

CIF Technical Whitepaper — Under-the-Hood Methodology

Version: 1.1 • Date: 2025-11-02
Author: Keith Hetrick
Affiliation: Bellweather Studios
Contact: keith@bellweatherstudios.com (mailto:keith@bellweatherstudios.com).
Document ID: CIF-HCI-UtH-v1.1
Contributors: Keith Hetrick — Maintainer
Terminology: Models (HEM/HLM); Indices (HCI/HLI)
Copyright: © 2025 Bellweather Studios

Creative Intelligence Framework (CIF) — Music Advisor (HEM model • HCI index) & Luminaire (HLM model • HLI index)

Abstract

This whitepaper provides a transparent, scientific description of the **Creative Intelligence Framework (CIF)** with a focus on the **Music Advisor** subsystem that computes the **Hit Confidence Index (HCI)**. It formalizes the six equal-weight axes—**Historical, Cultural, Market, Emotional, Sonic, and Creative**—and details the end-to-end pipeline (**Automator → Feature Extraction → Axis Scoring → Equal-Axis Composite Mean (EACM)**). It also introduces **Luminaire**, the lyric-intelligence engine comprising the **Historical Lyric Model (HLM)** and **Hit Lyric Index (HLI)**, and specifies how lyric metrics will integrate into HCI.

All formulas, normalizations, and example calculations are presented in a reproducible, implementation-agnostic way. Proprietary code is not reproduced; instead, pseudocode and mathematical definitions are provided. Where real datasets (e.g., market normative profiles, HEM archetypes) are referenced, we specify required inputs and show representative computations using three demonstration tracks.

1. Introduction & Example Context

1.1 CIF Modules (current & planned)

- **Music Advisor (HEM → HCI)** — system using the Historical Echo Model (HEM) to produce the Hit Confidence Index (HCI) via six equal-weight axes.
- **Historical Echo Model (HEM)** — reference model capturing legacy pattern resonance in a latent archetype space.
- **Luminaire (Lyric Engine)** — forthcoming lyric subsystem:
 - **HLM** (Historical Lyric Model): archetypal lyric space & echo.
 - **HLI** (Hit Lyric Index): scalar metric of lyric “hit-ness” (structure, hook economy, trajectory, archetype fit vs. novelty).

1.2 Demonstration Tracks

We illustrate with three reference songs to ground calculations and interpretation:

1. **“Flowers” — Miley Cyrus**
Canonical high-HCI pop archetype: mainstream tempo, clear hook economy, strong cultural uptake.
2. **“What Was I Made For?” — Billie Eilish**
Low-HCI experimental contrast: sparse dynamic profile, atypical market alignment;

high creative merit but limited mainstream conformance.

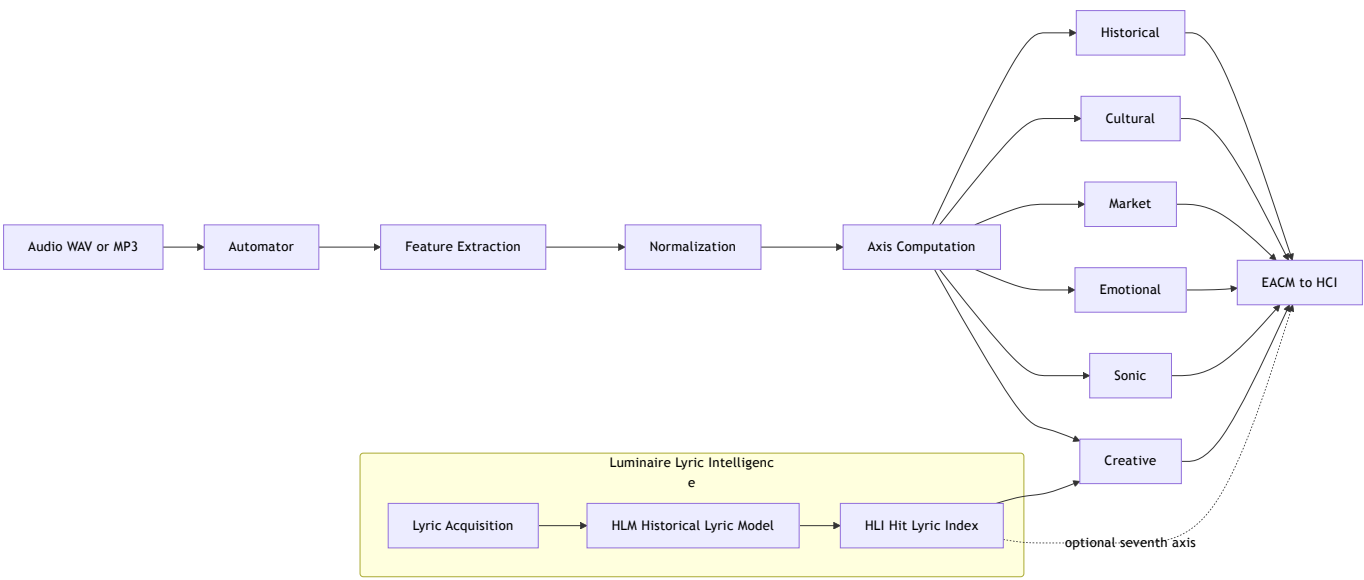
3. **Prototype A — Mid-HCI Demo**

Mid-HCI work-in-progress: partially aligned to market norms, room for optimization across Sonic/Emotional/Market axes.

