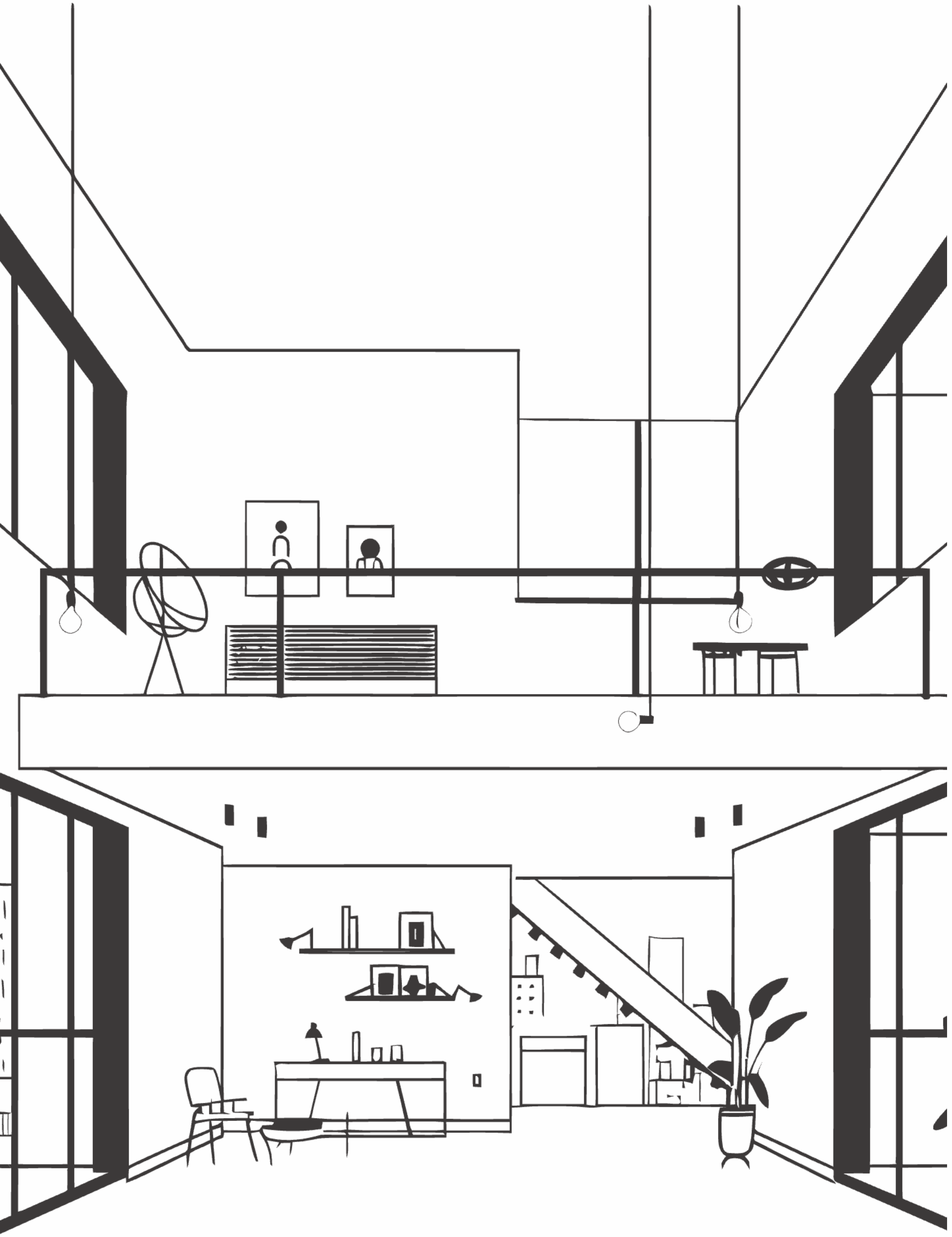


# Fashion-MNIST Clustering using Unsupervised learning

**Unsupervised Learning Final Project**

**Exploring pattern discovery in fashion images without labeled data**

CU Boulder 2025

Presented by Kimi Doan

# About me

## Kimi Doan

Chief Innovation Officer, Earable Neuroscience

Kimi leads innovation at Earable Neuroscience, a 3x CES Innovation Awards winner (2023-2025) pioneering AI wearables and digital therapeutics based on brainwave technology.

## Background & Expertise

- 15+ years in global tech leadership across marketing, business development, and strategic partnerships
- Tech Evangelist and Business Connector with deep expertise in computer science and neuroscience
- Former Global Chief Marketing Officer at VinFast EV (NASDAQ: VFS, 2020-2022)

## Education

- MSc in Computer Science, University of Colorado Boulder (Expected Graduation 2026)
- MBA (First-Class Honours), University of Gloucestershire, UK, 2011
- BSc in Computer Science and Telecommunications, Helsinki University of Technology, 2010

## Research Interests

Applied AI and neuroscience therapy approach to enhance longevity and unlock human potential.

