The 40th Anual AAAI Conference on Artificial Intelligence
January 20 - January 27, 2026 | Singapore

Bridging Streaming Continual Learning via In-Context Large Tabular Models

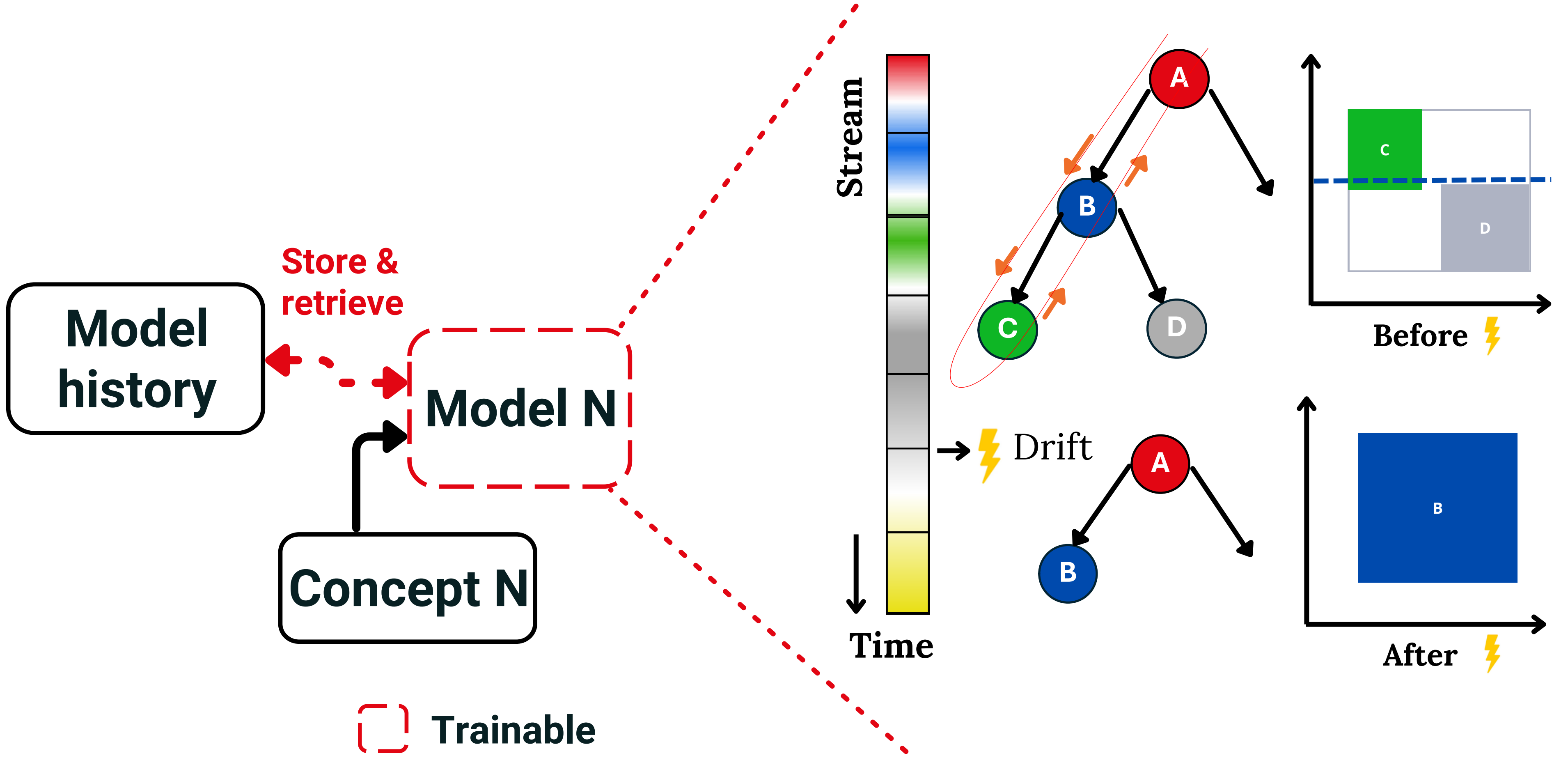
Afonso Lourenço[1], João Gama[2], Eric P. Xing[3,4], Goreti Marreiros[1]

1 GECAD, ISEP, Polytechnic of Porto, Rua Dr. António Bernardino de Almeida, Porto, 4249-015, Portugal
2 INESC-TEC, FEP, University of Porto, Rua Dr. Roberto Frias, Porto, 4200-465, Portugal
3 Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
4 Carnegie Mellon University, Pittsburgh, PA, USA