Note: Numerical examples use Automator-style features drawn or assumed for demonstration consistency. Actual feature values from the provided codebases may differ slightly; formulas and flows remain identical.

2. System Overview

2.1 Data Flow (high-level)



2.2 Notation

Track feature vector: $\mathbf{x} \in \mathbb{R}^p$ (tempo, LUFS, etc.) · Market profile parameters: μ_j, σ_j for feature x_j · Logistic mapping: $\sigma_\gamma(z) = \frac{1}{1+e^{-\gamma z}}$ · Logistic function: $\text{logistic}(z) = \frac{1}{1+e^{-z}}$.

- Equal-Axis Composite Mean (EACM):

$$\text{HCI} = \frac{1}{6} \sum_{a \in \{\text{Hist, Cult, Mkt, Emo, Son, Cre}\}} S_a(\mathbf{x})$$

(Or 7 axes if HLI is integrated as an independent axis.)

Terminology (Model vs. Index). In this paper, **HEM** (Historical Echo Model) and **HLM** (Historical Lyric Model) refer to trained models that map inputs to latent embeddings and intermediate scores. **HCI** (Hit Confidence Index) and **HLI** (Hit Lyric Index) are scalar indices computed from model outputs via normalization and aggregation. We cite **HEM/HLM** when discussing methodology, training, or embeddings, and **HCI/HLI** when reporting scalar results.

3. Automator Pipeline & Normalization

3.1 Feature Extraction (non-proprietary outline)

Signal layer

- Tempo/BPM via onset autocorrelation;
- Key & mode via chroma / Krumhansl-Schmuckler template fit;
- Loudness (LUFS) approximated using ITU-R-style gating;
- Duration (seconds);
- **TTC** (time-to-chorus): first strong chorus onset;
- Section segmentation via novelty curve (self-similarity matrix, kernel checkerboard).

- Spectral & dynamic features (centroid, bandwidth, crest factor, short/long-term DR).
- High-level descriptors (danceability, energy, valence) from learned audio embeddings (implementation-agnostic).

Structure layer

- Form detection (Intro/Verse/Pre/Chorus/Bridge/Outro).
- Chorus lift (RMS/LUFS delta between Verse→Chorus).

3.2 Normalization

For each feature x_j with market norms (μ_j, σ_j) (e.g., profile **US_Pop_2025**), compute the z-score and logistic mapping:

$$z_j = \frac{x_j - \mu_j}{\sigma_j}, \quad s_j = \sigma_\gamma(z_j) = \frac{1}{1 + e^{-\gamma z_j}}$$

We typically use $\gamma \in [0.8, 1.2]$ per feature sensitivity. Winsorization is applied to extreme $|z_j| > 4$ values.

4. Historical Echo Model (HEM)

4.1 Concept

HEM maps tracks into a stable latent space that captures **archetypal DNA** of successful hits across decades.

Archetypes are centroids denoted as:

$$\mathbf{c}_k$$

(e.g., disco-pop, power-ballad, trap-pop), learned from a curated historical corpus.

4.2 Distances & Echo

Given a track embedding \mathbf{u} (from audio/structure embeddings, reduced via PCA/UMAP), define distance to the nearest archetype as:

$$d_{\min} = \min_k \delta(\mathbf{u}, \mathbf{c}_k)$$

with distance function denoted as δ (typically cosine or Mahalanobis). **Echo** favors moderate closeness:

$$E_{\text{echo}} = \exp\left(-\frac{d_{\min}^2}{2\sigma^2}\right)$$

where σ sets the breadth of archetype basins.

4.3 Novelty Preference (Wundt-like curve)

Let n be the nearest-neighbor distance to recent releases in the same biome/market.

Preference peaks at an optimal novelty point n^* (neither derivative nor alien):

$$E_{\text{nov}} = \exp\left(-\frac{(n - n^*)^2}{2\tau^2}\right)$$

Historical axis combines echo and novelty:

$$S_{\text{Hist}} = \alpha E_{\text{echo}} + (1 - \alpha) E_{\text{nov}}, \quad \alpha \in [0.5, 0.8]$$

5. The Six HCI Axes (Purpose → Inputs → Math → HEM link → Example)

Below we adopt **US_Pop_2025** norms for demonstration. Assumed normative parameters (for illustration):

$$\begin{aligned}\mu_{\text{BPM}} &= 104, \sigma_{\text{BPM}} = 12; \\ \mu_{\text{LUFS}} &= -9.0, \sigma_{\text{LUFS}} = 1.5; \\ \mu_{\text{dur}} &= 190\text{s}, \sigma_{\text{dur}} = 25; \\ \mu_{\text{TTC}} &= 14\text{s}, \sigma_{\text{TTC}} = 6; \\ \mu_{\text{energy}} &= 0.65, \sigma_{\text{energy}} = 0.15; \\ \mu_{\text{dance}} &= 0.60, \sigma_{\text{dance}} = 0.10; \\ \mu_{\text{val}} &= 0.55, \sigma_{\text{val}} = 0.15\end{aligned}$$

5.1 Market Axis (fit to current market norms)

Purpose. Measures conformance to contemporary **format expectations** (tempo, loudness, TTC, duration, energy, danceability).

