

# AI & EHR Course Glossary

| Key Term  | Definition   |
|---|--|
| <b>Aleatoric Uncertainty</b>                        | Otherwise known as statistical uncertainty and are known unknowns. This type of uncertainty is inherent and just a part of the stochasticity that naturally exists.  |
| <b>AUC</b>  | Area under the ROC Curve measures the entire two-dimensional area underneath the entire ROC curve.   |
| <b>Black Box</b>                                    | Term often used in software when inputs go in and outputs come out of an algorithm and it is not clear how the outputs were arrived at.  |
| <b>Business Associates Agreement/Addendum (BAA)</b> | this is the contract between a covered entity and BA   |
| <b>Business Associates</b>                          | A business associate is a person or entity that performs certain functions or activities that involve the use or disclosure of protected health information on behalf of, or provides services to, a covered entity. |
| <b>Cardinality</b>                                  | refers to the number of unique values that a feature has and is relevant to EHR datasets because there are code sets such as diagnosis codes in the order of tens of thousands of unique codes.                      |
| <b>CCS</b>  | Clinical Classifications Software used to map diagnosis or procedure codes from ICD code sets.   |
| <b>Covered Entities</b>                             | are a group of industry organizations defined by HIPAA to be one of three groups health insurance plans, providers, or clearinghouses. You can see from the table the types of entities in each category.            |
| <b>CRISP-DM</b>                                     | This stands for “cross-industry standard process for data mining” and is a common framework used for data science projects and includes a series of steps from business understanding to deployment.                 |
| <b>Crosswalk</b>                                    | A connection between two different code sets or versions of drugs in the same code set.  |
| <b>Data Leakage</b>                                 | Inadvertently sharing data between test and training datasets.   |
| <b>De-identifying a Dataset</b>                     | refers to the removal of identifying fields like name, address from a dataset. De-Identification is done to reduce privacy risks to individuals and support the secondary use of data for research and such.         |
| <b>DPA</b>  | The Data Protection Act really builds off of and add to GDPR   |
| <b>EDA</b>  | Exploratory Data Analysis  |
| <b>EHR Data</b>                                     | is the data being collected when we see a doctor, pick up a prescription at the pharmacy, or even from a visit to the dentist.   |
| <b>EHR</b>  | Electronic Medical Record  |
| <b>EMR</b>  | Electronic Medical Record  |
| <b>Encounter Level</b>                              | Also known as the visit level, which is the aggregated information from the previously mentioned line level for one encounter. This information can be collapsed into a single row or arrays.                        |
| <b>Encounter</b>                                    | An interaction between a patient and healthcare provider(s) for the purpose of providing healthcare service(s) or assessing the health status of a patient.  |
| <b>Epistemic Uncertainty</b>                        | Also known as systemic uncertainty and are unknown unknowns. This type of uncertainty can be improved by adding parameters/features that might measure   |

|                                    |   |
|------------------------------------|---|
|                                    | something in more detail or provide more knowledge.   |
| <b>EU</b>                          | European Union  |
| <b>Expert Determination Method</b> | Completed by a statistician to determine there is a small enough risk that an individual could be identified.   |
| <b>F1</b>                          | Harmonic mean between precision and recall  |
| <b>GDPR</b>                        | The General Data Protection Regulation is generally considered more stringent than even HIPAA when it comes to protections for patients.  |
| <b>Health Record</b>               | A patient's documentation of their healthcare encounters and the data created by the encounters across time   |
| <b>HIPAA</b>                       | Health Insurance Portability and Accountability Act   |
| <b>HIPAA</b>                       | The Health Insurance Portability and Accountability Act is the key industry regulation that you should be familiar with within the U.S.   |
| <b>HITECH</b>                      | The Health Information Technology for Economic and Clinical Health Act is also important to note and this is really just an update to HIPAA that accounts for technology.         |
| <b>ICD10-CM</b>                    | International Classification of Diseases 10 - Clinical Modification   |
| <b>ICD10</b>                       | International Classification of Diseases 10   |
| <b>Line Level</b>                  | A denormalized representation of all the things that might happen in a medical visit or encounter.  |
| <b>Longitudinal Level</b>          | Also known as the patient level view. This level aggregates the patient history and can show how the culmination of visits/encounter lead to some clinical impact.                |
| <b>MAE</b>                         | Mean Absolute Error is a measure of errors between paired observations.   |
| <b>MAPE</b>                        | Mean Absolute Percentage Error is a measure of quality for regression models loss.  |
| <b>MAR</b>                         | Missing at Random and this is the opposite case where there <b>**is some systematic relationship**</b> between data and the probability of missing data.                          |
| <b>MCAR</b>                        | Missing Completely at Random. This means that the data is missing due to something unrelated to the data and there is <b>**no systematic reason**</b> for the missing data.       |
| <b>Medical Encounter</b>           | An interaction between a patient and healthcare provider(s) for the purpose of providing healthcare service(s) or assessing the health status of a patient.                       |
| <b>Medical Procedure</b>           | a course of action to achieve an intended result for a patient in healthcare.   |
| <b>MNAR</b>                        | Missing Not at Random and this usually means there is a relationship between a value in the dataset and the missing values.   |
| <b>Model Agnostic</b>              | methods that can be used on deep learning models or traditional ml models   |
| <b>MS-DRG</b>                      | Medicare Severity-Diagnosis Related Group   |
| <b>NDC</b>                         | National Drug Code  |
| <b>P(A   Evidence)</b>             | Updating the prior with new evidence  |
| <b>P(B)</b>                        | posterior probability   |
| <b>Payers</b>                      | a group that consists of companies like healthcare insurance. Payers can also include entities such as the government for things like Medicaid and Medicare in the United States. |
| <b>PHI</b>                         | Protected Health Information  |
| <b>Precision</b>                   | The fraction of relevant instances among the retrieved instances  |
| <b>Primary Diagnosis Code</b>      | The code that takes up the most resources to treat.   |

|                                  |   |
|----------------------------------|---|
| <b>Principal Diagnosis Code</b>  | The diagnosis that is found after hospitalization to be the one that is chiefly responsible.  |
| <b>Prior distribution - P(A)</b> |   |
| <b>Procedure Codes</b>           | The categorization of these medical codes during an encounter.  |
| <b>Providers</b>                 | These are entities/groups/individuals that provide care for patients. Providers can range in size from entire hospital networks to individual doctors and medical professionals.                        |
| <b>Recall</b>                    | The fraction of the total amount of relevant instances that were actually retrieved.  |
| <b>Representative Splitting</b>  | Having accurate labels and demographics in your data splits that reflect the real world.  |
| <b>RMSE</b>                      | Root Mean Square Error- a measure of the differences between values predicted by a model.   |
| <b>ROC</b>                       | Receiver Operating Characteristic Curve or ROC curve that shows a graph of the performance of a classification model. It is the True Positive Rate Vs. False Positive Rate across different thresholds. |
| <b>RXNorm</b>                    | which does what its name implies and groups medication together.  |
| <b>Safe Harbor</b>               | The removal of 18 identifiers like name, zip code, etc.   |
| <b>Secondary Diagnosis Codes</b> | The other diagnosis codes listed on an encounter.   |
| <b>SNOMED CT</b>                 | Systematized Nomenclature of Medicine—Clinical Terms  |
| <b>Unintended Biases</b>         | A bias that is not intentional and often is not even apparent to the creator of a model   |
| <b>WHO</b>                       | World Health Organization   |