K&O CURATOR TECHNOLOGIES GROUP LTD

Official Decision Memo – Fashion Datasets for CuratorAI

Date: October 6, 2025

To: CuratorAI Development Team

From: Keneth Olupot, Founder – K&O Curator Technologies Group Ltd

Subject: Final Selection of Commercially-Compliant Fashion Datasets for CuratorAI

Overview

We appreciate the detailed assessment and recommendations outlined in the Fashion Datasets Analysis for CuratorAI document prepared by Samuel Ssekizinvu. After internal review, we concur that DeepFashion datasets are not suitable for commercial deployment due to their restrictive academic license. In alignment with CuratorAI's commercial roadmap and AI infrastructure plan, we have finalized four datasets that collectively match and extend DeepFashion's capabilities while remaining fully compliant for commercial use.

1. Street2Shop (Hugging Face) – Visual Matching Core

License: CC BY 4.0 | Commercial Use: Permitted with attribution

Scale: ≈ 35,000 matched pairs

Purpose: Train CuratorAI's cross-domain visual retrieval engine to power "Find This Look," similar-style searches, and visual embeddings.

Key Strengths:

- - Provides real-world street-to-shop pairings for contrastive training.
- - Enables accurate outfit recognition across poses, lighting, and backgrounds.
- - Establishes the foundation for CuratorAI's visual similarity and retrieval layer.

2. iMaterialist (Fashion Attributes) – Attribute Learning Backbone

License: MIT License | Commercial Use: Fully permitted

Scale: 1 million+ images, 228 attributes across 8 categories

Purpose: Train CuratorAI's attribute and style classification engine, supporting semantic search and personalization.

Key Strengths:

- - Fine-grained attributes (color, fabric, fit, neckline, pattern, sleeve type).
- - Enables text-to-image and filter-based search queries.

- - Provides large-scale supervised data for attribute tagging and recommendation systems.

3. Fashionpedia – Segmentation and Part-Level Understanding

License: CC BY 4.0 | Commercial Use: Permitted with attribution

Scale: ≈ 48,000 expert-annotated images with 46 categories and 294 fine attributes

Purpose: Support CuratorAI's clothing-part segmentation and structure recognition modules.

Key Strengths:

- - High-quality polygon masks for garment parts (collars, sleeves, hems, pockets).
- - Adds spatial awareness and enhances matching accuracy.
- - Improves CuratorAI's visual reasoning and compositional understanding of fashion items.

4. Kaggle Fashion Product Images (Myntra Dataset) – Structured Commerce Bridge

License: CC0 Public Domain | Commercial Use: Fully permitted

Scale: ≈ 44,000 product photos with complete metadata

Purpose: Provide CuratorAI's retail taxonomy, catalog structure, and metadata foundation for real-world commercial application.

Key Strengths:

- - Includes detailed fields: gender, masterCategory, subCategory, articleType, baseColour, season, year, usage, and product name.
- - Offers ready-to-integrate data for catalog indexing, recommendations, and user context.
- - 100% open for modification and redistribution with no attribution requirements.

Dataset Alignment with CuratorAI Core Functions

Function	Street2Shop (HF)	iMaterialist (Fashion)	Fashionpedia	Kaggle Fashion Images
Visual Matching & "Find This Look"	□ Cross-domain pairs for retrieval	—	□ Clean segmentation assists detection	□ Catalog destination for matching
Attribute Recognition & Tagging	—	□ Attribute backbone	□ Part-aware attributes	□ Taxonomy validation
Personalized Styling &	—	□ Style vectors for	□ Fit & part-	□ Season, usage, and

Recommendations		personalization	level matching	demographic context
Outfit Parsing & Structure Analysis	—	—	Polygon-based segmentation	—
Catalog Enrichment & Trend Insights	—	Attribute analytics	Shape and detail trends	Retail metadata and product analytics

Summary Table

Dataset	License	Commercial Use	Role in CuratorAI	Key Advantage
Street2Shop (HF)	CC BY 4.0	Yes	Visual retrieval	Cross-domain matching (street ↔ shop)
iMaterialist (Fashion)	MIT	Yes	Attribute modeling	Large-scale attribute learning
Fashionpedia	CC BY 4.0	Yes	Segmentation / part parsing	Structural and visual understanding
Kaggle Fashion Images	CC0	Yes	Metadata / taxonomy bridge	Real-world commerce alignment

Conclusion

This four-dataset framework offers CuratorAI a complete, scalable, and license-compliant data foundation. Together, they replicate DeepFashion's core advantages—visual retrieval, attribute recognition, segmentation, and metadata depth—without its legal restrictions. This ensures commercial safety, diversity, data richness, and real-world integration. CuratorAI will proceed with this dataset stack for Phase 1 training and system foundation, integrating user-generated and synthetic data in later optimization phases unless the development team has further suggestions and views.

Approved by:

Keneth Olupot
Founder – K&O Curator Technologies Group Ltd
Date: October 6, 2025