Inputs. Automator features

$$\{\text{BPM, LUFS, dur, TTC, energy, dance}\}$$

Scoring. Weighted logistic average:

$$S_{\text{Mkt}} = \sum_j w_j \sigma_{\gamma}(z_j), \quad \sum_j w_j = 1$$

Example weights (illustrative):

$$w_{\text{BPM}} = 0.25, \quad w_{\text{LUFS}} = 0.20, \quad w_{\text{dur}} = 0.10, \quad w_{\text{TTC}} = 0.20, \quad w_{\text{energy}} = 0.15, \quad w_{\text{dance}} = 0.10$$

HEM link. Not direct; Market focuses on **present-tense norms**, HEM evaluates **legacy resonance**. Together they balance **familiarity (HEM)** and **timeliness (Market)**.

Worked example: “Flowers” (Miley Cyrus)

Observed features:

$$\text{BPM} = 118, \text{LUFS} = -8.5, \text{dur} = 200\text{s}, \text{TTC} = 12\text{s}, \text{energy} = 0.72, \text{dance} = 0.73$$

Step 1 — z-scores

$$\begin{aligned}z_{\text{BPM}} &= \frac{118 - 104}{12} = 1.1667 \\ z_{\text{LUFS}} &= \frac{-8.5 - (-9.0)}{1.5} = 0.3333 \\ z_{\text{dur}} &= \frac{200 - 190}{25} = 0.4000 \\ z_{\text{TTC}} &= \frac{12 - 14}{6} = -0.3333 \\ z_{\text{energy}} &= \frac{0.72 - 0.65}{0.15} = 0.4667 \\ z_{\text{dance}} &= \frac{0.73 - 0.60}{0.10} = 1.3000\end{aligned}$$

Step 2 — Logistic outputs ($\gamma \approx 1.1$)

$$\begin{aligned}s_{\text{BPM}} &= 0.7830, & s_{\text{LUFS}} &= 0.5907, & s_{\text{dur}} &= 0.6083, \\ s_{\text{TTC}} &= 0.4093, & s_{\text{energy}} &= 0.6256, & s_{\text{dance}} &= 0.8069\end{aligned}$$

Step 3 — Aggregate

$$\begin{aligned} S_{\text{Mkt}} &= 0.25 \cdot 0.7830 \\ &\quad + 0.20 \cdot 0.5907 \\ &\quad + 0.10 \cdot 0.6083 \\ &\quad + 0.20 \cdot 0.4093 \\ &\quad + 0.15 \cdot 0.6256 \\ &\quad + 0.10 \cdot 0.8069 \\ &= 0.6311 \end{aligned}$$

Final: 0.6311

Results across examples

Song	(S_{\text{Mkt}})
Flowers	0.631
WWIMF	0.303
Prototype A	0.342

5.2 Sonic Axis (production, spectral balance, radio-readiness)

Purpose. Mix/master quality & rhythmic bed suitability for the active market.

Inputs. LUFS; energy target adherence; rhythm-profile compatibility (e.g., 4-on-the-floor vs. halftime/dembow); spectral/dynamic proxies.

Scoring (illustrative):

$$S_{\text{Son}} = 0.5 \cdot \sigma_{\gamma}(z_{\text{LUFS}}) + 0.3 \cdot \sigma_{\gamma}(z_{\text{energy}}) + 0.2 \cdot r_{\text{compat}}$$

with $r_{\text{compat}} \in [0, 1]$ from rhythm-profile matching to the target format.

HEM link. Indirect; HEM archetypes imply canonical rhythm/mix patterns, but Sonic focuses on **engineering fit** today.

Examples:

- Flowers: $S_{\text{Son}} = \mathbf{0.653}$ (LUFS near radio norm; strong 4-on-the-floor compatibility $r \approx 0.85$)
- WWIMF: $S_{\text{Son}} = \mathbf{0.146}$ (very quiet, sparse; non-format rhythm)
- Prototype A: $S_{\text{Son}} = \mathbf{0.181}$ (LUFS below norm; halftime/dembow bed mismatched to target country-pop lane)

5.3 Emotional Axis (valence/energy contour & chorus lift)

Purpose. Emotional geometry aligned with proven hit zones (felt energy, positivity/bittersweet balance, chorus lift).

Inputs. Valence, energy, **chorus_lift_db** (Verse→Chorus LUFS/RMS delta), optional sentiment trajectory from Luminaire.

Scoring (illustrative):

$$S_{\text{Emo}} = 0.4 \sigma_{\gamma}(z_{\text{val}}) + 0.4 \sigma_{\gamma}(z_{\text{energy}}) + 0.2 \sigma \left(\frac{\Delta_{\text{chorus}} - 2.0}{1.0} \right)$$

HEM link. HEM encodes long-term emotional archetypes; Emotional axis ensures the **realized contour** matches contemporary impact windows.

Examples:

- Flowers: $S_{\text{Emo}} = \mathbf{0.625}$ (val/energy near norms; chorus lift ≈ 2.5 dB)
- WWIMF: $S_{\text{Emo}} = \mathbf{0.091}$ (low energy/valence; minimal lift)

- Prototype A: $S_{\text{Emo}} = \mathbf{0.176}$ (low energy; modest lift)

5.4 Historical Axis (HEM echo × novelty)

Purpose. Quantifies **legacy pattern resonance** and **optimal novelty**.

