Date Issued: 8/13/2019
Date Modified: 1/16/2020
Version: 1.3
Approved by: Scheme Committee

CertNexus Certified Artificial Intelligence Practitioner™ (CAIP) Exam AIP-110

Exam Information

Candidate Eligibility:

The *Certified Artificial Intelligence Practitioner™ (CAIP)* exam requires no application fee, supporting documentation, or other eligibility verification measures for you to be eligible to take the exam. An exam voucher will come bundled with your training program or can be purchased separately here. Once purchased, you will receive more information about how to register for and schedule your exam through Pearson Vue. You can also purchase a voucher directly through Pearson Vue. Once you have obtained your voucher information, you can register for an exam time here. By registering, you agree to our Candidate Agreement included here.

Exam Prerequisites

While there are no formal prerequisites to register for and schedule an exam, we strongly recommend you first possess the following knowledge and skills:

- Explain how artificial intelligence (AI) and machine learning (ML) can solve business problems.
- Execute an applied ML workflow.
- Summarize outcomes of accepted learning algorithms.
- Formulate mathematical representations of business problems using domain insight.
- Develop and test hypothesis using experimental design.
- Distinguish benefits and drawbacks of various machine learning models and given a scenario select appropriate model and define tradeoffs.
- Given a scenario, select appropriate tool sets (both proprietary and open source).
- Demonstrate responsibility based upon ethical implications when sharing data sources.
- Plan, manage, train and hand off an ML model as part of a (software) solution.
- Communicate the findings of an AI and ML workflow and solution back to the organization.
- Identify the impact that propagating biases has within AI.
- Select and implement an appropriate algorithm for a given business problem.
- Select and implement the appropriate techniques for a given ML problem.
- Demonstrate a working level knowledge of development tools such as Python and R.

You can obtain this level of skill and knowledge by taking the following courseware, which is available through training providers located around the world, or by attending an equivalent third-party training program:

- *AIBIZ™ (Exam AIZ-110)*
- *CertNexus Certified Artificial Intelligence (AI) Practitioner™ (Exam AIP-110)*

Exam Specifications

Number of Items: 80

Passing Score: 60%

Duration: 120 minutes (**Note**: exam time includes 5 minutes for reading and signing the Candidate Agreement and 5 minutes for the Pearson VUE testing system tutorial.)

Exam Options: In person at Pearson VUE test centers or online with Pearson OnVUE online proctoring

Item Formats: Multiple Choice/Multiple Response

Exam Description

Target Candidate:

This certification exam is designed for practitioners who are seeking to demonstrate a vendor-neutral, cross-industry skill set within AI and with a focus on ML that will enable them to design, implement, and hand off an AI solution or environment.

Exam Objective Statement:

This exam will certify that the candidate has the knowledge and skill set of AI concepts, technologies, and tools that will enable them to become a capable AI practitioner in a wide variety of AI-related job functions.

To ensure exam candidates possess the aforementioned skills, the *Certified Artificial Intelligence Practitioner™ (CAIP)* exam will test them on the following domains with the following weightings:

Domain	% of Examination
1.0 Applied Artificial Intelligence and Machine Learning in Business	5%
2.0 Problem Formulation	25%
3.0 Data Collection, Comprehension, Cleaning, and Engineering	20%
4.0 Algorithm Selection and Model Training	35%
5.0 Model Handoff	10%
6.0 Ethics and Oversight	5%
Total	**100%**

The information that follows is meant to help you prepare for your certification exam. This information does not represent an exhaustive list of all the concepts and skills that you may be tested on during your exam. The exam domains, identified previously and included in the objectives listing, represent the large content areas covered in the exam. The objectives within those domains represent the specific tasks associated with the job role(s) being tested. The information beyond the domains and objectives is meant to provide examples of the types of concepts, tools, skills, and abilities that relate to the corresponding domains and objectives. All of this information represents the industry-expert analysis of the job role(s) related to the certification and does not necessarily correlate one-to-one with the content covered in your training program or on your exam. We strongly recommend that you independently study to familiarize yourself with any concept identified here that was not explicitly covered in your training program or products.

Objectives:

Domain 1.0 Applied Artificial Intelligence and Machine Learning in Business

Objective 1.1 Identify and describe how artificial intelligence and machine learning are used to solve business problems.

- Business benefits of applied artificial intelligence and machine learning
- Relationship between machine learning and data science
    - ML is a subset of data science
- Data mining
- NLP
    - Speech recognition
    - Text analysis
- Computer Vision
    - Image recognition
    - Video tracking

Domain 2.0 Problem Formulation

Objective 2.1 Given a business problem, select an appropriate machine learning model and outcome.