## Personal Interests

Fashion and branding

# The Problem: Discovering Patterns Without Labels

## Challenge

How can we automatically group similar fashion items when we don't have category labels? This is unsupervised clustering of Fashion-MNIST images - discovering natural groupings without supervision.

Unlike supervised learning where we train with known categories, our models must identify meaningful patterns independently, mimicking how humans naturally recognize similar items.

## Real-World Applications

- **Fashion Recommendation Systems**: Suggesting similar items based on visual similarity
- **Inventory Management**: Automatically categorizing new products
- **Trend Analysis**: Identifying emerging fashion patterns and styles
- **Visual Search**: Finding products that match customer preferences

# Why Unsupervised Learning?

### No Labels Required

Works with unlabeled data, eliminating expensive manual annotation. Perfect when labeling thousands of items isn't feasible.

### Hidden Patterns

Discovers patterns and relationships that might not be obvious, revealing natural groupings in data.

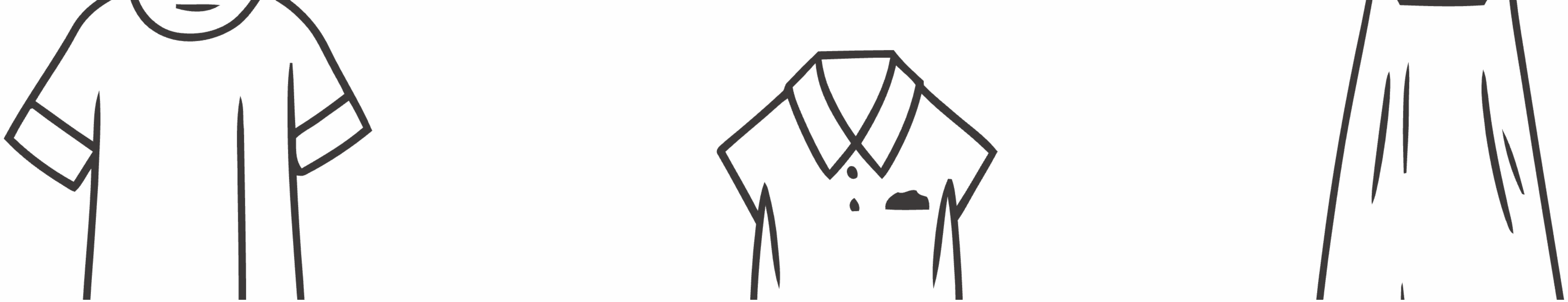
### Cost Effective

Saves time and resources when labeling is expensive or requires domain expertise from fashion specialists.

### Scalability

Can handle massive unlabeled datasets that would be impractical to manually categorize.

In fashion retail, where new products arrive constantly and visual similarity matters more than rigid categories, unsupervised learning provides a scalable solution for organizing and understanding inventory.



# Dataset Overview: Fashion-MNIST

## Dataset Statistics

- **60,000 training images** for learning representations
- **10,000 test images** for evaluating clustering
- **28×28 grayscale images** (784 pixel features)
- **10 fashion categories** from Zalando's catalog
- **Perfectly balanced** with 10% per class

**Source:** Fashion-MNIST was created by Zalando Research as a drop-in replacement for the traditional MNIST handwritten digits dataset, providing a more challenging benchmark for machine learning algorithms.

## Why Fashion-MNIST?

This dataset is ideal for clustering research because:

1. More complex than MNIST digits
2. Real-world fashion application
3. Standardized benchmark
4. Perfect class balance
5. Manageable computational requirements