Conventional in-weights learning

Stream Learning (SL): relies on structural expansion

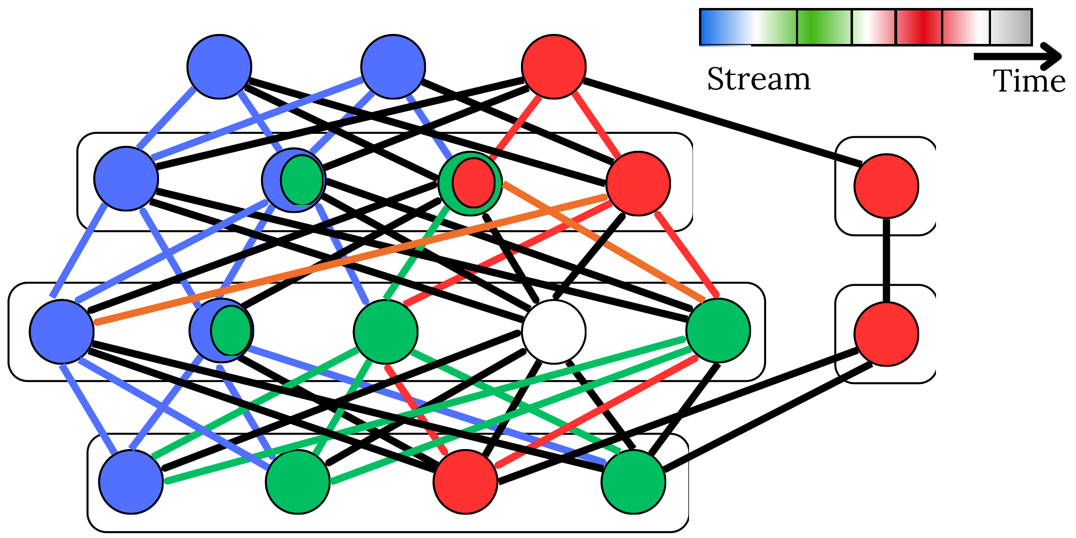
- Ensembles of Incremental Decision Trees (IDTs) as the long-standing SOTA, with fast convergence due to few parameters, using statistical split bounds and subtree replacement
- Limitations of IDT-based SL:
  - 1) limited representational capacity (single-view, axis-aligned splits)
  - 2) plasticity loss from locally optimal greedy splits
  - 3) catastrophic forgetting of class-conditional statistics
- Existing solutions are ad-hoc and narrow, typically via heuristics ensembling (divide-and-conquer), e.g. swapping models upon drift alarms, without assessing relevance of all stored models



OR

Continual Learning (CL): a partial answer

- Relies on structural expansion, and sophisticated parameter adaptation/activation DL schemas
- But, in tabular streams, DL inductive biases poorly match irregular tabular patterns and entangled parameters converge slowly under streaming updates



Streaming continual learning desiderata

Fundamental Mismatch Between SL and CL → Recent hybrids exist, but no unifying framework

SL: rapid adaptation, strict real-time constraints, weak high-order dependencies and forgetting

VS.

CL: long-term retention, mitigating forgetting, relaxed time constraints, weak adaptation

