classifiers exhibited the second-highest level of accuracy. In future investigations, it is advisable not to dismiss any of these models, as they all hold promise as potential candidates for effective classification. It's important to note that as datasets evolve, the most suitable model may also change, emphasizing the need for a case-specific analysis when selecting the optimal AI model.

While it is always prudent to maintain such a reference point, even without it, the results indicate significant potential in employing these AI models for the classification of new competencies based on a dataset of LOs. Although the achieved accuracy falls slightly short of what we would like to see (ideal threshold above 0.8), this result must be evaluated within the context of several key factors that exerted considerable influence on the model's performance and its suitability for this case study:

1. **Data Inequalities**: The presence of imbalances in the dataset significantly impacted the models' performance, posing a challenge in achieving higher accuracy.
2. **Outliers and Irregular Cardinality**: The existence of outliers and irregular data cardinalities introduced complexities that affected the models' ability to generalize effectively.
3. **Non-Unique Classification Labels**: The fact that some LOs shared classification labels due to their association with one or more certificates or curricula added intricacy to the classification task.
4. **Granularity of LO Descriptions**: The varying granularity of LO descriptions, ranging from general to highly specific, further influenced the model's performance.
5. **High Number of Unique Labels**: The extensive variety of curricula offered by the school resulted in a substantial number of unique classification labels, which presented an additional challenge for the models.

In the NPS case study, we opted for simplicity by working with the raw, unbalanced class frequencies, considering that this specific study was an addition to the original project plan. However, it is worth noting that there are techniques available to address these challenges within the dataset, which could be employed to enhance the prediction accuracy of the models in future research.

In general, generative models possess greater inherent capability than conditional models because they calculate conditional probabilities from joint probabilities rather than the other way around. However, this increased potency comes at a cost. Since Naïve Bayes, being more potent, involves more free parameters that must be learned while keeping the training set size constant, it leads to a scenario where there is less data available for training each parameter's value. Consequently, a generative model may not perform as effectively as a conditional model in predicting the most probable label for a given input.

V. CONCLUSIONS

This research proposes a framework that leverages AI to assist in both the creation and assessment of OBE/CBE programs, particularly those with interdisciplinary and multidisciplinary aspects. The primary objective is to enable educational institutions to align their offerings with contemporary and emerging competency demands. By implementing a functional system based on this framework, the intention is to facilitate well-timed strategic adjustments within educational programs, curricula, and courses. These adjustments are designed to effectively address the dynamic landscape of higher education, driven by the emergence of novel technologies and the imperative of upskilling.

Before defining the framework, we conducted an in-depth analysis of OBE as a strategic approach for education and training, emphasizing the importance of student's LOs. Additionally, we investigated AI concepts and methodologies to determine their suitability within the proposed framework. As part of the framework's requirements elicitation process, we defined use cases for potential stakeholders, examined AI analytical methods, and assessed their compatibility with our desired solution. We also considered data structuring standards to ensure compliance with DOD policies and instructions regarding data interoperability and the adoption of AI systems.

The resulting framework comprises of two primary components: the **SDM** component and the **AI-Assisted** component. Within the SDM component, educational program elements and their relationships are methodically organized, resulting in a database that will support advanced queries for data exploration and analysis and incorporation of potential new competencies within an educational program.

The AI-Assisted component's primary function is to categorize specific competencies within existing educational or training offerings. This component includes an AI model trained on the LOs of these offerings, providing anticipated classifications and a quantifiable level of accuracy. Given that LOs and competencies are expressed in natural language, we also used NLP techniques.

The framework's central strategic objectives are systematically delineated in a strategic map, offering crucial insights into its multifaceted significance. This map representation encapsulates four pivotal perspectives, each encompassing key objectives that articulate the framework's strategic utility: Data, AI-Model, Classification, and Recommendation.

In addition to proposing the framework, we conducted a comprehensive case study using the NPS educational model. Our methodology involved selecting specific curricula and applying the conceptual framework to assess its effectiveness. We explored the operational dynamics and performance of four AI-based classifier, which processes natural language competency descriptions as input and assigns classifications or predictions concerning their compatibility with existing educational offerings.

The AI-classifiers were trained using existing LOs in representative degree programs at NPS. Despite the challenges posed by data quality, we achieved approximately 70% accuracy on the test dataset predictions for one of the AI classifiers, even with a limited and less than robust dataset. These results not only demonstrate the feasibility of our proposed approach based on the promising results, but also highlight the potential benefits of this AI application for DOD education and training institutions. We expected to test different LOs data corpus and AI algorithms in the future, with the possibility of prototyping a tool based on the proposed framework.

LIST OF REFERENCES

1. CNO NAVPLAN 2021. (January 2021). [URL] https://media.defense.gov/2021/Jan/11/2002562551/-1/-1/1/CNO%20NAVPLAN%202021%20%20FINAL.PDF

2. Force Design 2030 Report. DON. (March 2020). [URL] https://www.hqmc.marines.mil/Portals/142/Docs/CMC38%20Force%20Design%202030%20Report%20Phase%20I%20and%20II.pdf?ver=2020-03-26-121328-460

3. One Navy-Marine Corps Team: Strategic Guidance from The Secretary of The Navy. (October 2021). [URL] https://nps.edu/documents/106660594/132228011/SECNAV+ONE+NAVY-MARINE+CORPS+TEAM+STRATEGY.pdf/ddf7fc24-369f-ee5d-ca41-fe19a3982a7d?t=1633736345394

4. Naval Education Strategy 2023. Secretary of the Navy. (June 2023). URL https://media.defense.gov/2023/Jun/23/2003246843/-1/-1/0/ASN_MRE_NES_DIGITAL.PDF

5. Pelfrey Sr, W. V., & Pelfrey Jr, W. V. (2009). Curriculum Evaluation and Revision in a Nascent Field: The Utility of the Retrospective Pretest–Posttest Model in a Homeland Security Program of Study. *Evaluation Review*, 33(1). 10.1177/0193841X08327578. [URL] http://erx.sagepub.com

6. Dede, C., Etemadi, A., & Forshaw, T. (2021). Intelligence augmentation: Upskilling humans to complement AI. The Next Level Lab at the Harvard Graduate School of Education. President and Fellows of Harvard College: Cambridge, MA.

