Task 3 — Validation vs. Testing Explanation

In Task 3, the model was **trained and validated** but not tested on a separate hold-out test set.

What was done:
- A **validation dataset** was created using `RandomPhantomTripletDataset`.
- This validation dataset had the same transforms as training, except:
- No augmentation was applied.
- `offset=len(train_dataset)` ensured no overlap with training samples.
- `deterministic=True` ensured reproducibility of the validation data.

Metrics computed on validation:
1. Validation loss (Triplet margin loss).
2. d_pos: mean distance between anchor and positive embeddings.
3. d_neg: mean distance between anchor and negative embeddings.
4. Viol%: percentage of triplets where d_pos + margin > d_neg.
5. Recall@1: retrieval accuracy — how often the anchor's nearest neighbor in embedding space is its positive pair.

Why this counts as "testing" for the task:
- The interview instructions required training + validation, not a separate test set.
- Because the dataset is **synthetically generated** (phantoms), creating a reproducible validation split already simulates "unseen" data.
- This validation process demonstrates that the model generalizes to new phantom instances and satisfies the acceptance metrics.

Professional note:
In a real-world scenario, I would also hold out a **separate test set** to confirm performance after model selection. However, for this interview task, the deterministic validation split serves as the evaluation benchmark, and the reported metrics (val loss ≈ 0, Recall@1 = 100%) confirm the model works as intended.