

# FLUX WIKI

## KEEPER MODEL

---

SPECIFICATION FOR THE FLUX PERSONAL TASTE MODEL — WHAT IT LEARNS, HOW IT IS TRAINED, AND WHAT IT CANNOT REPLACE.

[flux.dantesisofo.com/wiki/keeper-model/](https://flux.dantesisofo.com/wiki/keeper-model/)

FLUX\_WIKI\_v2.0

JUNE 2026

FLUX DOCUMENTATION SYSTEM Layer 5 - INTELLIGENCE | keeper-model  
flux.dantesisofo.com/wiki/keeper-model/

## KEEPER MODEL

---

### 1. THE CORE QUESTION

---

The keeper model is built to answer one question:

"What would Dante probably keep?"

Not:

"What is objectively great photography?"

This distinction is canonical. It is not a caveat. It is the entire design philosophy.

The model is not learning universal aesthetics. It is not learning what the photography establishment values. It is not learning what performs well on Instagram. It is not learning what wins awards or gets published.

It is learning: - personal visual taste - selection philosophy - aesthetic thresholds - recurring motifs - sequencing tendencies - the things that consistently get rejected

The keeper archive is the training dataset. The full corpus is the test set. The model is a mirror.

---

### 2. WHY THIS DISTINCTION MATTERS

---

A model trained on "great photography" would learn the aesthetics of whoever labeled the training data. That is a different photographer's eye. It would systematically disagree with the keeper archive.

A model trained on "what would Dante keep" learns from demonstrated behavior across years of practice.

The archive is a self-portrait. The model should learn one photographer's eye. Whose eye depends entirely on whose keeper archive is used.

Every FLUX node with a keeper archive builds its own model. The model is not universal. It is personal.

This is the only correct framing. A model that claims universal photographic quality is lying.

---

### 3. TRAINING DATA

---

Positive examples: ~15,000 (keeper archive)

Negative examples: ~385,000 (full corpus, non-keeper)  
Class imbalance: 3.75% positive  
Near-misses: the most valuable signal (images reviewed and rejected)

The positive class is the existing keeper archive: approximately 15,000 photographs selected from the full corpus during years of active practice, stored in /FLUX\_ARCHIVE/KEEPERS/ in chronological folder structure.

The negative class is the remainder of the full corpus: approximately 385,000 photographs that were not selected. These are not "bad" photographs. They are photographs that did not meet the keeper threshold.

The most valuable training signal is the near-miss: photographs that were reviewed and rejected. A photograph that is nearly a keeper but not quite encodes more information about the selection threshold than a photograph that was never considered.

The full corpus of 400,000 images provides implicit negatives. No explicit negative labeling is required – any photograph not in the keeper archive is a negative example.

---

#### 4. WHAT THE MODEL LEARNS

The model learns patterns in what gets selected and what gets rejected.

**Selection patterns the model detects:** - Compositional preferences (layering, compression, foreground/background relationships) - Tonal preferences (high contrast, fog, flat light – which conditions produce more keepers) - Subject preferences (people at distance vs. close, specific recurring motifs) - Temporal patterns (time of day, season, session length vs. keeper density) - Location patterns (which streets produce more keepers, which building types) - Technical thresholds (acceptable blur, acceptable exposure, acceptable focus distance)

**What the model cannot learn directly:** - Why a specific photograph was made (intention) - What was happening on that walk (narrative) - Emotional weight of specific images - Sequential relationships between frames

The model learns correlations in visual features. It does not learn meaning.

---

#### 5. MODEL ARCHITECTURE

The keeper model is a **ranking model**, not a binary classifier.

A binary classifier would output: keeper / not keeper. A ranking model outputs: keeper score (0.0-1.0) expressing confidence.

Ranking is preferred because the keep/reject decision exists on a spectrum. Some photographs are clear keepers. Some are clear rejects. Most are somewhere between. The model should reflect this.

**Architecture: pairwise comparison**

Input: (embedding\_A, embedding\_B, label: "A is more likely a keeper")

Output: ranking score for A and B relative to each other

Training: construct pairs from (keeper, non-keeper) examples

Inference: score each photograph independently against the learned preference

Alternative: contrastive learning – train the model to push keeper embeddings closer together and farther from non-keeper embeddings in a learned projection space. The keeper score is then the distance from the cluster centroid of keeper embeddings.