7. Rose, S. (2020, May). How Artificial Intelligence helps to improve Decision Making? [URL] https://becominghuman.ai/how-artificial-intelligence-helps-to-improve-decision-making-347a67f9127f

8. Fridgeirsson, T., et al. (2021). An Authoritative Study on the Near Future Effect of Artificial Intelligence on Project Management Knowledge Areas. Sustainability, 13(23), 2345. https://doi.org/10.3390/su13042345. [URL] https://www.mdpi.com/journal/sustainability

9. Karandish, D. (2021, June). 7 Benefits of AI in Education. Karandish. [URL] https://thejournal.com/articles/2021/06/23/7-benefits-of-ai-in-education.aspx

10. Zawacki-Richter, et al. (2021). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019

11. Ahmad, et al. (2021). Artificial Intelligence and its Role in Education. *Sustainability, 13*(21), 12902. https://doi.org/10.3390/su132212902

12. Gunawan, K. D. H., Liliasari, L., Kaniawati, K., & Setiawan, W. (2021). Implementation of Competency Enhancement Program for Science Teachers Assisted by Artificial Intelligence in Designing HOTS-based Integrated Science Learning. *IPA JPPI, 7*(1), 55-65. p-ISSN 2477-1422 e-ISSN 2477-2038.

13. Somasundarama, M., Lathab, P., & Saravana Pandianc, S. A. (2020). Design Using Artificial Intelligence (AI) Back Propagation Method. In *9th World Engineering Education Forum, WEEF. Procedia Computer Science, 172*, 134-138.

14. Rudzajs, P., & Kirikova, M. (2021/2022). Monitoring Services to Support Continuous Curriculum Engineering. *Complex Systems Informatics and Modeling Quarterly (CSIMQ)*, 29(4), 53-73. Published online by RTU Press. https://doi.org/10.7250/csimq.2021-29.04.

15. Naval Education Strategy. (2023). [URL] https://media.defense.gov/2023/Jun/23/2003246843/-1/-1/0/ASN_MRE_NES_DIGITAL.PDF

16. Ready Relevant Learning (RRL). (n.d.). Retrieved from https://www.netc.navy.mil/RRL/

17. U.S. Joint Chiefs of Staff. (2022, April 1). Outcomes-Based Military Education Procedures for Officer Professional Military Education, CJCSM 1810.01.

18. Department of Defense. (2017). DOD Competency Definition. *DOD Instruction 1400.25*, June 2017. Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/140025/140025_vol250.pdf

19. C-BEN. (2017). Quality Framework for Competency-Based Education Programs. Released September 2017. Retrieved from https://www.cbenetwork.org/wp-content/uploads/2018/09/Quality-Framework-for-Competency-Based-Education-Programs-Updated.pdf

20. Advanced Distributed Learning. (2023.). [URL] https://adlnet.gov/

21. Advanced Distributed Learning. (2018). Competency-based Learning. [URL] https://adlnet.gov/publications/2019/08/Competency-Based-Learning/

22. Advanced Distributed Learning (ADL). (2020, March 3). Competency Framework Development Process Report. Retrieved from https://adlnet.gov/assets/uploads/2020-03%20Competency%20Framework%20Development%20Process%20Report.pdf

23. Advanced Distributed Learning (ADL). (n.d.). Competency and Skills System (CaSS). Retrieved from https://adlnet.gov/projects/cass/

24. IMS Global Learning Consortium. (n.d.). Competencies and Academic Standards Exchange. Retrieved from https://www.imsglobal.org/activity/case

25. Achievement Standards Network (ASN). (n.d.). Achievement Standards Network (ASN) Description Framework (ASN-DF). Retrieved from http://standards.asn.desire2learn.com/

26. O*NET Resource Center. (n.d.). O*NET Program. Retrieved from https://www.onetcenter.org/overview.html

27. Pelfrey Sr, W. V., & Pelfrey Jr, W. V. (2009). Curriculum Evaluation and Revision in a Nascent Field: The Utility of the Retrospective Pretest–Posttest Model in a Homeland Security Program of Study. *Evaluation Review, 33*(1). https://doi.org/10.1177/0193841X08327578

28. United States Department of the Navy. (2019, July). OPNAV INSTRUCTION 1520.23C: Graduate Education.

29. United States Department of the Navy. (2019, April). SECNAV 1524.2D: Policies Concerning the Naval Postgraduate School.

30. United States Department of Defense. (2008, April 29). DOD INSTRUCTION 1322.10: Policy on Graduate Education for Military Officers.

31. Naval Postgraduate School. (2016, February 11). NPS INSTRUCTION 1550.1G: Guidelines for Conducting Curricular Reviews.

32. Credential Engine. (2023). Credential Engine Schemas Handbook, October 2023 CTDL-ASN Release. [URL] https://credreg.net/ctdl/handbook#ctdlasnschema

33. Xiao, P. (2022). Artificial Intelligence Programming with Python: From Zero to Hero. John Wiley & Sons, Inc.

34. Aggarwal, C. C. (2021). Artificial Intelligence: A Textbook (1st ed. 2021). Springer International Publishing. ISBN: 9783030723590.

35. Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2020). Fundamentals of ML for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies (2nd ed.). The MIT Press.

36. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (Adaptive Computation and ML series). The MIT Press. Hardcover, 800 pages. ISBN-10: 0262035618, ISBN-13: 978-0262035613.

37. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

38. Cortes, C., & Vapnik, V. (1995). Support-vector networks. ML, 20(3), 273-297.

39. Breiman, L. (2001). Random forests. ML, 45(1), 5-32.

40. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21-27.

41. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

42. Bishop, C. M. (2006). Pattern Recognition and ML. Springer.

43. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232.

44. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.

45. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.

46. Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). Practical Natural Language Processing. O'Reilly Media, Inc.

47. Tunstall, L., von Werra, L., & Wolf, T. (2022). Natural Language Processing with Transformers, Revised Edition. O'Reilly Media, Inc.

48. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media, Inc.

49. Bengfort, B., Bilbro, R., & Ojeda, T. (2018). Applied Text Analysis with Python: Enabling Language-Aware Data Products with ML. O'Reilly Media, Inc.

50. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.

51. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532-1543.

52. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 1, 4171-4186.

53. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. OpenAI. Retrieved from https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

54. Sanh, V., Debut, L., Chaumond, J., & Courbariaux, M. (2019). DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. arXiv preprint arXiv:1910.01108.