# The Ten Fashion Categories

01

## T-shirt/top

**Short-sleeved casual tops**

02

## Trouser

**Pants and trousers**

03

## Pullover

**Sweaters without buttons**

04

## Dress

**One-piece garments**

05

## Coat

**Outerwear jackets**

06

## Sandal

**Open-toe footwear**

07

## Shirt

**Button-up tops**

08

## Sneaker

**Athletic shoes**

09

## Bag

**Handbags and purses**

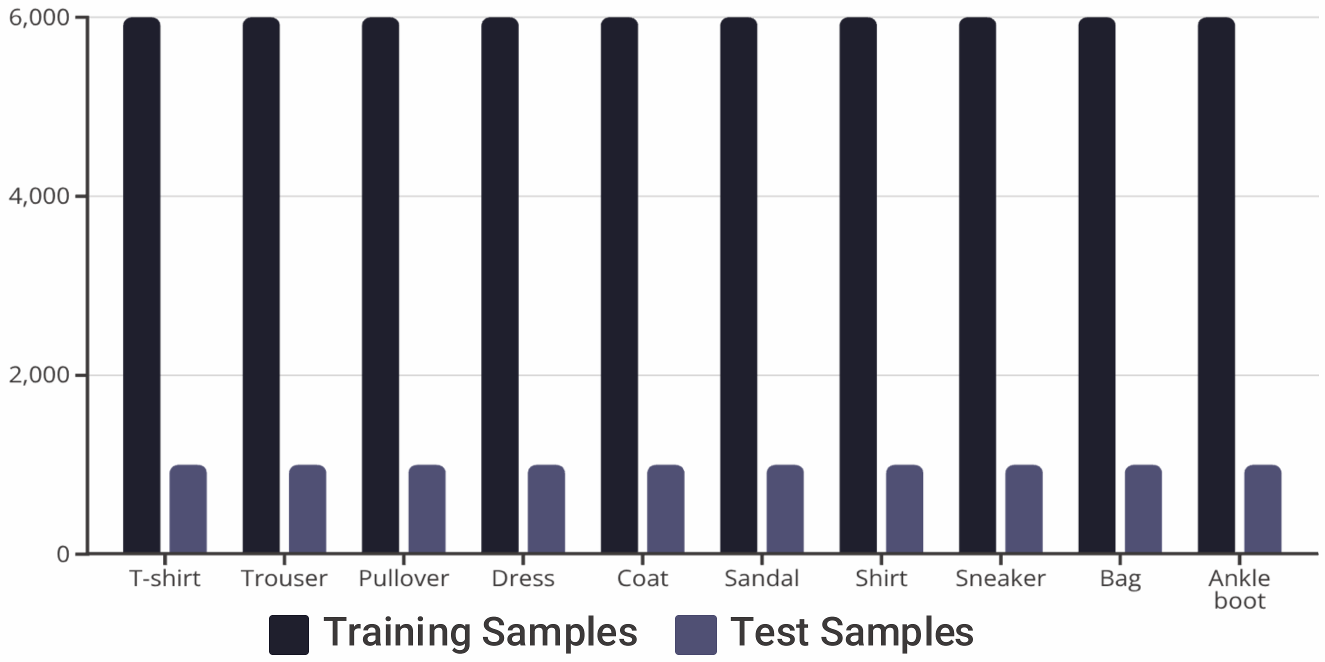
10

## Ankle boot

**Short boots**

Each category presents unique visual characteristics. Some items like T-shirts and Shirts share similar silhouettes, creating interesting challenges for unsupervised clustering algorithms to distinguish between closely related categories.

# EDA: Perfectly Balanced Dataset



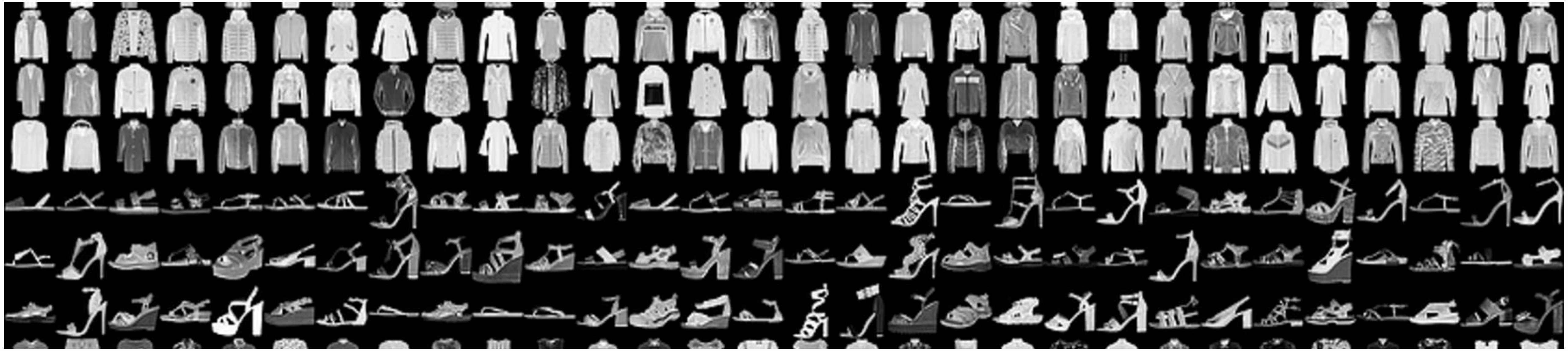
## Distribution Analysis

The exploratory data analysis reveals exceptional dataset quality:

- Each class contains exactly **6,000 training samples** (10.0%)
- Each class contains exactly **1,000 test samples** (10.0%)
- **Total: 70,000 images** across all categories
- Perfect balance eliminates class imbalance issues

**Key Finding:** The perfectly balanced distribution means our clustering algorithms won't be biased toward overrepresented categories, allowing fair evaluation of pattern discovery capabilities.

# Sample Images from Each Category



These sample images illustrate the visual diversity within Fashion-MNIST. Each image is **28×28 pixels**, providing **784 features** per image. The grayscale format simplifies computation while retaining essential shape and texture information needed for clustering.

Notice how some categories like T-shirts and Shirts share similar structures, while others like Bags and Footwear have distinctly different silhouettes. This variation creates both opportunities and challenges for unsupervised learning algorithms.

