Technical Glossary

Key Terms and Definitions for AI-Driven Design Thinking

Week 1 Appendix

Machine Learning Fundamentals

- **Machine Learning (ML)**: Computer systems that learn patterns from data without explicit programming. Like teaching by examples rather than rules.
- **Algorithm**: Step-by-step procedure for solving a problem. Think of it as a recipe that computers follow.
- **Training**: Process of teaching an ML model by showing it examples. Like a student learning from practice problems.
- **Model**: The learned representation that can make predictions. Like a mental model built from experience.
- **Pattern Recognition**: Identifying regularities in data. Like recognizing faces - you know one when you see it.
- **Feature**: A measurable property of data. Like height, weight, or color in describing a person.

Types of Learning

- **Supervised Learning**: Learning with labeled examples. Like learning with an answer key.
  - Example: Email spam detection (spam/not spam labels)
- **Unsupervised Learning**: Finding patterns without labels. Like organizing books without categories.
  - Example: Customer segmentation without predefined groups
- **Classification**: Predicting categories. Is this email spam? (Yes/No)
- **Regression**: Predicting continuous values. What will the temperature be? (73.5°F)
- **Clustering**: Grouping similar items together. Like sorting laundry by color.

Natural Language Processing (NLP) - Part 1

- **NLP**: Teaching computers to understand human language. Like Google Translate or Siri.
- **Tokenization**: Breaking text into words or pieces. "Hello world" $\rightarrow$ ["Hello", "world"]
- **Sentiment Analysis**: Determining emotional tone. Is this review positive or negative?
- **Topic Modeling**: Discovering themes in text automatically. Finding what documents are about.
- **Named Entity Recognition (NER)**: Finding names, places, dates in text. "Apple Inc." = Company
- **Stemming**: Reducing words to root form. "running", "runs", "ran" $\rightarrow$ "run"

Natural Language Processing (NLP) - Part 2

- **Lemmatization**: Smarter stemming using dictionary. "better" → "good"
- **Stop Words**: Common words often removed. "the", "is", "at", "which"
- **TF-IDF**: Term Frequency-Inverse Document Frequency. Measures word importance in documents.
- **Corpus**: Collection of text documents. Your dataset of text.
- **N-gram**: Sequence of N words. "New York" is a 2-gram (bigram).
- **Bag of Words**: Treating text as unordered word collection. Ignoring grammar, keeping frequency.

Advanced Machine Learning Concepts

- **Neural Network**: ML model inspired by brain structure. Networks of connected nodes.
- **Deep Learning**: Neural networks with many layers. "Deep" = many layers.
- **Embedding**: Converting words/items to numbers. "Cat" $\rightarrow$ [0.2, -0.5, 0.8, ...]
- **Attention Mechanism**: Model focuses on relevant parts. Like highlighting important text.
- **Transformer**: Architecture behind ChatGPT. Processes all words simultaneously.
- **Fine-tuning**: Adapting pre-trained model to specific task. Like specializing general knowledge.

Statistical Foundations

- **Mean/Average**: Sum divided by count. Central tendency measure.
- **Variance**: How spread out data is. High variance = more spread.
- **Standard Deviation**: Square root of variance. Average distance from mean.
- **Correlation**: Relationship between variables. Height and weight often correlate.
- **Distribution**: How data spreads across values. Bell curve is normal distribution.
- **Entropy**: Measure of uncertainty/chaos. High entropy = more random.

Clustering Algorithms

- **K-means**: Divides data into K groups. You specify number of clusters.
  - Example: Segment customers into 5 groups
- **Hierarchical Clustering**: Builds tree of clusters. Can see relationships at different levels.
- **DBSCAN**: Density-based clustering. Finds arbitrary shaped clusters.
- **Centroid**: Center point of cluster. Average position of all points.
- **Silhouette Score**: Measures how well clustered data is. Higher = better separation.

Generative AI and LLMs

- **Generative AI (GenAI)**: AI that creates new content. ChatGPT, DALL-E, etc.
- **Large Language Model (LLM)**: AI trained on vast text. GPT, Claude, BERT.
- **Prompt**: Input text given to AI. Your question or instruction.
- **Token**: Piece of text (word or part). "Tokenization" $\rightarrow$ ["Token", "ization"]
- **Temperature**: Controls AI creativity. Low = predictable, High = creative.
- **Context Window**: How much text AI can process. Like working memory.