55. Natural Language Toolkit (NLTK). (2023). NLTK 3.8.1 release. Retrieved from https://www.nltk.org/?utm_source=thenewstack&utm_medium=website&utm_content=inline-mention&utm_campaign=platform

56. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser. L. & Polosukhin, I. (2017). Attention Is All You Need. In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. Retrieved from https://arxiv.org/abs/1706.03762

57. Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. arXiv preprint arXiv:1801.06146.

58. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 (SPSS Inc.).

59. Wehrstein, L., & Bachmann, B. (2020). CRISP-DM ready for ML Projects. Dec 19, 2020. Posted in [URL] https://towardsdatascience.com/crisp-dm-ready-for-machine-learning-projects-2aad9172056a

60. NAVPERS 15839I. (2022). Manual of Navy Officer Manpower and Personnel Classifications. Volume I. Major Code Structures. Oct 2022.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A: PREDICTIVE DATA ANALYSIS PROJECT LIFECYCLE

It is important for any project to follow a methodology that will ensure its successful execution. In the case of AI-related projects, we suggest the adoption of the well-known and widely accepted Cross-Industry Standard Process for Data Mining (CRISP-DM) to reflect ML specifics. Figure 12 illustrates the adapted CRISP- DM method for this case, which contains all the phases of the original CRISP-DM [58] with the addition of Model Management.

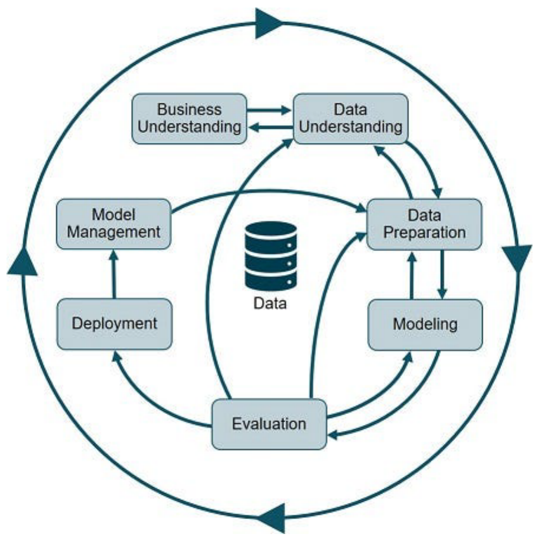

Figure 12. CRISP-DM with Model Management (copyright: Wehrstein, L.[59]).

Each phase is briefly described as follows:

a. Business Understanding: In this phase, the project objectives and requirements are defined to establish a clear understanding of the business goals and constraints of the data mining project.
b. Data Understanding: During this phase, data sources are explored and assessed to gain insights into the available data and identify potential issues for further analysis.
c. Data Preparation: This phase involves cleaning, transforming, and integrating data to create a suitable dataset for analysis, ensuring its quality and relevance.
d. Modeling: Models are built and evaluated based on the selected algorithms, with the goal of identifying patterns, relationships, or trends in the data.
e. Evaluation: In this phase, the models are assessed to determine their effectiveness in meeting the business objectives and requirements, using appropriate evaluation techniques.
f. Deployment: The chosen model is deployed into the operational environment, integrating it with the business processes to achieve the desired impact.
g. Model Management: this phase includes the monitoring and management of the final model in ML.

APPENDIX B: NPS DATA MODEL

A. DEFINITIONS AND RESPONSIBILITIES [60]

Educational Skill Requirement (ESR). Specific and measurable statement of what the student must know or be able to do as the outcome of participation in an education program.

Core Skill Requirement (CSR). Specify the functional areas covered by a subspecialty discipline. They are a set of quantifiable skills, traits, and experiences that a subspecialist must possess to perform acceptably in a coded billet.

Major Area Sponsor. Within the Navy Subspecialty System framework, a Navy flag officer responsible for the requirements and resources of a broad range of curricula grouped into a particular category, including defining core skill requirements, educational skill requirement, billets, and quotas.

Subject Matter Expert. Within the Navy Subspecialty System framework, a person responsible for administration and management of educational skill requirements for specific curricula as assigned by the major area sponsor.

Subspecialty. An additional set of skills acquired by a member that are necessary for optimal performance of assigned duties in a Navy billet.

Source of graduate education: Naval Postgraduate School (NPS) is the Navy's primary source of graduate education. Each program is specifically designed to match educational skill requirements with the knowledge, skills, and abilities required by the major area sponsor. The Naval War College is the Navy's primary source for graduate-level Navy professional military education (PME) and joint professional military education (JPME). Civilian institutions and other military education institutions that are able to meet the major area sponsor curricular requirements, and that provide cost-effective, efficient delivery of timely, relevant, quality education programs, may be used as a source of graduate education.

Responsibilities highlight: the Vice-Chief of Naval Operations conduct a biennial assessment of graduate education and prepare report for Assistant Secretary of the Navy (Manpower and Reserve Affairs) submission to the Under Secretary of Defense (Personnel and Reserve Affairs) by 30 November of even numbered years. The Chief of Naval Personnel establish and maintain metrics to measure the return on education investments in collaboration with the major area sponsors (MAS).

Among several other responsibilities, the President of NPS, is accountable for:

a. Maintain a fully accredited academic institution whose curricula and programs fulfill validated education requirements to increase combat effectiveness of the Navy and Marine Corps.
b. Conduct curriculum reviews, in conjunction with major area sponsor, type commanders, and other stakeholders, at least biennially to ensure programs are academically sound and are being conducted per accreditation standards and title 10, United States Code, direction pertaining to NPS. The curriculum review process shall ensure maintenance of fundamental graduate level educational requirements despite changes to rigor or length of

time of educational programs. NPS shall maintain a majority voice in how a curriculum is best delivered and shall publish guidance on the curriculum review process.

B. DATA MODEL

Figure 13 presents the Entity Relationship diagram for the NPS case.

The data modeling diagram showed in Figure 14 is used to create and view graphical representations of relational database schemas. In this representation, rectangles symbolize entities that will be converted into tables, while diamonds denote the relationships between these entities along with their respective cardinalities. Depending on the cardinality, these relationships can be expressed either as separate tables or as foreign keys within the tables that represent the associated entities.