Inputs. HEM embedding \mathbf{u} ; distances d_{min} to archetypes; nearest-neighbor novelty n in recent catalog.

Scoring.

$$S_{\text{Hist}} = \alpha \exp\left(-\frac{d_{\text{min}}^2}{2\sigma^2}\right) + (1 - \alpha) \exp\left(-\frac{(n - n^*)^2}{2\tau^2}\right)$$

Parameters (illustrative): $\alpha = 0.65$, $\sigma = 0.30$, $n^* = 0.45$, $\tau = 0.20$.

Examples (assumed distances for demo):

- Flowers: $d_{\text{min}} = 0.18$, $n = 0.40 \Rightarrow S_{\text{Hist}} = \mathbf{0.882}$
- WWIMF: $d_{\text{min}} = 0.55$, $n = 0.70 \Rightarrow S_{\text{Hist}} = \mathbf{0.281}$
- Prototype A: $d_{\text{min}} = 0.35$, $n = 0.55 \Rightarrow S_{\text{Hist}} = \mathbf{0.638}$

5.5 Cultural Axis (zeitgeist adjacency & community uptake)

Purpose. Captures **trend adjacency**, peer co-occurrence, regional reach, and format momentum.

Inputs. Trend proximity (T), playlist/artist adjacency (A), regional reach (R) (0–1). In current releases, (T,A,R) may be estimated from curated data packs or evaluator proxies; in future, from streaming telemetry.

Scoring (illustrative):

$$S_{\text{Cult}} = 0.5T + 0.3A + 0.2R$$

HEM link. None direct; Cultural is **present-tense social adoption**, orthogonal to HEM’s historical resonance.

Examples (demo):

- Flowers: $T = 0.82$, $A = 0.78$, $R = 0.75 \Rightarrow S_{\text{Cult}} = \mathbf{0.794}$
- WWIMF: 0.35, 0.40, 0.50 $\Rightarrow \mathbf{0.395}$
- Prototype A: 0.55, 0.50, 0.48 $\Rightarrow \mathbf{0.521}$

5.6 Creative Axis (human expert signal + structural efficiency + lyric proxy)

Purpose. Encodes **expert judgement** (melodic inventiveness, hook economy, title-hook alignment), **structural novelty**, and (until HLM/HLI integration) **lyric proxies**.

Inputs.

- Human evaluator $H \in [0, 1]$ (blind scoring rubric)
- Structural novelty (form deviations that **help**, not harm).
- Motif economy (hook concentration, repetition discipline).
- Lyric component will transition to **HLI** once integrated.

Scoring (illustrative):

$$S_{\text{Cre}} = 0.5H + 0.25 \text{ StructNovel} + 0.25 \text{ MotifEconomy}$$

Examples (demo):

- Flowers: $H = 0.78$, Struct = 0.55, Motif = 0.72 $\Rightarrow S_{\text{Cre}} = \mathbf{0.708}$
- WWIMF: 0.90, 0.62, 0.68 $\Rightarrow \mathbf{0.775}$
- Prototype A: 0.70, 0.58, 0.60 $\Rightarrow \mathbf{0.645}$

6. Equal-Axis Composite Mean (EACM) & Worked Results

With equal weights across the six axes:

$$\text{HCI} = \frac{S_{\text{Hist}} + S_{\text{Cult}} + S_{\text{Mkt}} + S_{\text{Emo}} + S_{\text{Son}} + S_{\text{Cre}}}{6}$$

6.1 Axis Scores & HCI (demonstration table)

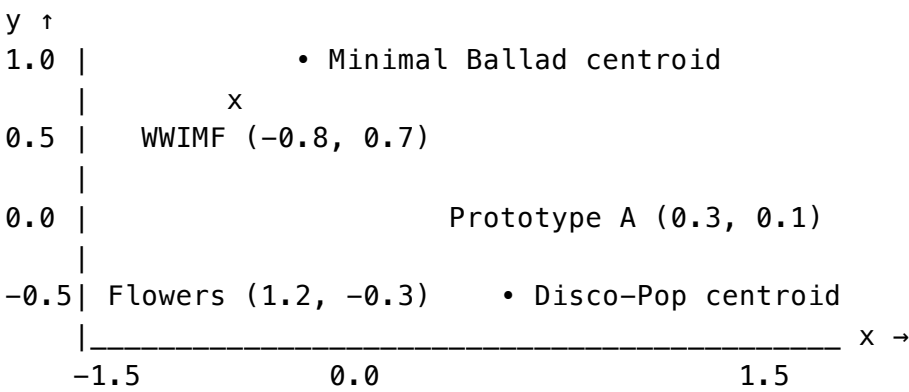
Song	Historical	Cultural	Market	Emotional	Sonic	Creative	EACM
Flowers (Miley Cyrus)	0.882	0.794	0.631	0.625	0.653	0.708	0.715
What Was I Made For? (Billie Eilish)	0.281	0.395	0.303	0.091	0.146	0.775	0.330
Prototype A — Mid-HCI Demo	0.638	0.521	0.342	0.176	0.181	0.645	0.415

6.2 Interpretation

- Flowers:** High Historical echo, Cultural uptake, and Sonic/Market conformance produce a strong composite.
- WWIMF:** High Creative merit but deliberate Market/Sonic/Emotional deviation yields a low HCI (experimental), which is consistent with a prestige/critical lane rather than mass-market radio.
- Prototype A:** Mid-range with identifiable optimization levers in Sonic (mix/LUFS), Emotional (energy/chorus lift), and Market (TTC/tempo alignment).