- Supervised learning
    - Discrete classification
        - Unbalanced class distribution
        - Resampling
        - Class probability distributions
    - Continuous regression
        - Different types of regressions
            - Lasso
            - Ridge
            - Linear

        - Leave one out regression
        - Relationships between Lasso/Ridge and L1/L2 norms
    - Deep learning
- Unsupervised learning
    - Clustering
    - No labels present
    - Identifying groups or segments
    - Dimensionality reduction
- Semi-supervised
- Reinforcement learning

Objective 2.2 Use an experimental design approach to develop and test a hypothesis.

- Design of experiments
- Statement of hypotheses
    - AB testing
    - Alpha vs. Beta values
    - P values
    - Confidence intervals
- Understanding business KPIs
- Actionable insights over higher accuracy models
    - Precision vs. recall
        - F Score
        - Confusion matrix
- Utilizing judgement to recognize when to stop

Objective 2.3 Select appropriate tools to solve a given machine learning task.

- Open source AI tools
    - Tensorflow
        - NumPy
        - Keras
        - TensorBoard
    - PyTorch
    - NLTK
    - Scikit-learn
    - Pandas
    - Spark ML
        - Spark Core
        - Spark SQL
- Proprietary AI tools
    - Microsoft Azure AI tool
        - Azure Machine Learning
        - Azure Databricks
        - Azure Search
        - Cognitive Services
    - Amazon Web Services AI Services
        - Amazon SageMaker
        - ML Framework
        - AI Services

        - AWS DeepRacer
    - IBM Watson
        - Watson OpenScale
        - Watson Machine Learning
        - NLU
    - Google AI
        - ML Kit
        - DeepDream
        - Data Search
        - Cloud AI
        - Cloud AutoML

Objective 2.4 Evaluate the applicability of new AI technologies to perform a given business task.

- Process to evaluate the applicability of new technologies
- Future innovations
    - Distributed artificial intelligence
    - Hyper-heuristic
    - Federated learning

Domain 3.0 Data Collection, Comprehension, Cleaning, and Engineering

Objective 3.1 Collect and prepare a dataset to use for training and testing.

- Sources of data
    - Big data/data sets
        - Volume, variety, velocity
        - Data repositories
        - Data prep
        - Stream data
- Structure of data
- Extract, transform, and load (ETL)

Objective 3.2 Analyze a dataset to gain insights.

- Visualization
- Correlations between attributes
- Attribute characteristics
- Descriptive statistics applied to features
    - Standard deviation
    - Kurtosis
    - Mean, mode, median
    - Variance
    - Column distribution
- Classic data sets
    - Time Series
    - Bayesian

Objective 3.3 Clean data in preparation for use in machine learning.

- Data backups
- Data cleansing
- Typecasting
- Operations appropriate for different data types

- Data encryption

Objective 3.4 Engineer features of a dataset to prepare it for use in a machine learning model.

- Feature transformation
- Dimensionality reduction
- Center and spread measures
- Imputing missing values
- Duplicates
- Data binning
- Normalization and standardization
- String manipulation
- Summarization
- Embedded spaces

Domain 4.0 Algorithm Selection and Model Training

Objective 4.1 Select and implement an appropriate algorithm to solve a given business problem.

- Decision Tree
  - Hyperparameters
    - Maximum depth
    - Minimum sample per leaf node
  - Number of random splits
  - Information gain/entropy measure
  - GINI Index
  - Continuous variable discretization
  - Random Forest
    - Hyperparameters
      - Number of trees
      - Maximum depth
      - Minimum samples required to split node
      - Minimum samples required to be at leaf node
    - Variable selection
    - Assumption
    - Sampling is representative
- SVM
  - Kernels
    - RBF
    - Gaussian
    - Polynomial
    - Linear kernel
    - Sigmoid
    - TanH
  - Can tackle non-linearly separable data
  - Uses hyperplanes for separation
- Neural networks
  - ANN
    - Layers of a traditional network

                - Input
                - Hidden
                - Output
            - Activation functions
                - Tanh
                - Sigmoid
                - ReLU
        - CNN
            - Image processing
        - RNN
            - Natural language
            - LSTM
                - Natural language
                - Forecasting
        - GAN
            - Adversarial content creation
- Clustering
    - K-means clustering
        - Issues clustering circular data
        - Methods for deciding K-means
            - Elbow
            - Silhouette plot
- Classification
    - Logistic regression
    - Softmax regression
- Regression
    - Linear regression
- KNN
    - How do you decide the K?
    - Difference between K-Means and KNN

Objective 4.2 Select and implement the appropriate techniques for a given machine learning problem.