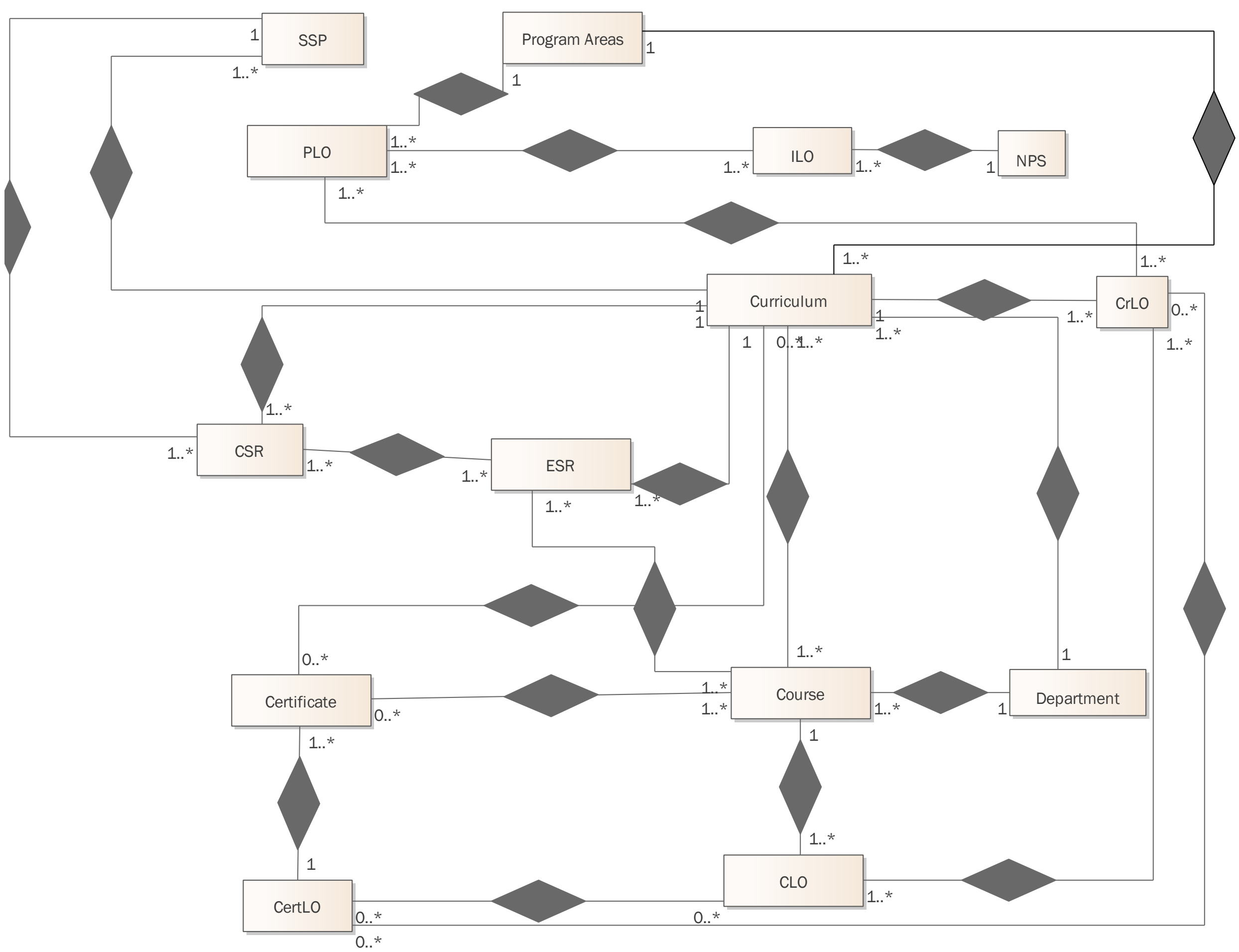

Figure 13. Entity Relationship diagram for NPS case.

Figure 14. Definition Data Model for NPS Case: Relationships are represented in Joint tables.

C. MOVES CURRICULUM MODEL

This subsection presents the structural model for the curriculum 399 -MOVES. It includes the set of diagrams designed to represent MOVES's main elements and their relationships, following the NPS education structure presented in Figure 15.

The MOVES curriculum is associated with two different SSPs: USMC 8825 and USN 6202. For each SSP, there is a different set of CSRs. Figure 15 presents the MOVES model top-level diagram.

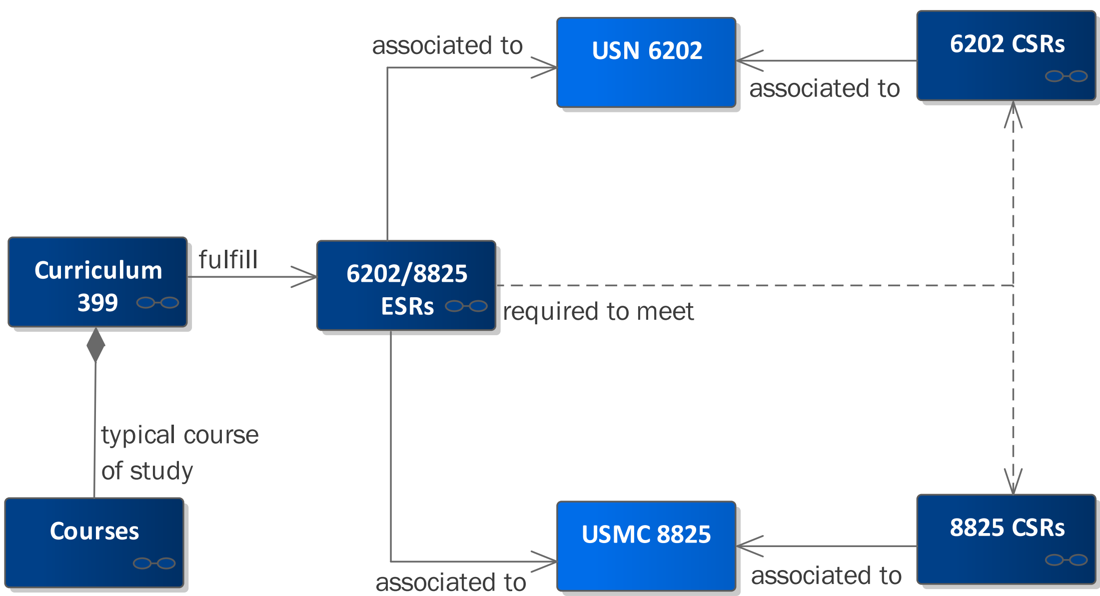

Figure 15. MOVES Top-level Structure Diagram.