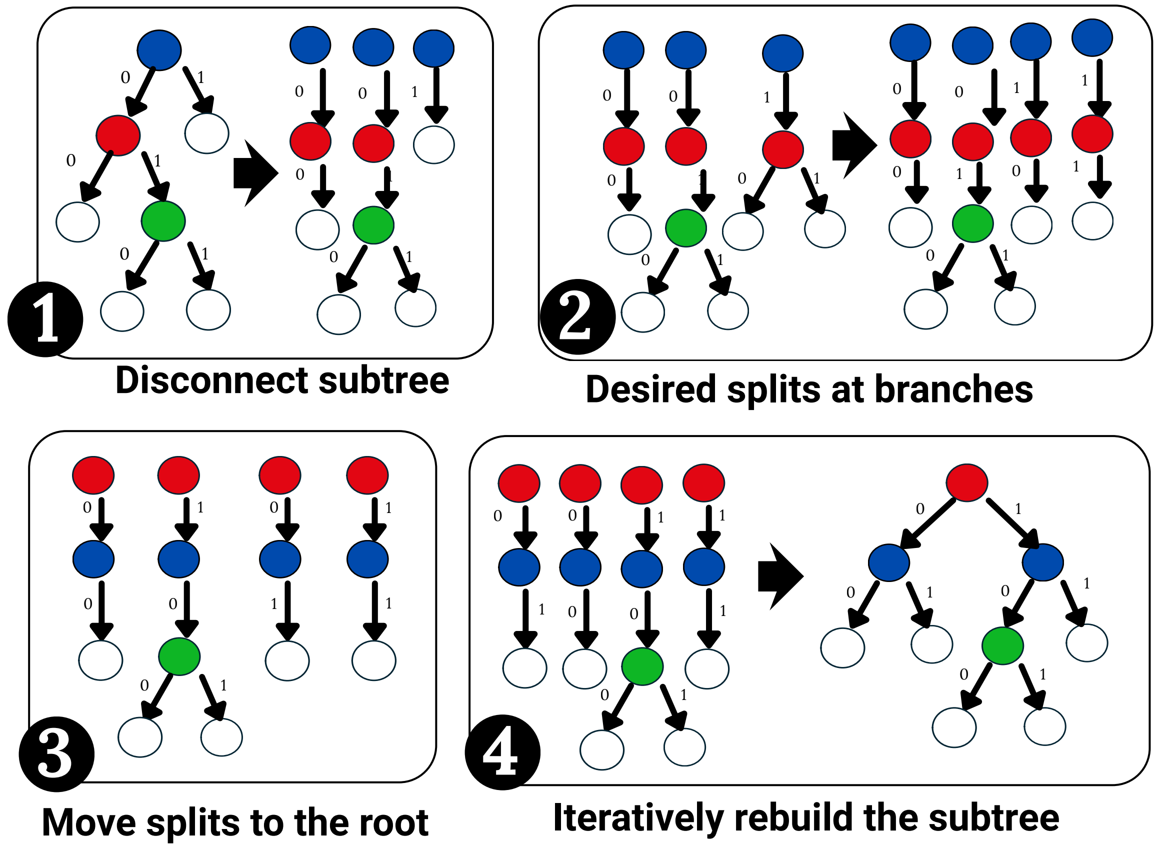
Shared SL–CL desiderata tension:

- **Plasticity**: adapt to current distribution
- **Stability**: retain past knowledge

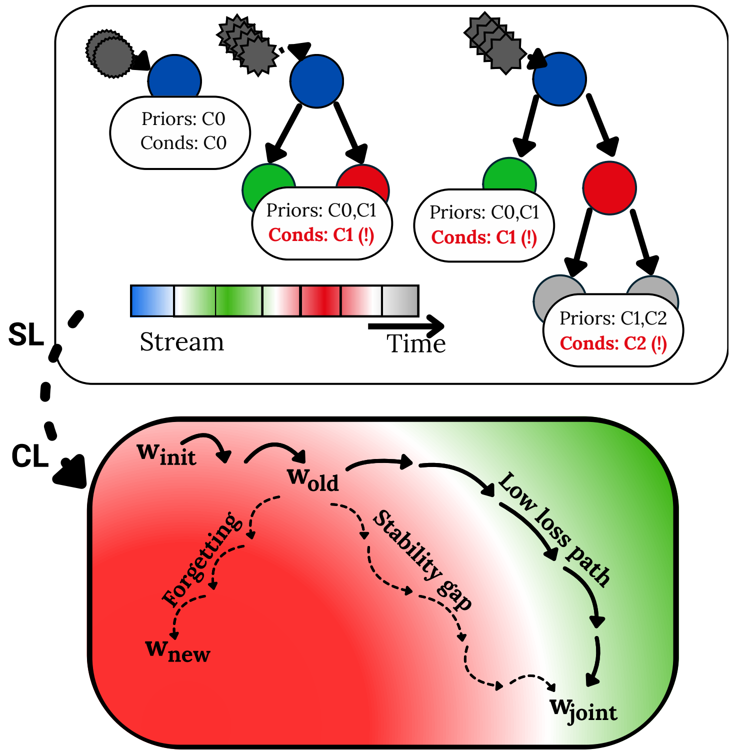
Under memory constraints, this yields two desiderata:

- **Diversification**: avoid redundant stored information
- **Retrieval**: re-activate relevant past experience

Plasticity: Restructuring IDTs by their intrinsic non-overlapping rule decomposition covering the full space