Architecture decision deferred until full corpus embeddings are available for experimentation.

---

## 6. KEEPER SCORE

The keeper score is a continuous float on [0.0, 1.0].

1.0 – confirmed keeper (from keeper archive; ground truth)

0.8+ – highly probable keeper; review recommended

0.5-0.8 – uncertain; human review required

0.0-0.5 – probable non-keeper

0.0 – definitive reject signal (clear technical failure or out-of-protocol)

The score threshold for auto-surfacing in the portal is TBD. It will be calibrated against the photographer's actual review behavior after the model's first training run.

The keeper score is stored in the photos table as `keeper_score REAL`. It is recomputed when the model is retrained. Historical scores are not archived by default (they are derived values, not source-of-truth values).

---

## 7. KEEPER MATCHING

The keeper archive (`/FLUX_ARCHIVE/KEEPERS/`) and the full corpus (`/FLUX_ARCHIVE/ORIGINALS/`) exist as separate folder structures. They must be linked.

Matching strategy (applied in order):

1. EXIF timestamp match (primary)

Match keeper EXIF `DateTimeOriginal` against corpus EXIF `DateTimeOriginal`.

Exact match = definitive link.

2. Original Ricoh filename match (secondary)

Match keeper's embedded original filename (`R0001234.JPG`) against corpus filename component.

Match = high-confidence link.

3. SHA-256 hash match (tertiary)

Compute SHA-256 of keeper file; compare against corpus SHA-256 index.

Exact match = definitive link regardless of filename or EXIF.

4. Image similarity fallback

For files where metadata was altered or stripped:

Compute embedding of keeper file; find nearest neighbor in corpus embeddings.

If similarity > 0.98 (near-identical), treat as a match.

Human review required for similarity < 0.99.

Unmatched keepers (in keeper archive, no corpus match found) are flagged for manual review. They may be from sessions not yet imported into the full corpus, or from edited versions that diverged from originals.

---

## 8. WHAT THIS ENABLES

---

Once the keeper model is trained:

- **Auto-ranking during ingest:** new photographs arriving in /FLUX\_INBOX/ are automatically scored. The portal shows keeper score next to each photograph before the photographer reviews.
  - **Issue draft suggestions:** when 36+ high-scoring photographs are available, the system can propose a draft issue sequence sorted by keeper score + chronological order.
  - **Motif clustering:** keeper photographs are clustered by visual similarity (using embeddings). Clusters represent recurring motifs. The photographer can see which motifs appear most frequently in their practice.
  - **Sequencing assistance:** keeper score + temporal order can generate weighted sequence suggestions (see: AUTONOMOUS SEQUENCING).
  - **Session analytics:** per-session keeper density. Which sessions produced the most keepers? Which locations? Which times of day?
- 

## 9. WHAT THIS DOES NOT DO

---

The keeper model does not replace curation.

The photographer still reviews and approves every photograph that enters a FLUX issue. The model ranks. The photographer selects.

The keeper model is wrong sometimes. It is trained on historical taste, which may not match current taste. It does not know about recent aesthetic shifts in the photographer's practice. It does not know what happened during a specific session that makes a technically weaker photograph emotionally important.

The keeper model does not claim universal validity. Its scores are only valid for the photographer whose keeper archive trained it. They are meaningless applied to another photographer's work.

The keeper model is deliberately wrong on the margins. Near-threshold decisions are exactly where human curation is most valuable. The model is confident about clear keepers and clear rejects. It is uncertain about everything in between. That uncertainty is honest.

---

## SEE ALSO

---

---

<b>Document</b>	<b>Layer</b>	<b>Relationship</b>
INTELLIGENCE	Layer 5 – Intelligence	Layer overview; keeper model is a subdocument
EMBEDDINGS	Layer 5 – Intelligence	Input features for the keeper model
TRAINING DATA	Layer 5 – Intelligence	The labeled dataset construction and matching strategy
METADATA ENRICHMENT	Layer 5 – Intelligence	Stores keeper_score per photograph in the SQLite schema
AUTONOMOUS SEQUENCING	Layer 5 – Intelligence	Uses keeper scores as input for sequence generation

---

FLUX\_WIKI\_v2.0 – [flux.dantesisofo.com/wiki/keeper-model/](http://flux.dantesisofo.com/wiki/keeper-model/)