MOVES typical course of study is presented visually in Figure 16. For each course listed in Figure 16, the model represent its relation to each of the course's LOs, visually showed in Figure 17.

Figure 18 presents the visual mapping between ESRs and courses in the curriculum 399. Both SPPs 6202 CSRs and SPP 8825 CSRs were mapped to ESRs. Figure 19 shows the resulting diagram.

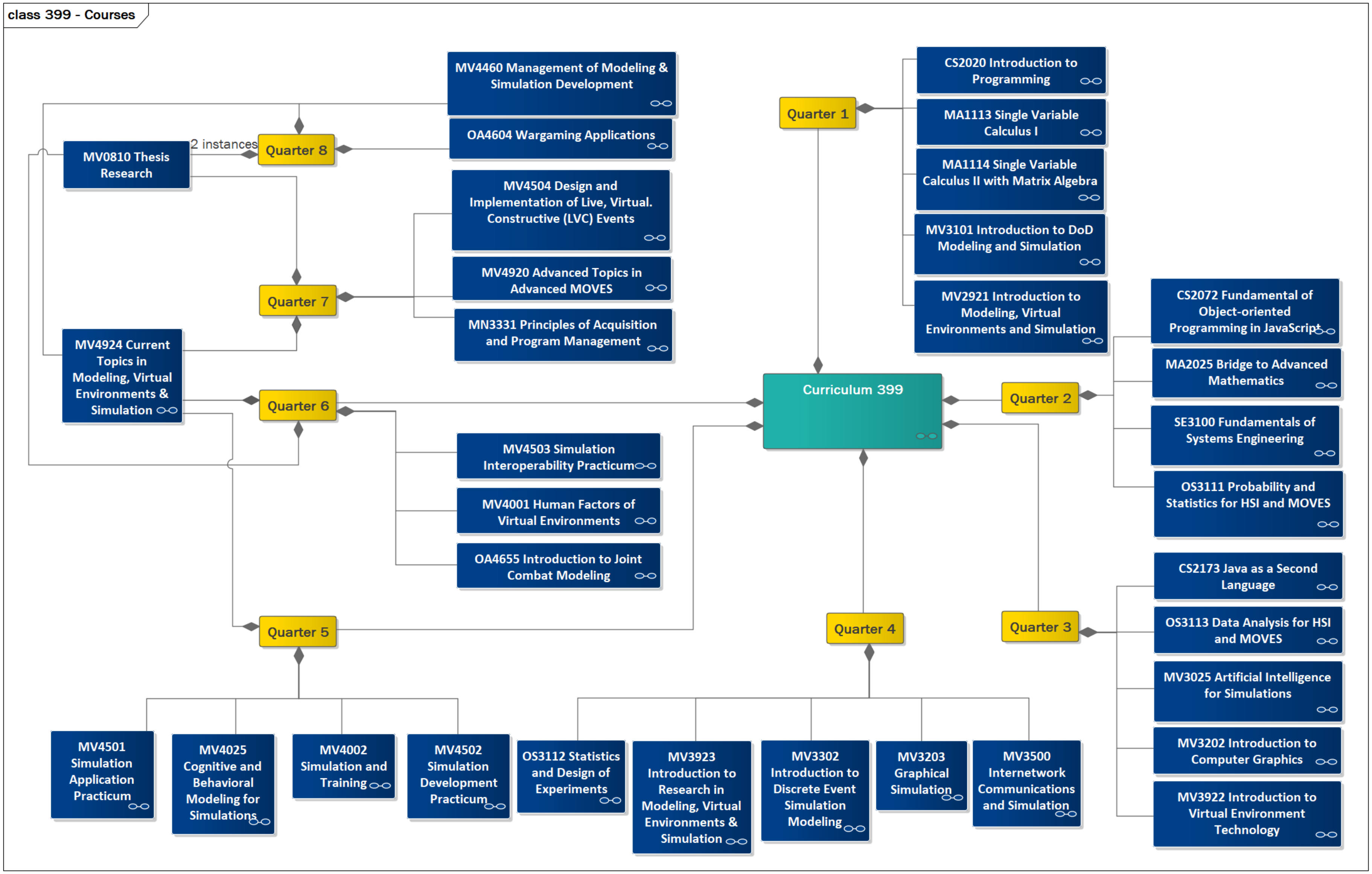

Figure 16. Curriculum 399 typical course of study.

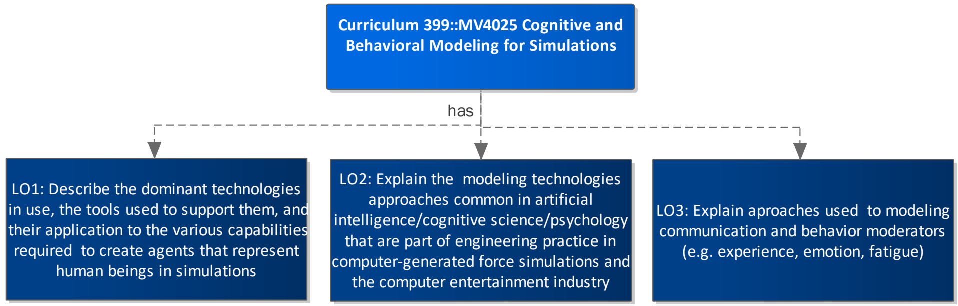

Figure 17. Course *Cognitive and Behavioral Modeling for Simulations'* LOs.