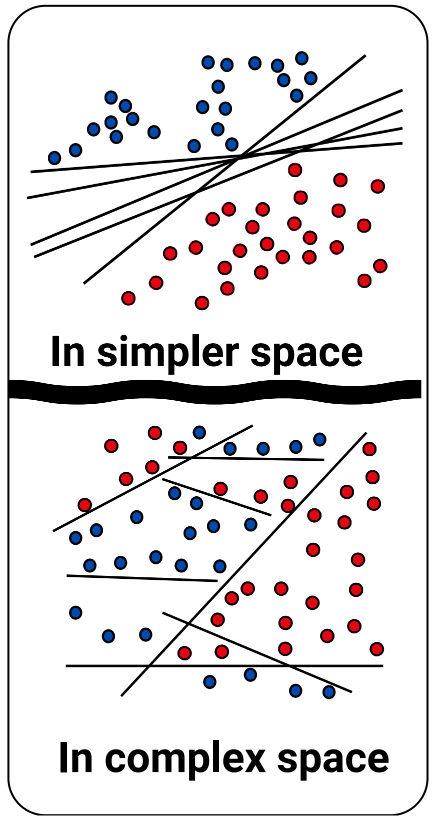


Stability: IDTs' conditional classification fails in class-incremental sequences

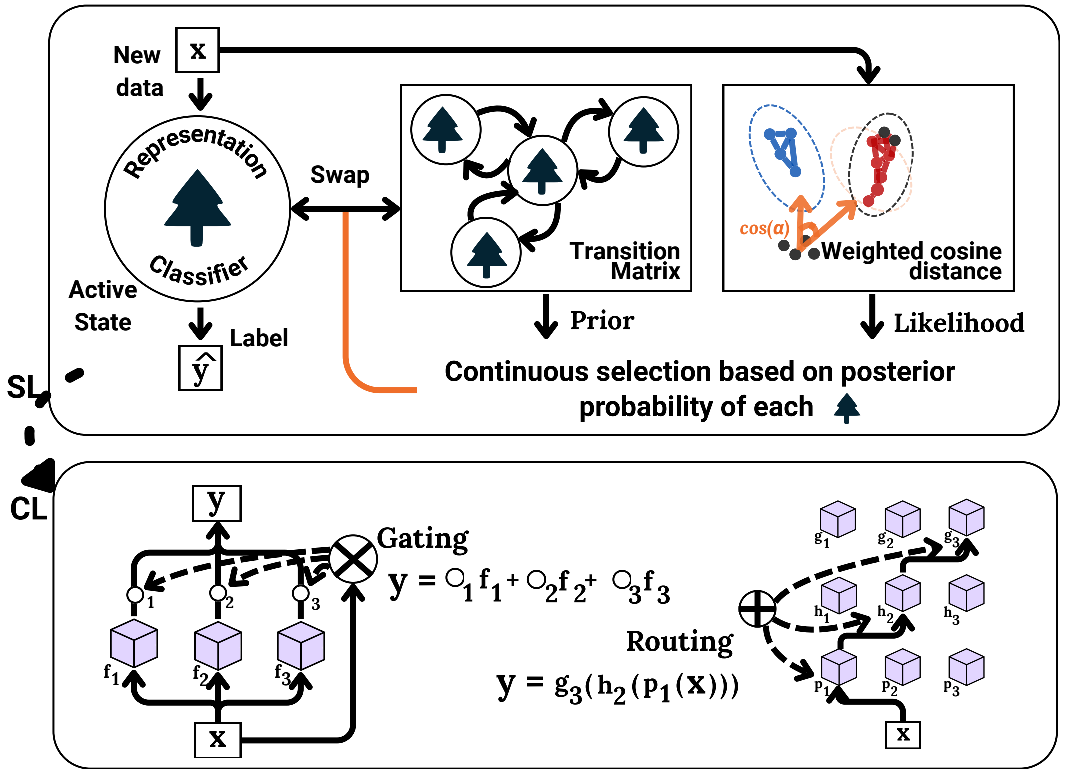


When optimizing the joint loss, SGD passes through a region of high loss on old tasks

Diversification: Multiple non-redundant perspectives



Retrieval: Posterior probabilities for all models, with signals of priors (when expected to reappear) and likelihoods (how it matches current data)



End-to-end differentiable retrieval in CL: mixture-of-experts, modular routing...