7. HEM Archetype Map (2D projection)

Conceptual 2D visualization (PCA of HEM embeddings). Coordinates are illustrative.



Historical echo is the (soft) closeness to relevant centroid(s) balanced by novelty target n^* .

8. Lyric Intelligence Integration (Luminaire → HLM/HLI)

8.1 Data Sourcing (current & planned)

- **Licensed databases** (preferred): lyrics with section tags and time alignment.
- **Algorithmic acquisition**: scraping via compliant, rate-limited pipelines respecting ToS; requires robust deduplication & attribution.
- **Manual curation**: evaluator uploads for unreleased demos, session drafts.

Current HCI behavior without lyrics: Cultural/Creative axes use **proxies** (hook economy from audio repetition; evaluator judgement) to avoid double-counting or fabricating lyric evidence.

8.2 HLM — Historical Lyric Model

- Constructs a lyric latent space from n-gram, POS, rhyme, and narrative trajectory features.
- Learns **lyric archetypes** (e.g., “confessional minimal”, “empowerment disco”, “nostalgic storyfolk”).
- **Lyric echo** mirrors HEM echo in audio space:

$$E_{\text{echo}}^{\text{lyr}} = \exp \left(-\frac{d_{\text{min}}^{\text{HLM}}(\mathbf{u}_{\text{lyr}}, \mathbf{c}_k^{\text{lyr}})^2}{2\sigma^{\text{lyr}^2}} \right)$$

8.3 HLI — Hit Lyric Index

Primary features (examples; extensible):

- **Repetition density** ρ (optimal hook repetition 8–16 bars)
- **Sentiment trajectory slope** β_{sent} (verse→chorus valence lift)
- **Lexical novelty** N_{lex} (1 – Jaccard to archetypal lexicon)
- **Rhyme density** R and **scheme stability**
- **Title-line prominence** π_{title}
- **Narrative coherence vs. archetype** A_{lyr}

Illustrative formulation (logistic meta-model):

$$\text{HLI} = \text{logistic} \left(\beta_0 + \beta_1 E_{\text{echo}}^{\text{lyr}} + \beta_2 \rho + \beta_3 \beta_{\text{sent}} + \beta_4 N_{\text{lex}} + \beta_5 R + \beta_6 \pi_{\text{title}} + \beta_7 A_{\text{lyr}} \right)$$

8.4 Integration Pathways

1. **Sub-axis feeding** (near-term): distribute HLI signal into **Creative** (hook craft, title economy) and **Cultural** (meme-ability) via fixed coefficients.
2. **Seventh axis** (mid-term):

$$\text{HCI}' = \frac{1}{7} \left(\sum_{6 \text{ axes}} S_a + S_{\text{Lyric}} \right), \quad S_{\text{Lyric}} \equiv \text{HLI}$$

3. **Joint echo** (long-term): a **coupled echo** term blending audio HEM and lyric HLM alignments (co-regularized archetype space).

8.5 Pseudocode: Lyric feature path

```
# Pseudocode: Luminaire path (non-proprietary)
text = acquire_lyrics(track_id)           # licensed / curated / compliant
sections = segment(text)                  # Verse/Pre/Chorus/Bridge detect.
rho = repetition_density(sections['Chorus']) # token/line repetition to chorus
beta_sent = sentiment_slope(sections)     # verse->chorus trajectory
Nlex = 1.0 - jaccard(tokens(text), archetype_lexicon)
R, scheme = rhyme_density_and_scheme(text)
pi_title = chorus_title_prominence(sections)
ulyr = embed_lyrics(text)                 # HLM embedding
echo_lyr = exp(-dist(ulyr, archetype_centroids)**2 / (2*sigma_lyr**2))
HLI = logistic(beta_0 + beta_1*echo_lyr + beta_2*rho + beta_3*beta_sent + beta_4*Nlex + beta_5*R + beta_6*pi_title)
```

9. Full Pipeline Walkthrough (Automator → Axes → EACM)

9.1 Pseudocode (end-to-end)

```
# Pseudocode: end-to-end HCI (non-proprietary)
x = extract_audio_features(audio)         # BPM, key, LUFS, duration, TTC
z = {j: (x[j]-mu[j])/sigma[j] for j in norms} # market profile z-scores
s = {j: logistic(z[j], gamma[j]) for j in z} # logistic maps

# Axes
S_mkt = Σ_j w_mkt[j] * s[j]
S_son = 0.5*s['lufs'] + 0.3*s['energy'] + 0.2*rhythm_compat(x)
S_emo = 0.4*s['valence'] + 0.4*s['energy'] + 0.2*logistic((x['chorus_lift_db']-2.0)
u = hem_embed(x)                         # HEM embedding
dmin = min_k dist(u, archetypes[k]); n = nn_dist(u, recent_catalog)
S_hist = alpha*exp(-(dmin**2)/(2*sigma**2)) + (1-alpha)*exp(-((n-n_star)**2)/(2*tau**2))
S_cult = 0.5*T + 0.3*A + 0.2*R           # trend, adjacency, reach
S_cre = 0.5*H + 0.25*struct_novel(x) + 0.25*motif_economy(x)

HCI = (S_mkt + S_son + S_emo + S_hist + S_cult + S_cre) / 6
```