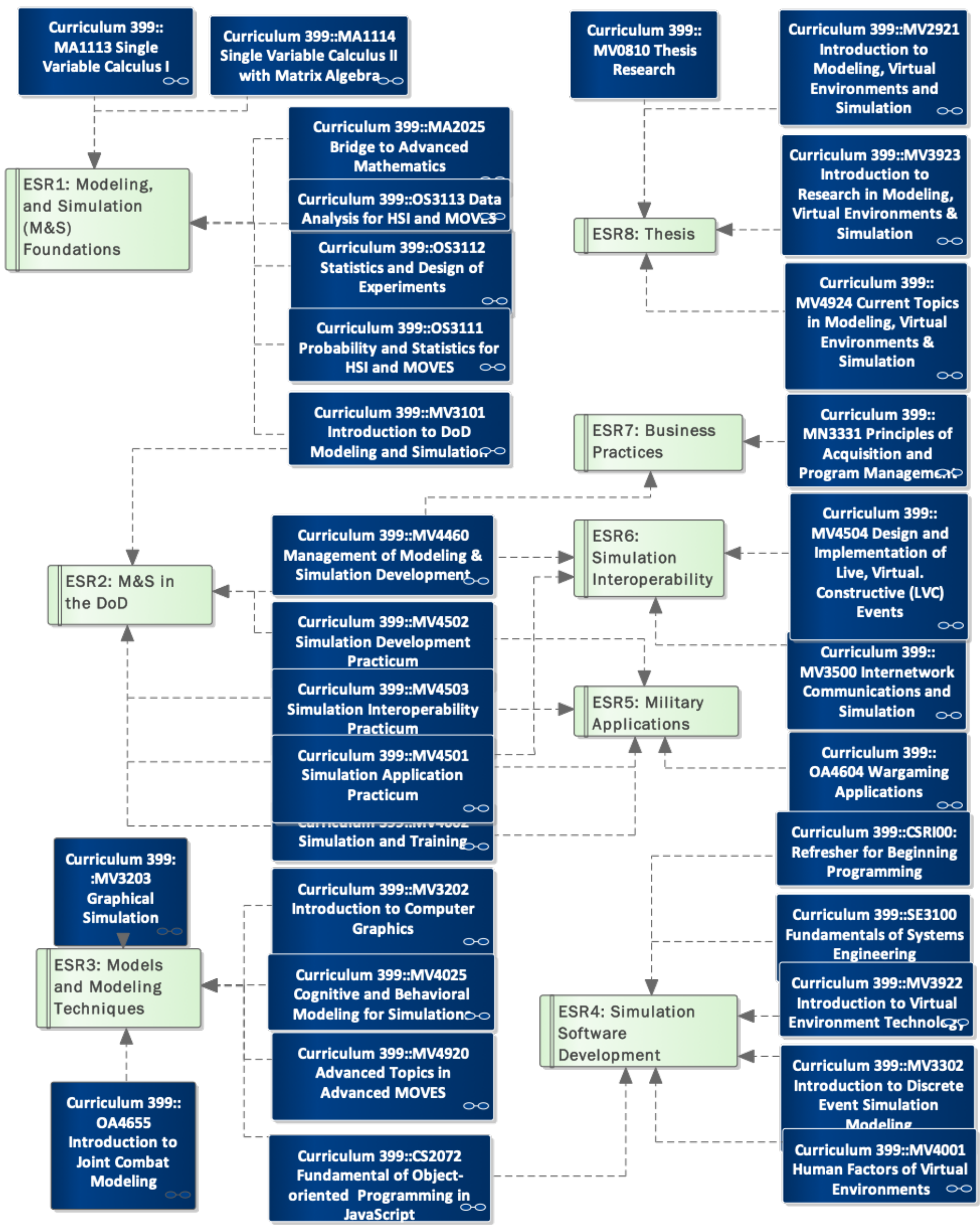

Figure 18. MOVES ESRs to Course Crosswalk

Figure 19. CSRs (SPP 6202 and 8825) and ESRs mapping.

APPENDIX C: NPS CASE AI MODELS PERFORMANCE

A. STUDY CASE LOs DATA CORPUS

Given that we are tackling a text classification challenge, it is prudent to scrutinize the distribution of LOs (LOs) across the curricula. A dataset with an imbalanced class distribution may necessitate a distinct approach in terms of training loss and evaluation metrics compared to a balanced dataset. In this case study, our LOs data corpus comprises 1,821 labeled LOs associated with their respective curricula. The frequency distribution of these LOs is visually presented in Figure 20.

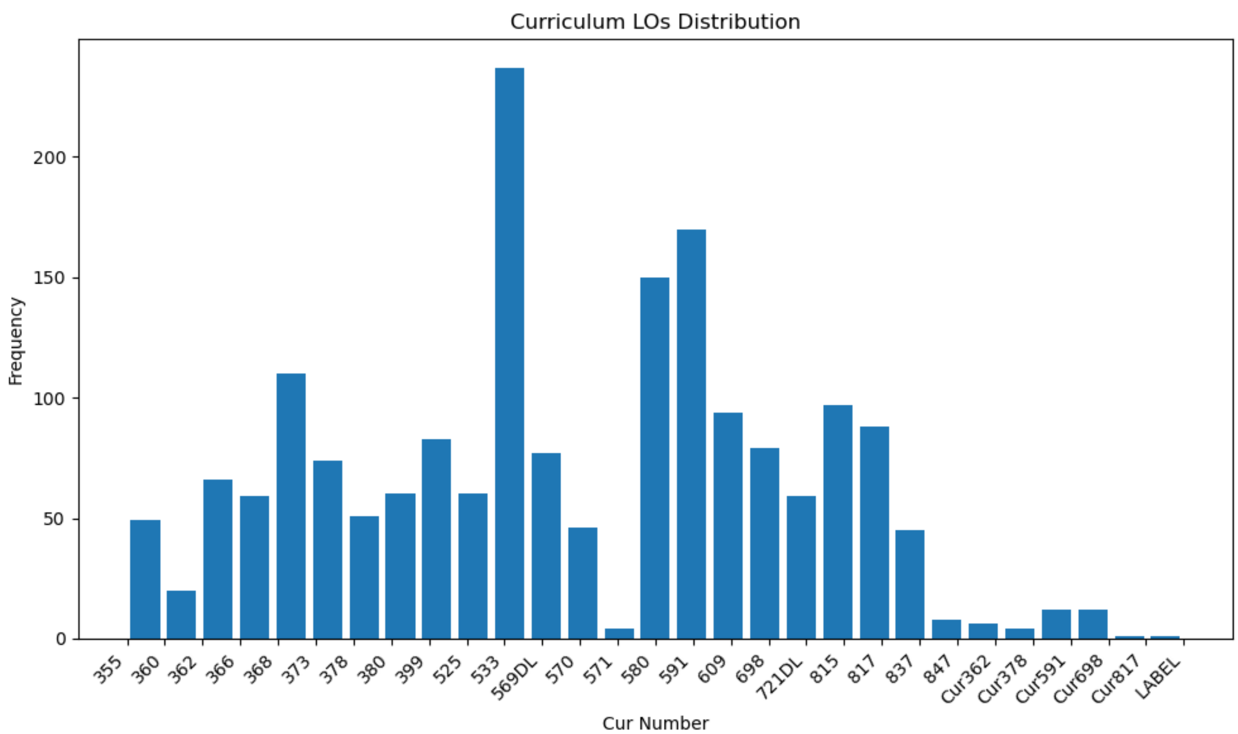

Figure 20. Labels frequency.