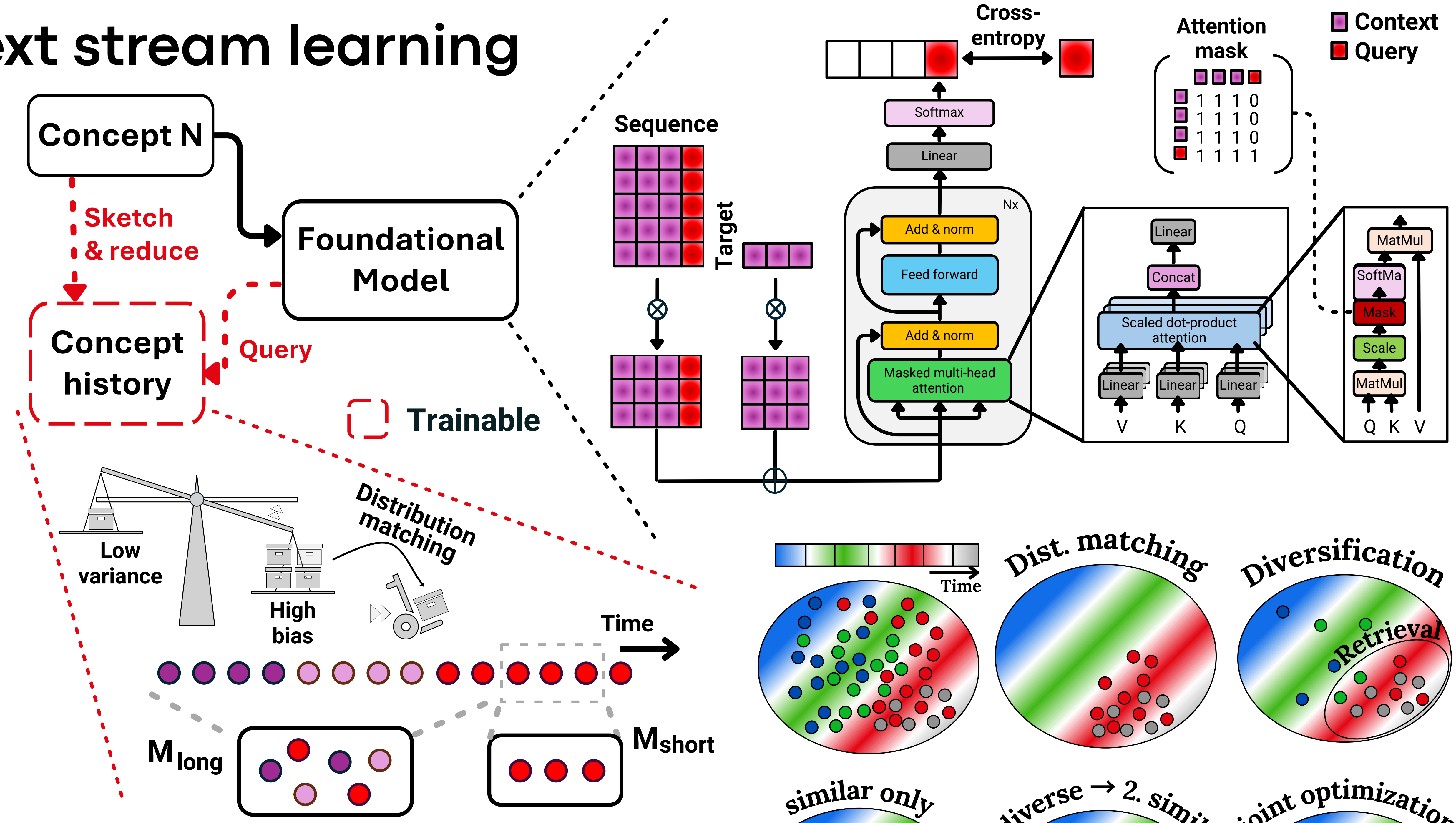
A new paradigm: In-context stream learning

- **Large Tabular Models** (LTMs):
  - Transformer pre-trained on synthetic tabular data
  - Perform instant classification on unseen datasets
  - No parameter updates
  - Huge contexts (500K+ samples, 50K+ features)

1 Two Axes of On-the-fly Sketching for LTMs:

- **Distribution Matching**: select data similar to target (controls the bias–variance trade-off)
  - **Plasticity**: prioritize recent samples for fast adaptation to new patterns, e.g. intra-class variability with short-term buffer
  - **Stability**: retain older samples for global understanding of past knowledge, e.g. inter-class balance with long-term buffer

- Effective matching requires inductive biases:
  - Local distribution smoothness
  - Concept-wise clustering
  - Manifold support
- Naturally leading to compression
  - maximize feature coverage
  - balance similarity and diversity
  - consider difficulty and sensitivity



2 **Distribution Compression**: reduce redundancy while keeping representational power

- **Diversification**: maintain representative, non-overlapping samples, to eliminate
- **(Optional) Retrieval**: select query-specific subsets from memory (enables divide-and-conquer)