Specific Algorithms Explained

- **LDA (Latent Dirichlet Allocation)**: Finds hidden topics in documents. Like auto-tagging articles.
- **BERT**: Bidirectional Encoder Representations from Transformers. Understands context both ways.
- **GPT**: Generative Pre-trained Transformer. Predicts next word repeatedly.
- **Word2Vec**: Converts words to vectors. Similar words have similar vectors.
- **Random Forest**: Many decision trees voting together. Wisdom of crowds for ML.
- **Gradient Descent**: How models learn. Like finding lowest point in valley.

Model Evaluation Terms

- **Accuracy**: Percentage correct predictions. 95% = 95 out of 100 correct.
- **Precision**: Of predicted positives, how many correct? Quality over quantity.
- **Recall**: Of actual positives, how many found? Finding all relevant items.
- **F1 Score**: Balance of precision and recall. Harmonic mean of both.
- **Cross-validation**: Testing model on different data splits. Ensures robust performance.
- **Overfitting**: Model memorizes training data. Works great on training, poorly on new data.

Design Thinking Integration

- **Empathy**: Understanding user feelings and needs. Foundation of design thinking.
- **Persona**: Fictional user representing a segment. "Sarah, 34, working mother"
- **User Journey**: Path user takes to achieve goal. All touchpoints and experiences.
- **Pain Point**: Problem causing user frustration. Opportunity for improvement.
- **Ideation**: Generating solution ideas. Brainstorming phase.
- **Prototype**: Early version to test ideas. Minimum viable solution.

Data Processing and Preparation

- **Data Cleaning**: Fixing errors, removing duplicates. Making data usable.
- **Normalization**: Scaling values to standard range. 0-100 $\rightarrow$ 0-1.
- **Feature Engineering**: Creating new features from existing. Combining age and income.
- **Pipeline**: Sequence of data processing steps. Input $\rightarrow$ Process $\rightarrow$ Output.
- **Batch Processing**: Processing data in groups. Opposite of real-time.
- **API**: Application Programming Interface. How programs talk to each other.

Advanced Concepts

- **Hyperparameter**: Setting that controls learning. Like learning rate or tree depth.
- **Ensemble Method**: Combining multiple models. Like getting second opinions.
- **Transfer Learning**: Using knowledge from one task for another. Like applying skills.
- **Reinforcement Learning**: Learning through trial and reward. How games learn to play.
- **Federated Learning**: Training on distributed data. Privacy-preserving ML.
- **Edge Computing**: Processing near data source. Phone ML vs cloud ML.

Privacy and Ethics Terms

- **Bias**: Systematic unfairness in predictions. Historical biases in data.
- **Fairness Metrics**: Measures of equal treatment. Demographic parity, equal opportunity.
- **Differential Privacy**: Adding noise to protect individuals. Privacy with utility.
- **SHAP**: SHapley Additive exPlanations. Explains model decisions.
- **LIME**: Local Interpretable Model-agnostic Explanations. Why did model decide this?
- **Explainable AI**: Models that can explain decisions. Transparency in AI.

Scale and Performance Concepts

- **Big Data**: Datasets too large for traditional tools. Millions/billions of records.
- **Scalability**: Ability to handle growing data. Works for 10 or 10 million.
- **Real-time Processing**: Immediate data analysis. Results in milliseconds.
- **Latency**: Delay between request and response. Lower is better.
- **Throughput**: Amount processed per time. Transactions per second.
- **Distributed Computing**: Using multiple computers together. Divide and conquer.

Common Mathematical Notation

- $\sum$ **(Sigma)**: Sum notation. $\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$
- μ **(Mu)**: Population mean (average)
- σ **(Sigma)**: Standard deviation
- $p(x)$: Probability of x occurring
- log: Logarithm. Inverse of exponential
- ∇ **(Nabla)**: Gradient. Direction of steepest increase

Quick Reference Guide

Most Important Terms:

- ML = Machine Learning
- AI = Artificial Intelligence
- NLP = Natural Language Processing
- LLM = Large Language Model
- API = Application Programming Interface

Key Concepts to Remember:

- Clustering = Grouping similar items
- Classification = Predicting categories
- Regression = Predicting numbers
- Training = Teaching the model
- Inference = Using the model

Remember: Every technical term is just a name for a simple concept!