B. CLASSIFIERS' RESULTS ANALYSIS

The best resulting metrics for each AI model during several training cycles are presented in Table 3. Figure 21 presents the confusion matrix using Decision Tree algorithm.

Table 3. AI models resulting metrics

Metric	Decision Tree	Max Entropy	Naïve Bayes	Transformers
Accuracy on the development test set:	**0.59**	**0.72**	**0.67**	**0.64**
Weighted F1 Score on the development test set	0.58	0.72	0.62	0.59

Precision on the development test set:	0.59	0.72	0.67	0.64
Recall (Sensitivity) on the development test set	0.59	0.72	0.67	0.64
Specificity on the development test set	0.59	0.72	0.67	0.64
False Positive Rate on the development test set	0.41	0.28	0.33	0.36
Accuracy on the test set	0.60	0.70	0.58	0.60
Weighted F1 Score on the test set	0.58	0.68	0.52	0.53
Precision on the test set	0.60	0.70	0.58	0.60
Recall (Sensitivity) on the test set	0.60	0.70	0.58	0.60
Specificity on the test set	0.60	0.70	0.58	0.60
False Positive Rate on the test set	0.40	0.30	0.42	0.40

Figure 21. Decision Tree Confusion Matrix.

Figure 21 displays the confusion matrix generated using the decision tree algorithm. This matrix provides valuable insights into the accuracy of our predictions. The numbers along the diagonal indicate correct predictions, while those outside the diagonal highlight instances where the classifier erred in predicting the correct label. Interestingly, even these mispredictions can be viewed as constructive since they often occur when curricula associated with the true label and the predicted label share significant similarities.

For instance, consider the confusion matrix generated using the max entropy algorithm, as showcased in Figure 22. Here, we observe that label 591 (Master of Science in Space Engineering) was predicted as label 533 (Master of Science in Applied Physics of Combat Systems) on five occasions. These two curricula share common courses and, consequently, overlapping LOs.

Moving forward, Figure 23 and Figure 24 present the confusion matrices for the Naïve Bayes algorithm and the Transformers method, respectively. Additionally, Figure 25 provides the confusion matrix for the Dummy Classifier for comprehensive analysis and comparison.

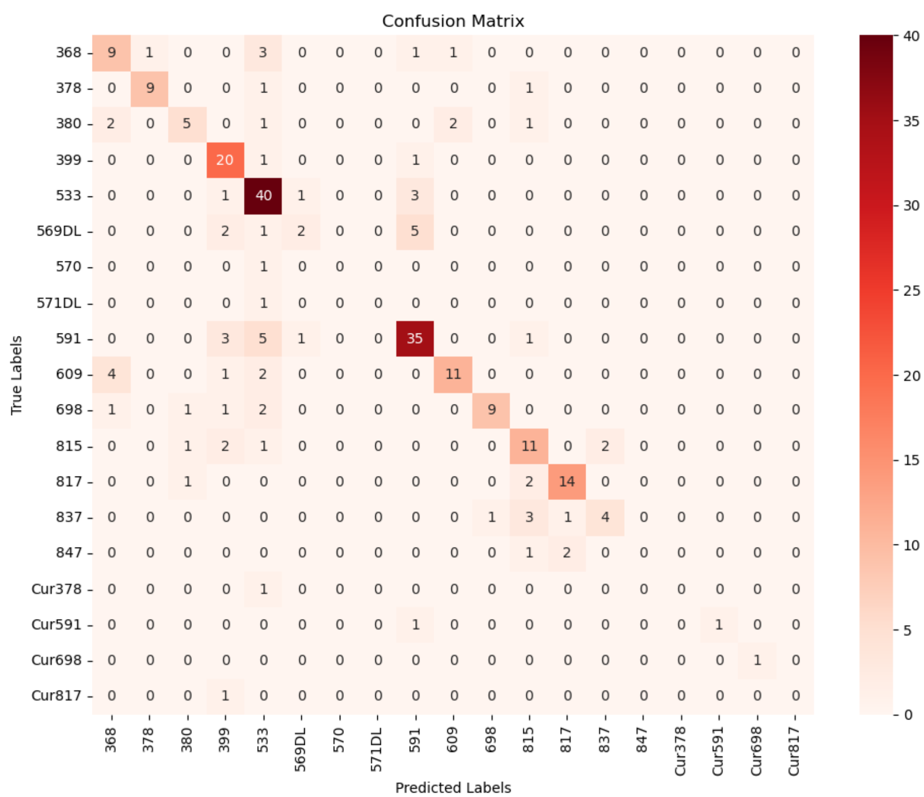

Figure 22. Max Entropy Confusion Matrix.

