



Structural DNA Analysis — Mathematical Framework

One-Pager | Creative Intelligence

Prerequisite: 250 criativos anotados com a Taxonomia Operacional v1.0 (50 códigos)

O Dataset

Cada criativo anotado produz três representações simultâneas:

1. Vetor de Presença (50 dimensões) — Binário. Cada posição = 1 tag. Criativo inteiro vira uma linha numa matriz 250×50.

2. Sequência Ordenada — Lista de tags na ordem em que aparecem. Ex:

`CURIOSITY_OPEN → SELF_SELECT → PAIN_ARTICULATE → SPOILER_MUP → ...` Cada criativo = uma cadeia de comprimento variável.

3. Vetor Posicional — Para cada tag presente, registra o percentil em que aparece (0.0 = início, 1.0 = final). Permite medir: mesma tag, posição diferente, resultado diferente.

As Três Camadas de Análise

Camada 1 — Prevalência (Tags obrigatórias vs. tóxicas)

Input	Método	Output
Matriz 250×50 + performance (Y)	Chi-squared test por tag + regressão logística	Lista de tags "obrigatórias" (>80% dos winners), "opcionais" e "tóxicas" (mais frequentes em losers)

Feature importance: Quais das 50 tags são mais preditivas de sucesso? Random forest com termos de interação captura combinações (ex: ENEMY_FRAME + SCARCITY_INFO + INVALIDATE_SOLUTIONS como "tríade conspiratória").

Camada 2 — Sequência Canônica (A fórmula estrutural dominante)

Input	Método	Output
250 sequências ordenadas + performance	Longest Common Subsequence (bioinformática) + Matrizes de transição de Markov	Sequência vencedora canônica + desvio tolerável

Matriz de Transição: $P(\text{tag}_j \text{ vem depois de } \text{tag}_i)$ calculada separadamente para winners e losers. A divergência entre as duas matrizes revela **onde os winners "viram"** estruturalmente — qual transição específica separa sucesso de fracasso.

Resultado esperado: "Criativos que desviam da sequência canônica em mais de N posições performam X% pior."

Camada 3 — Structural Fingerprint (Múltiplas fórmulas vencedoras)

Input	Método	Output
Vetores posicionais completos	Embedding + clustering (k-means / DBSCAN)	Arquétipos estruturais (ex: cluster "fear-heavy" vs. cluster "proof-heavy")

Cada criativo vira um embedding numérico que codifica presença + ordem + posição simultaneamente. Permite calcular **distância estrutural** entre qualquer par de criativos e descobrir se existem múltiplas fórmulas vencedoras operando por caminhos diferentes.

Variável de Performance (Y)

Cenário	O que funciona	O que não funciona
Com dados de performance (hook rate, CTR, CPA, ROAS)	Todas as 3 camadas com poder preditivo completo	—

Cenário	O que funciona	O que não funciona
Só winners sem variação de performance	Camadas 2 e 3 (extrair fórmula dominante + arquétipos)	Camada 1 limitada (sem grupo controle de losers)

Deliverables Esperados

1. **Tag Scorecard** — Ranking das 50 tags por impacto na performance, com intervalos de confiança
2. **Sequência Canônica** — A fórmula estrutural dominante com tolerância de desvio medida
3. **Mapa de Transições Críticas** — As 5-10 transições (tag_i → tag_j) que mais diferenciam winners de losers
4. **Arquétipos Estruturais** — 3-5 clusters de criativos com padrões distintos, cada um com sua mini-fórmula
5. **Structural Distance Calculator** — Ferramenta que mede quão "longe" um criativo novo está do arquétipo vencedor mais próximo

Framework v1.0 | Fev 2026 | Depende da conclusão da anotação dos 250 criativos com a taxonomia operacional