Azure AI-900 Artificial Intelligence (AI) Cheat sheet v2024.11.01

Willy's Notes

Translation
Conversion
Predictions

Guiding principles

FRAUPIT

- **Accountability** – People should be accountable for AI systems
- **Inclusivity** – Empower everyone and Engage people
- **Transparency** – Understandable decisions
- **Privacy & Security** – Be secure and respect privacy
- **Fairness** – Treat all people fairly
- **Reliability** – Perform reliably and Safely
- **Unintended Consequences** – Bias, Wrong, Illegal, Hallucinations, Harmful

AI Fairness Checklist

Seeing AI

Common Workloads

- **Anomaly Detection** – Is this weird?
- **Content Moderation** – Violence, Self-Harm, Sexual, Hate
- **Prediction & Forecasting**
- **Computer Vision**
- **Knowledge Mining** – AI S/W Navigator
- **Natural Language Processing (NLP)** – What info is in? – AI S/W Sentinel

Entra ID RBAC — Key (auth)

ENDPOINT
- REST
- SDK
- Region

Vision Workloads

- **Facial detection & recognition**
- **Optical Character Recognition** – Detect text
- **Image Classification** – What is image? A car!
- **Semantic Segmentation** – Detect object boundaries
- **Object Detection** – Detect multiple objects on image
- Tools

Cognitive Services
- Umbrella service
- Includes many other services

Computer Vision Services
- Pre-trained LM model
- Can **recognize** and **categorize** 10,000 objects
- Detect faces and text recognition

Custom Vision Services
- Model you can build and train
- Classification or object detection (**class name**, **probability score**, **bounding box**)

Face Services
- Recognize a human face on an image
- Returns coordinates of 1+ faces
- Recognizes pre-trained celebrities

Face Detection
- Gender
- Age
- Face boundaries

Face Verification
- Compare faces
- Identify if same individual

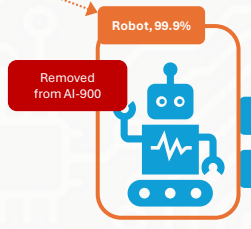
Similar Faces
- Identify similar faces

Face Grouping
- Divides a set of unknown faces into groups based on similarities

Identify API
- Identify (assign names) to faces

Form Recognizer Service
- Ideal for invoices and receipts
- **Pre-built** or **custom** model
- Pre-built = English in USA

AI Video Indexer
- Build on Az Media Services and Azure AI Services (**Face**, Translator, **Vision**, Speech)
- Allows extraction of insights from videos



Agents or Bots
- Conversational AI
- Computer and humans talk
- Operate over web, email, social media, and voice

Conversational Services
- QnA Maker Service – codeless chatbot
  Azure Bot Service – Enterprise grade

NLP Workloads

Understand written or spoken language

- **Key Phrase Extraction** – Key points in document
- **Speech Recognition & Synthesis** – Speech -> Text, Text -> Speech
- **Entity recognition** – Categorize by type and subtype; Today = date; Driveway = location
- **Language Modelling** – Common Language
- **Sentiment Analysis** – Prebuilt model 0-1 (neg-pos)
- **Translation** – 60+ Languages, Context aware
- Services

Text Analysis Service
- Key phrase extraction
- Entity detection
- Sentiment analysis (positive=1,neutral=0.5, negative=0)
- Language detection
- Returns NaN for undetermined

Language Understanding Service (LUIS)
- Understand natural language
- **Utterances** – *when do you close?*
- **Entities** – *today (date time)*
- **Intents** – *goal: total hours of operation*
- No intent – *what is the meaning of life?*

Speech Service
- Text to speech
- Speech to text
- Auto translation
- Custom voice models

Translator Text Service
- Translate between 70 languages
- Custom models
- Language detection
  - ISO 6391 Code
  - Language Name
  - Score

Generative AI Workloads

- **Others** – Discovery, Gaming, Security, Simulations
- **Image Creation** – Art, Impressionism, Face
- **Text Generation** – ChatGPT, Copilot
- **Video Generation** – Deep Fakes, Animation
- **Code Generation**
- **Music, Audio Generation** – Voice synthesis, Sound effects
- Responsible Generative AI

- Copilot Studio (low code)
- Azure AI Studion (PaaS)

Dangers
- Deep fakes
- Impersonation
- Harmful
- Discriminating

Process
1. Identify harm – *what is worst case scenario?*
2. Measure harms – *try to generate harmful content*
3. Mitigate harms - *fine-tune model and filters, meta prompts, grounding data*
4. Operate responsibly - *monitor, feedback, track telemetry*