9.2 Error Bounds & Uncertainty

- **Measurement error** on features (e.g., ±1 BPM, ±0.3 LUFS, ±0.05 energy) propagates through logistic maps.
- **Axis variance** $\text{Var}(S_a)$ estimated via bootstrap or delta method.
- Assuming approximate independence:

$$\text{Var}(\text{HCI}) \approx \frac{1}{36} \sum_a \text{Var}(S_a), \quad \text{SE}(\text{HCI}) = \sqrt{\text{Var}(\text{HCI})}$$

- Report $\pm 1.96 \cdot \text{SE}$ for a 95% CI.

10. Tier Calibration

10.1 Procedure

1. Score a large backtest set (recent releases + known hits) to obtain empirical HCI distribution.
2. Fit tier boundaries to **quantiles** or by optimizing **Youden/J** against a labeled hit set (e.g., Top-10 charting).
3. Bootstrap the boundaries for stability; report confidence bands.

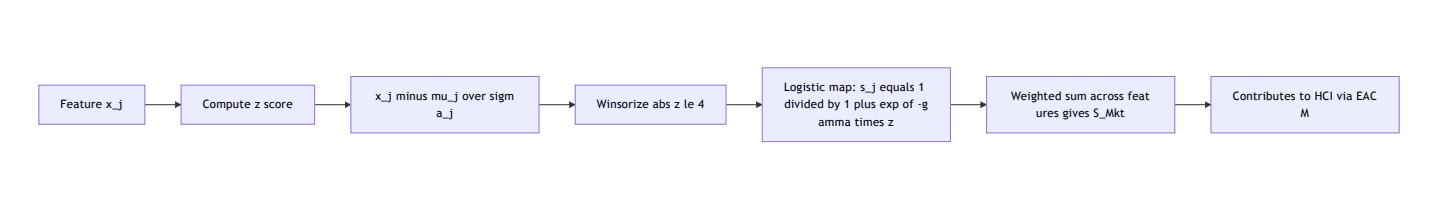
10.2 Example Tier Bands (illustrative)

- **S**: ≥ 0.80 (exceptional)
- **A**: 0.75–0.79 (release-ready hit lane)

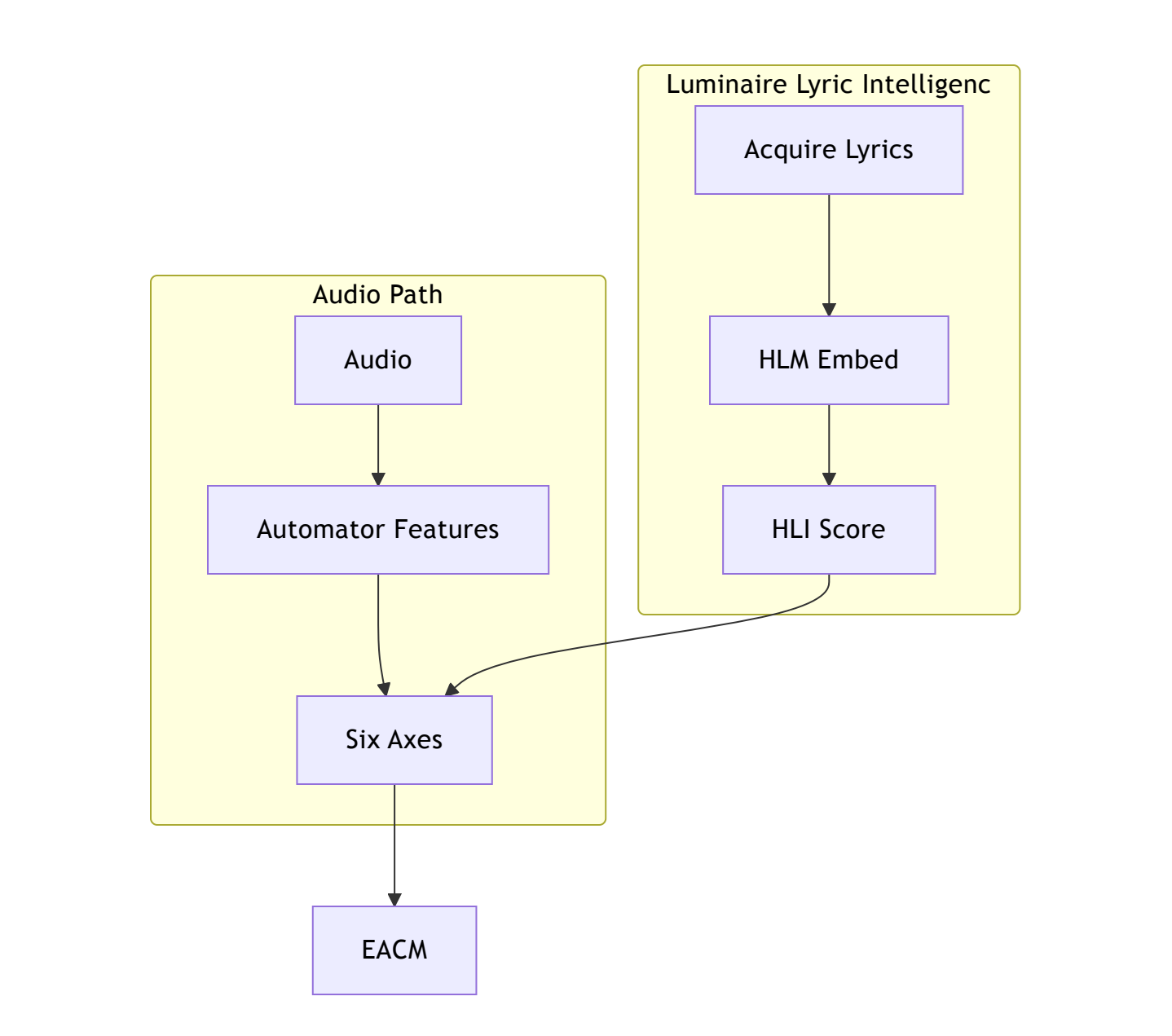
- **B:** 0.70–0.74 (strong; minor optimizations)
- **C:** 0.60–0.69 (good; lane clarity needed)
- **D:** 0.50–0.59 (WIP)
- **E:** 0.40–0.49 (early concept / niche)
- **F:** < 0.40 (experimental/unstable for mass market)