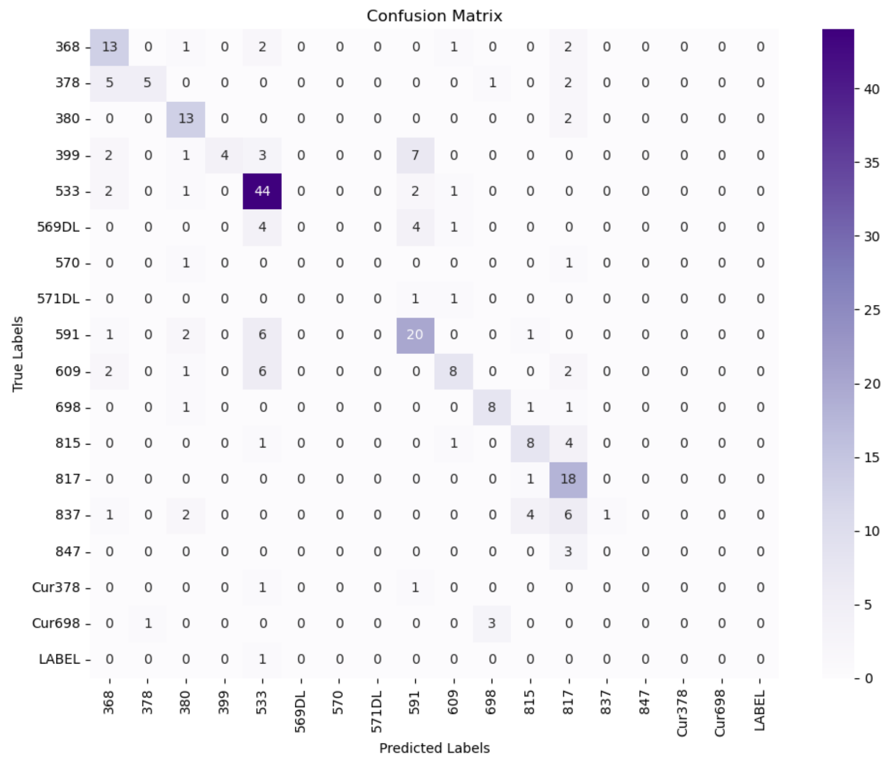

Figure 23. Naïve Bayes Confusion Matrix.

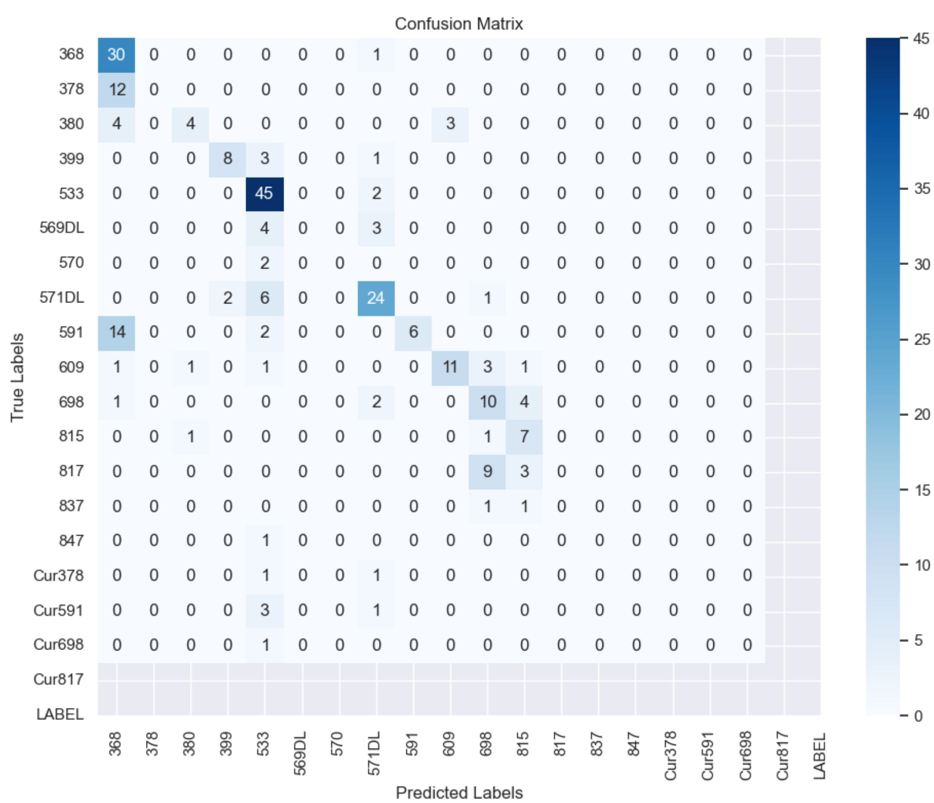

Figure 24. Transformers Confusion Matrix.

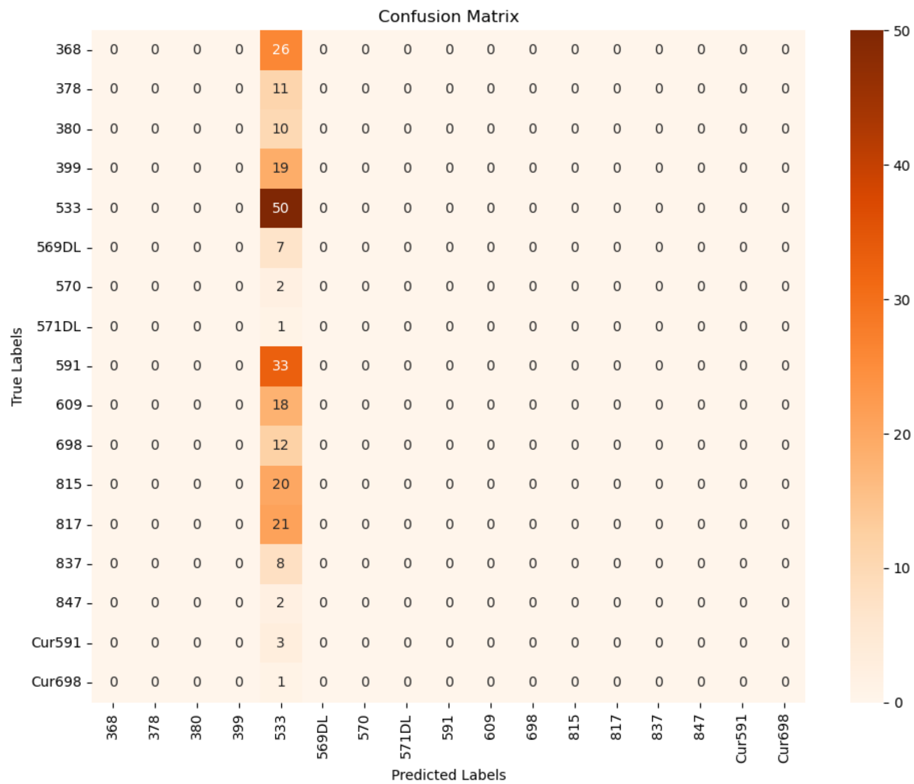

Figure 25. Dummy Classifier Confusion Matrix

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Research Sponsored Programs Office, Code 41
   Naval Postgraduate School
   Monterey, CA 93943

4. CAPT Hans Fosser and Mr. Kyle R Alcock
   OPNAV N711, Navy Education Strategy and Policy Branch
   Office of the Chief of Warfighting Development
   2000 Navy Pentagon, Room 5A268
   Washington, DC 20350-2000