# EDA: Key Statistical Findings

0.28

### Mean Pixel Value

Average pixel intensity across all images, indicating predominantly dark backgrounds with lighter fashion items

0.35

### Pixel Standard Deviation

Variation in pixel values, showing diverse brightness patterns across different garment types

784

### Features per Image

Total pixel count from 28×28 dimensions, representing high-dimensional input space for clustering

## Brightness Patterns

Different classes exhibit distinctive brightness characteristics:

- **Trousers**: Strong vertical structure with consistent pixel patterns
- **Bags**: Compact central shapes with concentrated brightness
- **Footwear**: Horizontal orientation with defined edges
- **Upper garments**: Varied patterns based on style and fit

## Average Images Analysis

Computing average images per class revealed distinctive visual signatures that help explain clustering performance. Classes with clearer average patterns (like Trousers and Bags) tend to cluster more effectively than classes with high intra-class variation.

# Methodology: Two Unsupervised Approaches

### Autoencoder + KMeans

Deterministic dimensionality reduction followed by centroid-based clustering

### Variational Autoencoder + GMM

Probabilistic latent space learning paired with mixture model clustering

## Three-Stage Pipeline

01

### Learn Representations

Train autoencoder to compress 784-dimensional images into low-dimensional latent space while preserving key information

02

### Extract Latent Space

Encode all test images into learned representations, creating meaningful feature vectors for clustering

03

### Perform Clustering

Apply clustering algorithm (KMeans or GMM) to latent representations to discover natural groupings

Both approaches learn compressed representations that capture essential features while reducing dimensionality. The key difference lies in how they model the latent space: deterministically or probabilistically.

# Model Architectures

## Autoencoder (AE)

### Input Layer

**784 dimensions** (28×28 pixels)

### Hidden Layer

**256 dimensions** with ReLU activation

### Latent Space

**32 dimensions** (compressed representation)

### Hidden Layer

**256 dimensions** with ReLU activation

### Output Layer

**784 dimensions** (reconstruction)

## Variational Autoencoder (VAE)

### Input Layer

**784 dimensions** (28×28 pixels)

### Encoder Hidden

**256 → 64 dimensions** with ReLU

### Latent Distribution

**16-dim mean & variance** (probabilistic)

### Decoder Hidden

**64 → 256 dimensions** with ReLU

### Output Layer

**784 dimensions** (reconstruction)

**Key Difference:** The Autoencoder learns a deterministic mapping from input to latent space, while the Variational Autoencoder learns a *probabilistic distribution* in the latent space, modeling both mean and variance. This probabilistic approach creates a more structured, continuous latent space that's better suited for clustering.

# Clustering Methods: KMeans vs GMM

### Autoencoder → KMeans

**Extract:** 32-dimensional encoded vectors directly from latent layer

**Cluster:** KMeans with k=10 clusters, finding centroids in latent space

**Rationale:** Deterministic clustering matches deterministic encoding

### VAE → GMM

**Extract:** 16-dimensional mean vectors from learned distributions

**Cluster:** Gaussian Mixture Model with 10 components

**Rationale:** Probabilistic clustering aligns with probabilistic latent space

## Why Different Methods?

The choice of clustering algorithm should match the nature of the learned representations:

- **KMeans** assumes hard cluster assignments with spherical boundaries - appropriate for deterministic AE embeddings
- **GMM** models soft cluster assignments with probabilistic boundaries - natural fit for VAE's probabilistic latent space
- Both methods use **10 clusters** to match the known number of fashion categories

This alignment between representation learning and clustering approach is crucial for optimal performance.

# Hyperparameter Optimization

### Autoencoder Best Config

- **Latent Dimension:** 64
- **Learning Rate:** 1e-3
- **Architecture:** 784→256→64→256→784

### VAE Best Config

- **Latent Dimension:** 16
- **Learning Rate:** 1e-3
- **Architecture:** 784→256→64→16→64→256→784

## Search Strategy

Systematic grid search was performed across key hyperparameters to identify optimal configurations:

### Latent Dimensions Tested

- 8, 16, 32, 64, 128 dimensions
- Balancing compression vs information retention
- Smaller dimensions force more abstraction

### Learning Rates Tested

- 1e-4, 5e-4, 1e-3, 5e-3 rates
- Too small: slow convergence
- Too large: unstable training

Interestingly, the VAE performed best with a much smaller latent dimension (16) compared to the AE (64), suggesting that the probabilistic regularization in VAE enables more efficient compression without information loss.

# Evaluation Metrics Explained

## NMI: Normalized Mutual Information

Measures agreement between discovered clusters and true labels using information theory principles.

- **Range:** 0 to 1 (higher is better)
- **1.0 =** Perfect agreement with true labels
- **0.0 =** No mutual information
- **Advantage:** Normalized for fair comparison across different cluster sizes

## ARI: Adjusted Rand Index

Measures similarity between two clusterings while accounting for chance agreement.

- **Range:** -1 to 1 (higher is better)
- **1.0 =** Identical clusterings
- **0.0 =** Random labeling
- **Advantage:** Corrects for random chance in cluster assignment

## Silhouette Score

Measures how well-separated and cohesive clusters are based on intra-cluster and inter-cluster distances.