- Dimensionality reduction
    - Methods for DR
        - PCA
        - Lasso
        - Random Forest
- Feature engineering
    - When feature engineering is useful
    - Benefits of deriving custom features
- Feature expansion
    - Benefits to tackling nonlinear data
    - Methods of feature expansion
- Hyper-parameter optimization
    - Awareness of core parameters for key models

        - Model tuning
        - Types of hyperparameter searches
            - Random
            - Grid
            - Bayesian
            - Genetic
    - Cross validation
        - Leave 1 out cross validation
            - Used for small data sets
        - N Fold cross validation
    - Regularization
        - Types
            - L1
            - L2
        - Know the differences between L1 and L2
    - Variance vs. bias
        - The variance bias tradeoff
    - Model generalization
        - Overfitting
        - Use regularization
        - Relationship with variance and bias
    - Embedded spaces
        - Embedded space extraction from CNN
        - Word vector open source tools
            - Fasttext
            - word2vec
            - doc2vec
    - Data sets
        - Structured
        - Unstructured

Objective 4.3 Manage the time needed to train a model.

- Estimating the time needed to run a batch over a certain number of epochs
- Optimizing the development environment (e.g., scaling up GPU) vs. making compromises in the model to reduce processing time
- Awareness around processing costs

Objective 4.4 Train and tune a machine learning model.

- Model performance evaluation
    - Confusion matrix
    - Classifier performance measurement
    - Accuracy
    - Precision
    - Recall
    - Precision and recall tradeoff
    - F1 Score
    - ROC Curve
    - Thresholds

        - AUC
        - PRC
    - Cross validation
    - Model generalization
    - Performance tuning

Domain 5.0 Model Handoff

Objective 5.1 Communicate the findings of a machine learning project back to the organization.

- Translating ML results into potential business actions
    - Prediction or classification problems
- Data visualization
    - Big data
    - Internet of things
    - Categorical
    - Quantitative
    - Types
        - Table
        - Graph
- Explain accuracy vs precision vs recall to non-ML practitioners
    - True positives
    - False positives
- Compromising on accuracy for model interpretability

Domain 6.0 Ethics and Oversight

Objective 6.1 Identify and describe the impact that propagating biases has within AI.

- Avoiding biases in data
    - Current data sources
    - Model explainability
    - Transparency
- Preventing propagation of preconceived notions
    - Open review of data sets and algorithmic approaches
- Recognizing that proxies may be indicative of larger social discriminations

Objective 6.2 Comply with laws and standards that are applicable to businesses that employ AI.

- Relevant data privacy laws
    - GDPR
    - California Consumer Privacy Act

Objective 6.3 Given a use case, identify and describe the ethical issues on sharing the data sources.

- The importance of Open Source Data Access and data integrity
- Respecting the privacy of data sources and targets
- Guidelines for protecting privacy when collecting, storing, and disposing of data

Objective 6.4 Comply with company policies to promote privacy, security, and ethical practices.

- Potential ethical issues resulting from AI practices
    - Recognizing that the output of AI even though derived can still infringe on privacy/copyright/intellectual property rights
    - Recognizing the importance of basic humanitarian principles before developing any AI capabilities
    - Ability to identify alternative uses of AI you create by bad actors

Recertification Requirements

The *Certified Artificial Intelligence Practitioner (CAIP)* certification is valid for 3 years from the time certification is granted. You must re-take the most up-to-date version of the exam prior to the 3-year period's end to maintain a continuously valid certification.

Certified Artificial Intelligence Practitioner (CAIP) Acronyms

Acronym	Expanded Form
AI	artificial intelligence
ANN	artificial neural network
AUC	Area Under the Curve
CNN	convolutional neural network
DR	dimensionality reduction
ETL	exact, transform, and load
GAN	generative adversarial network
GDPR	General Data Protection Regulation
KNN	K-nearest neighbors
KPI	key performance indicator
LSTM	Long Short-Term Memory
ML	machine learning
NLP	natural language processing
NLTK	Natural Language Toolkit
NLU	natural language understanding
PCA	principal component analysis
PRC	Precision-Recall Curve
RBF	radial basis function kernel
ROC	Receiver Operating Characteristic
RNN	recurrent neural network
SVM	Support Vector Machine