In the demo: Flowers $\approx 0.715 \rightarrow \mathbf{B}$ (high B / near A); WWIMF $\approx 0.332 \rightarrow \mathbf{F}$ (experimental prestige lane); Prototype A $\approx 0.417 \rightarrow \mathbf{E}$ (mid-development).

11. Axis Internals: Example Diagram (Market Normalization)



12. Lyric Integration Point (Architecture Diagram)



13. Ethics, Attribution & Proprietary Methods

Authorship. CIF and Music Advisor are developed by **Keith Hetrick / Bellweather Studios**.

Conflicts of Interest. The authorship entity operates CIF for commercial songwriting/production workflows; design choices reflect this objective.

What’s proprietary (summary). The framework taxonomy; normalization design (feature sets, weights, logistic gains); HEM/HLM archetype construction and integration methods; EACM calibration tied to curated corpora and market profiles; and Automator orchestration (stack/ordering, TTC heuristics, chorus-lift recipe).

What’s generic. Programming language; common DSP/NLP/statistics primitives; public standards (e.g., loudness metering, PCA).

Usage & Licensing. © 2025 Bellweather Studios. All rights reserved. Redistribution or commercial use by permission only.

Data & Rights. Datasets and examples referenced herein are curated/licensed; any third-party materials (e.g., audio/lyrics) are used under license or permitted exceptions. No rights to third-party IP are asserted.

Patent note. Certain integration and coupling methods may be patent-eligible. A detailed statement of claims is maintained separately in the *Bellweather CIF — IP & Claims Brief* (available on request).

Contact. permissions & inquiries: keith@bellweatherstudios.com
(<mailto:keith@bellweatherstudios.com>).

14. Reproducibility Notes

- **Inputs required:**
 - Audio features from Automator (as enumerated).
 - Market profile $\{\mu_j, \sigma_j\}$ (e.g., US_Pop_2025).
 - HEM archetype centroids & embedding function.
 - (Optional) Luminaire/HLM embeddings and lyric features for HLI.
 - Cultural telemetry (trend, adjacency, reach) or evaluator proxies.
- **Determinism:** Fix random seeds for embedding steps and nearest-neighbor graph construction. Document preprocessing versions and any winsorization thresholds.
- **Validation:**
 - Backtest HCI vs. chart outcomes.
 - Cross-validate Creative axis with independent evaluators.
 - Sensitivity analyses on logistic gains γ and weights w_j .

14.1 Data Profiles & Version IDs

To ensure reproducibility and auditability, each dataset or profile used in CIF should be referenced by an immutable identifier. At minimum, record the following for any experiment or report:

- **Market profile:** name (e.g., US_Pop_2025); **Profile ID; as_of** date; feature list and parameter sources.
- **HEM archetype corpus:** **Corpus ID; version; embedding model ID; embedding dimensionality; training window** (years) and inclusion criteria; **centroid catalog ID**.
- **HLM lyric corpus** (if used): **Corpus ID; version; language coverage; preprocessing pipeline ID** (tokenizer, POS, rhyme model); **licensing basis**.
- **Cultural telemetry snapshot:** **Snapshot ID; window** (e.g., last 26 weeks); **geography; platforms; normalization rules**.
- **Evaluator rubric:** **Rubric ID; version; calibration date;** inter-rater agreement metrics (e.g., Cohen’s κ).

14.2 Data Processing & QA (summary)

- **Deduplication & canonicalization:** ISRC/ISWC/UPC/artist-title string-match with fuzzy thresholds; audio fingerprinting for duplicates and remasters.