- **Range:** -1 to 1 (higher is better)
- **+1 =** Dense, well-separated clusters
- **0 =** Overlapping clusters
- **-1 =** Incorrect assignments

Using multiple metrics provides a comprehensive view of clustering quality. NMI and ARI compare against ground truth labels, while Silhouette Score evaluates cluster structure independently.

# Results

## Comprehensive Evaluation Across Two Approaches

After training both autoencoder architectures and applying clustering algorithms to their learned latent representations, we evaluated performance using three complementary metrics. The results reveal interesting patterns about how different architectural choices impact unsupervised learning effectiveness.

1. ### Clustering Performance Metrics

   NMI, ARI, and Silhouette scores comparing AE+KMeans vs VAE+GMM

2. ### Visual Quality Assessment

   Reconstruction quality and latent space structure visualizations

3. ### Comparative Analysis

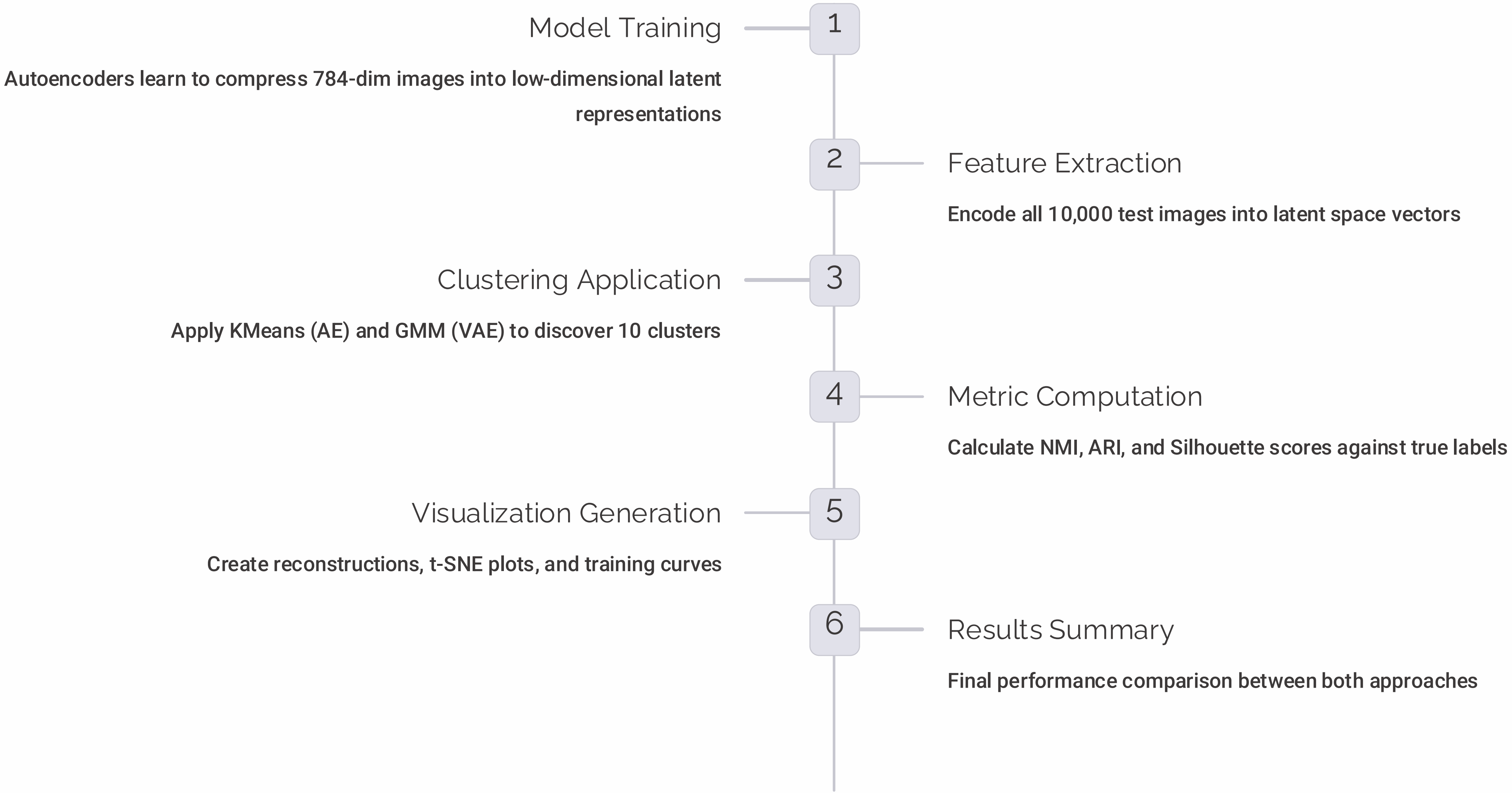
   Understanding why one approach outperformed the other

4. ### Hyperparameter Impact

   How architecture choices affected final performance

# Baseline Model Performance Evaluation

We train both baseline models and observe their performance in real-time. The output shows training progress, reconstruction quality, and latent space visualization as the models learn to compress and cluster fashion images.