AI Search
- skillset → Data
- Indexes → Knowledge Store, Text, Vector

Azure Portal → All Services → AI + Machine Learning

Feature
- Input variable
- Life insurance example: Smoker, Gender

Label
- Thing we predict
- Life insurance example: Life expectancy

Training
- Split date 50:50
- Training : Actual
- Compare actual to predicted
- Mean Square Error

Azure Portal → All Services → AI + Machine Learning → AI Machine Learning
1. New workspace
2. Launch studio
3. Create compute "instance" to work
4. Create compute "cluster" for ML training
5. Automate ML (no-code option) or ML Designer (low-code option)

Tools
- Notebooks
- Automated ML
- Designer

Compute
- Compute Instances
- Compute Clusters
- k8s Clusters
- Attached Compute
  - Managed outside of Azure
  - Managed yourself

Common Machine Learning Types

https://aka.ms/mlcheatsheet

- Common Types
- Experiment
- **Deep learning** – Neural networks, Large dataset
- **Clustering** – Unsupervised, unlabeled data; Find **similar** groups; How is this organized?
- **Classification** – Supervised, Historical data, **Binary**; Is it an Apple?; Is it A, B, or C?; Is it a ...?
- **Regression** – Supervised, Historical data, **Numeric**; How much?; How many?

X Sentiments

Marketing campaigns

Weather predictions

Metrics
- Average Distance to Other Center
- Average Distance to Cluster Center
- Combined Evaluation
- Maximal Distance to Cluster Center
- Number of Points

General
- Voting Ensembles combines predictions of several models.
- Model Interpretability helps understand models.
- Block Algorithms specifies one or more algorithms not to use.
- Automate model retraining with Azure ML Pipelines.

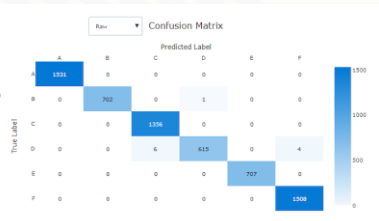
Metrics
- Coefficient of determination
- Mean Absolute error (MAE)
- Root mean squared error (RMSE)
- Relative Absolute error (RAE)
- Relative Squared error (RSE)

Models
- **Random Forest** - mix of categorical and numerical variables.
- **Lasso Regression** is a variant if Linear Regression.
- **Naïve Bayes** is a classification algorithm
- **Poisson regression** is used for modeling count variables.

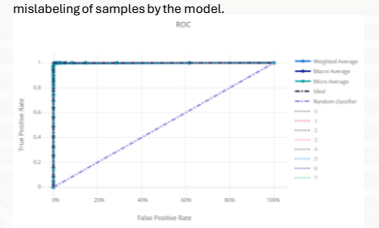
Training
- Bring/Gather Data
- Transform Data
  - Feature Select
  - Find &Clean outliers
  - Impute missing values
  - Normalize

Transfer Learning
- Leverage pre-trained model.

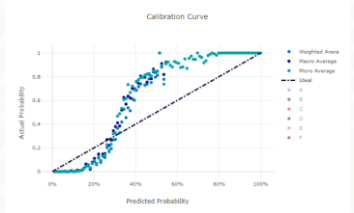
Metrics
- Accuracy [0,1] (! for AI Vision)
- Area Under curve (AUC) [0,1]
  - AUC=1 is best fitted model.
  - AUC<0.5 is worse than random
- Confusion Matrix
- F1 Score
- Precision [0,1]
- Recall [0,1]
- Selectivity [0,1]
- Weighted accuracy [0,1]



Confusion matrices visually demonstrate the systematic errors of a classification model, with "confusion" referring to the mislabeling of samples by the model.



The **receiver operating characteristic** (ROC) curve illustrates the connection between the true positive rate (TPR) and the false positive rate (FPR) as the decision threshold is varied. The area under the curve (AUC) represents the fraction of correctly classified instances.

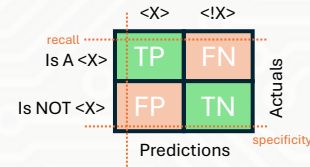


A **calibration curve** shows a model's confidence in predictions versus the proportion of correct positive samples at each confidence level. An ideal model accurately classifies 100% of predictions with 100% confidence, 50% with 50% confidence, and 20% with 20% confidence. A perfectly calibrated model's curve matches the y = x line, where probabilities are predicted flawlessly.

Images from Azure Machine Learning | Microsoft Learn

experiment results - Azure Machine Learning | Microsoft Learn

Confusion Matrix




```
Accuracy    = (TP+TN)/(TP+TN+FP+FN)
Recall      = TP/(TP+FN)
Precision   = TP/(TP+FP)
Specificity = (TN/(FP+TN)
F1 Score    = (2*Recall*Precision)/
              (Recall + Precision)
```