1 Model Training

Autoencoders learn to compress 784-dim images into low-dimensional latent representations

2 Feature Extraction

Encode all 10,000 test images into latent space vectors

3 Clustering Application

Apply KMeans (AE) and GMM (VAE) to discover 10 clusters

4 Metric Computation

Calculate NMI, ARI, and Silhouette scores against true labels

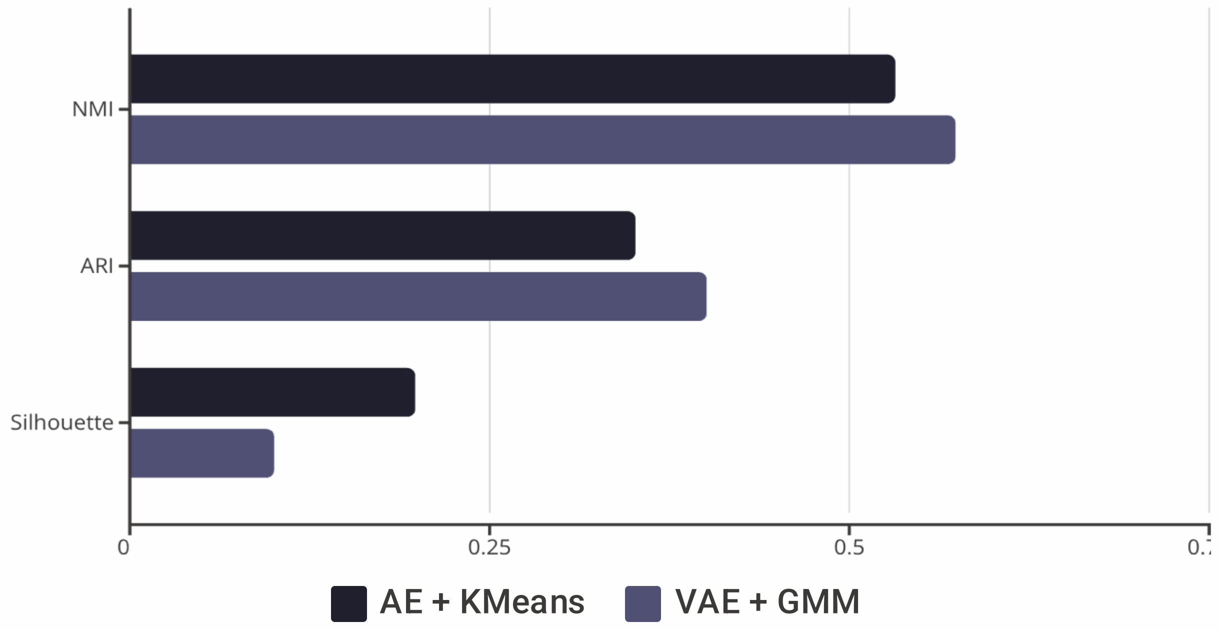
5 Visualization Generation

Create reconstructions, t-SNE plots, and training curves

6 Results Summary

Final performance comparison between both approaches

# Clustering Performance Comparison



## Performance Summary

### AE + KMeans

- **NMI:** 0.5316
- **ARI:** 0.3510
- **Silhouette:** 0.1981

Strong cluster separation but moderate label agreement

### VAE + GMM ★ Best

- **NMI:** 0.5732 ← Winner
- **ARI:** 0.4002 ← Winner
- **Silhouette:** 0.0998

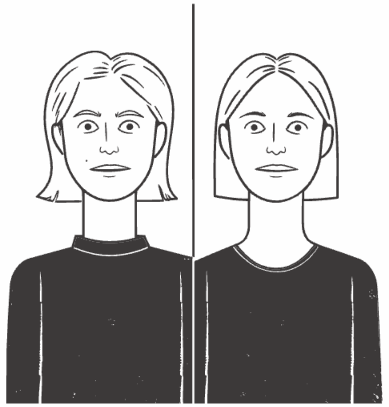
Superior label agreement despite softer boundaries

The VAE + GMM approach achieved the best performance on metrics that compare against true labels (NMI and ARI), demonstrating that probabilistic modeling creates more meaningful semantic clusters. While the Silhouette score is lower, this reflects softer cluster boundaries rather than poor p erformance—the probabilistic nature of GMM creates overlapping clusters that better capture fashion item ambiguity.

# Reconstruction Quality Assessment

## Original vs Reconstructed Images

Both autoencoders successfully learned to compress and reconstruct fashion images, demonstrating that meaningful features are captured in the latent space.



### Autoencoder Reconstructions

- Sharp, detailed reconstructions
- Preserves fine textures and edges
- Occasionally overfits to training patterns

### VAE Reconstructions

- Slightly blurrier but more robust
- Better generalization to variations
- Smoother transitions in latent space

**Key Insight:** The VAE's slightly blurrier reconstructions are not a weakness—they reflect the probabilistic nature of the model, which learns distributions rather than point estimates. This regularization actually helps clustering by creating a more structured latent space.

Visual inspection confirms that both models learned meaningful compressed representations. The quality of reconstructions validates that clustering is performed on genuinely informative latent features rather than noise.

# Latent Space Visualization with t-SNE

## Discovered Structure

The t-SNE projection of the latent space reveals that similar fashion items naturally cluster together in the learned representations:

- **Footwear clusters** (Sneakers, Sandals, Ankle boots) form distinct groups
- **Lower garments** (Trousers) show clear separation
- **Bags** occupy their own region due to unique shapes
- **Upper garments** show some overlap between similar items

## Confusion Patterns

Some expected confusion between visually similar categories:

- **T-shirts vs Shirts**: Similar silhouettes, different necklines
- **Pullovers vs Coats**: Overlapping in texture and coverage
- **Dresses vs Coats**: Variable length creates ambiguity

This confusion makes sense—even humans might struggle to distinguish between some of these categories in low-resolution grayscale.

# Key results

0.5732

VAE + GMM NMI

Best clustering performance - highest agreement with true labels

0.5316

AE + KMeans NMI

Strong baseline performance from deterministic approach

70,000

Total Images

60,000 training + 10,000 test images

784

Input Features

Pixels per image (28×28 dimensions)

16

VAE Latent Dim

Optimal compressed representation size

10

Fashion Categories

Perfectly balanced classes

## Performance Metrics

- VAE ARI: **0.4002** (best)
- AE ARI: **0.3510**
- VAE Silhouette: **0.0998**
- AE Silhouette: 0.1981

# Key Findings from the Analysis

### VAE + GMM Wins

Achieved best performance with **NMI: 0.5732** and **ARI: 0.4002**, demonstrating superior semantic clustering

### Learned > Raw Features

Latent representations dramatically outperform clustering on raw 784-dimensional pixel space

### Probabilistic Advantage

Probabilistic models provide better-structured latent spaces for discovering natural groupings

### Meaningful Visualizations

t-SNE plots confirm that learned representations capture semantic similarities between fashion items

## Expected Challenges

Some confusion between similar categories is expected and understandable:

- T-shirts and Shirts share structural similarity
- Pullovers and Coats have overlapping features
- Low resolution (28×28) limits fine detail
- Grayscale removes color information
- Intra-class variation creates ambiguity
- Some items genuinely fall between categories

# Why VAE + GMM Performed Better

### Probabilistic Latent Space

VAE learns distributions rather than point estimates, creating a smoother, more continuous latent space with fewer gaps

### Built-in Regularization

KL divergence term in VAE loss function encourages the latent space to follow a standard normal distribution, preventing overfitting

### Natural GMM Alignment

Gaussian Mixture Models work perfectly with VAE's probabilistic representations—both reason about distributions

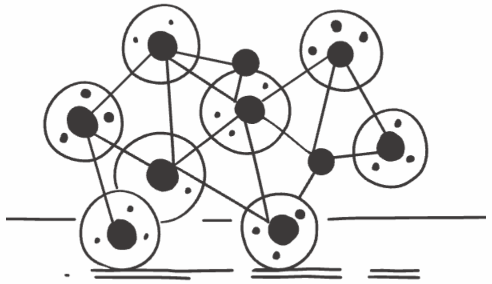
### Better Generalization

Probabilistic modeling handles intra-class variation more effectively, capturing uncertainty about ambiguous items

## Standard AE Limitations

While effective, traditional Autoencoders face challenges:

- **Deterministic mapping** creates point estimates without uncertainty
- **Potential gaps** in latent space between training examples
- **Less regularization** can lead to overfitting on training data
- **Hard boundaries** don't capture ambiguity between similar items



The VAE's probabilistic framework fundamentally changes how the model represents uncertainty, making it better suited for unsupervised tasks where ambiguity is inherent.

# Limitations and Constraints

1

### Supervised Learning Gap

Unsupervised clustering performance (NMI ~0.57) doesn't match supervised classification approaches, which can exceed 90% accuracy on Fashion-MNIST. This is expected, without labels during training, discovering perfect semantic categories is fundamentally harder.

2

### Visual Similarity Confusion

Some items are genuinely difficult to distinguish in 28×28 grayscale images. T-shirts vs Shirts, Pullovers vs Coats, and similar-looking items create unavoidable confusion even for well-trained models.

3

### Hyperparameter Sensitivity

Performance varies significantly with architecture choices like latent dimension size and learning rate. Finding optimal configurations requires careful tuning and computational resources.

4

### Computational Cost

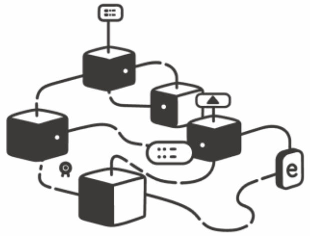
Training deep autoencoders, especially VAEs with their more complex loss functions, requires substantial computation. Hyperparameter search multiplies this cost.

5

### Evaluation Challenges

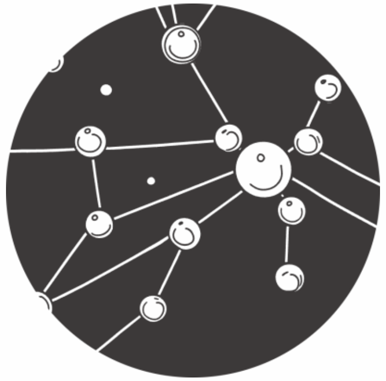
Assessing clustering quality in truly unsupervised scenarios (without ground truth) is difficult. We benefit from having labels for evaluation, but real applications may lack this luxury.

# Future Research Directions



## Convolutional Autoencoders

Leverage spatial structure with CNN layers for better feature extraction. Convolutional architectures naturally capture local patterns like edges, textures, and shapes, critical for fashion items.

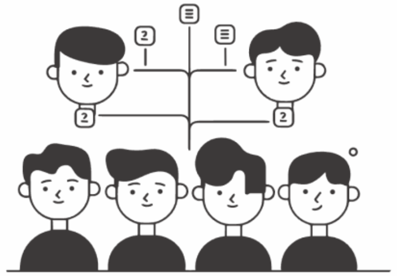


## Deep Clustering

Joint optimization of representation learning and clustering objectives. Instead of two separate stages, train the autoencoder and clustering simultaneously for end-to-end learning.

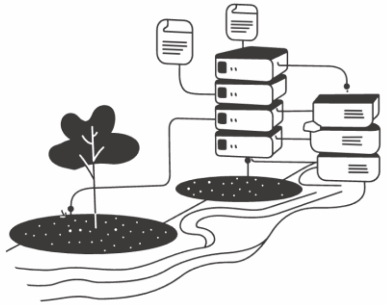
## Semi-Supervised Methods

Incorporate small amounts of labeled data to guide unsupervised learning. Even 1-5% labeled examples could dramatically improve clustering while maintaining scalability benefits.



## Attention Mechanisms

Discover multi-level category structures. Fashion has natural hierarchies (Clothing → Upper Body → T-shirts/Shirts) that hierarchical methods could exploit.



## Hierarchical Clustering

Focus on discriminative regions of images. Attention could help models focus on key features like necklines (T-shirt vs Shirt) or silhouette (Dress vs Coat).

# Conclusion

## Project Achievements

This project successfully demonstrated that unsupervised learning can discover meaningful patterns in fashion images without requiring labeled data during training.

### □ VAE + GMM achieved **NMI of 0.5732**

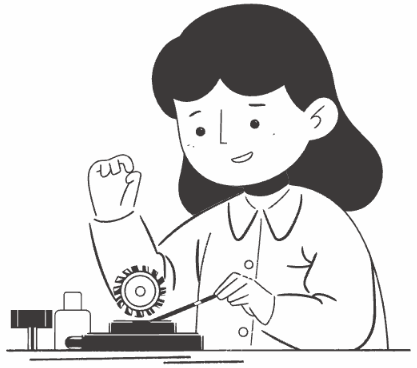
Best-performing approach with meaningful latent representations

### □ Learned features outperform raw pixels

Dimensionality reduction captured essential semantic information

### □ Probabilistic models show advantages

VAE's structured latent space enables better clustering than deterministic AE



> "Unsupervised learning bridges the gap between abundant unlabeled data and the need for intelligent organization, enabling scalable solutions for real-world fashion applications."

[Link to complete repo: https://github.com/kieumyaidev/Unsupervised-lab](https://github.com/kieumyaidev/Unsupervised-lab)

# Reference

**1. Dataset**

- **Fashion MNIST dataset:** https://www.kaggle.com/datasets/zalando-research/fashionmnist

**2. Key Concepts and Methods from CU Boulder’s Unsupervised Learning Course Note**

# AI ACKNOWLEDGMENTS

I would like to acknowledge the use of AI tools in the development of this project:

**Cursor**: Used for debugging code and resolving technical issues during implementation and helping to populate the README file after I finish my work

**ChatGPT**: Assisted with restructuring and proofreading content. I provided the overall structure and bullet points for each section, and ChatGPT helped with minor language revisions and proofreading to improve clarity and flow

**Gamma AI**: Used for formatting presentation slides. The content of the slides was derived from this notebook, and Gamma AI assisted with the visual layout and formatting

*All core concepts, methodology, experimental design, analysis and presentation content are my own work. The AI tools were used primarily for code debugging, language refinement, and presentation formatting assistance.*

## THANK YOU

# Questions?

This presentation covered unsupervised learning approaches for clustering, including exploratory data analysis, methodology design, model evaluation, and future directions.

**Contact:** kieu.doan@colorado.edu
or kimi@earable.ai