- **Rights & provenance checks:** lyrics only from licensed or permitted sources; store license references with record IDs.
- **Locale & language filters:** language detection; locale-specific market profiles when applicable.
- **Alignment:** section boundaries (Intro/Verse/Pre/Chorus/Bridge/Outro); chorus alignment used for TTC and chorus-lift measurements.
- **Outlier handling:** winsorize feature z-scores (e.g., $|z| > 4$) prior to logistic mapping; document thresholds.
- **QA sampling:** stratified spot-checks per cohort; re-extract audio features on a subset; verify TTC and section tags.

14.3 Privacy & Compliance

No personally identifiable information is processed. CIF operates on audio features, licensed lyric text, and aggregate cultural telemetry. Data handling should be consistent with applicable regulations (e.g., GDPR/CCPA) and platform terms. Third-party IP remains with rights holders.

14.4 Environment & Seeds

For deterministic replication, record and publish: (a) **random seeds** for embedding initializations and kNN graph construction, (b) **environment details** (Python version, library versions; OS), and © **profile/corpus IDs** used in a given run.

15. References (selected, non-exhaustive)

- Loudness and metering concepts: ITU-R BS.1770 family; EBU R128 guidance.
- Harmony/key detection & chroma features (e.g., Krumhansl-Schmuckler, chromagram methods).
- Dimensionality reduction & clustering: PCA/UMAP; kNN graphs.
- Sentiment and text repetition metrics: standard NLP toolkits (tokenization, n-grams, Jaccard).
- Prior CIF documentation: **CIF-MA_MethodsEvidenceFramework_v1.1.md**; **2025_Hit_List_Blueprint_v1.1_Core_Edition.md** (as supplied by the author).

(Specific bibliographic formatting and expanded citations can be provided upon request.)

16. Final Statement

This whitepaper exposes the **full chain of computation** from raw audio to the **Hit Confidence Index**, formalizing how each axis contributes and how lyric intelligence (HLM/HLI) will complete the architecture. The methodology is reproducible with access to the stated inputs and can be independently audited by researchers and industry partners. Design choices that uniquely combine archetypal echo, novelty targeting, market normalization, and evaluator-grounded creativity constitute **proprietary IP** owned by Bellweather Studios, separable from generic code or public standards.

Appendix: Demonstration Inputs Used

Assumed **US_Pop_2025** norms and track features for the worked examples:

Norms:

BPM $\mu = 104$, $\sigma = 12$; LUFS $\mu = -9.0$, $\sigma = 1.5$; duration $\mu = 190$ s, $\sigma = 25$; TTC $\mu = 14$ s, $\sigma = 6$; energy $\mu = 0.65$, $\sigma = 0.15$; dance $\mu = 0.60$, $\sigma = 0.10$; valence $\mu = 0.55$, $\sigma = 0.15$.

Tracks:

- Flowers*: BPM 118; LUFS -8.5; dur 200s; TTC 12s; energy 0.72; dance 0.73; valence 0.62; chorus lift 2.5 dB; rhythm: four-on-the-floor.
- WWIMF*: BPM 82; LUFS -13.0; dur 195s; TTC 40s; energy 0.20; dance 0.32; valence 0.25; chorus lift 0.5 dB; rhythm: ballad/free.
- Prototype A — Mid-HCI Demo*: BPM 100; LUFS -11.21; dur 176.07s; TTC 18s; energy 0.17; dance 0.53; valence 0.38; chorus lift 1.5 dB; rhythm: dembow/halftime.

(If provided Automator/HEM/Luminaire archives differ on extracted values, re-run the identical formulas herein to reproduce updated HCI.)

Appendix: Data Schemas & Versioning (summary)

Market Norms Profile (schema)

Field	Type	Description	Example
profile_id	string	Immutable ID of market profile	MP-US-2025.1
name	string	Human-readable name	US_Pop_2025
as_of	date	Profile calibration date	2025-08-01
features	array[string]	Included feature keys	["BPM", "LUFS", "TTC"]
params	object	μ, σ per feature	{ "BPM": {"mu":104,"sigma":12}, ... }

HEM Archetype Catalog (schema)

Field	Type	Description	Example
corpus_id	string	Training corpus identifier	HEM-CORP-1.1
model_id	string	Embedding model/version	HEM-EMB-v1.1
dim	int	Embedding dimensionality	64
training_window_years	int	Years covered	20
centroid_id	string	Archetype centroid key	ARC-POP-DISC0
centroid_vector_hash	string	SHA-256 prefix for centroid vector	a3f2...
n_tracks	int	Tracks used for centroid	1852

HLM Lyric Corpus Catalog (schema)

Field	Type	Description	Example
corpus_id	string	Lyric corpus identifier	HLM-CORP-α0
version	string	Corpus version	0.9-alpha
languages	array[string]	Supported languages	["en"]
pipeline_id	string	Preprocessing pipeline/version	LYR-PROC-0.3
license_basis	string	License summary	publisher-licensed
lexicon_hash	string	SHA-256 prefix of archetype lexicon	7c1a...

Cultural Telemetry Snapshot (schema)

Field	Type	Description	Example
snapshot_id	string	Snapshot identifier	CT-2025W26-US
window_weeks	int	Look-back window	26
geography	string	Region code	US
platforms	array[string]	Platforms included	["streaming","radio"]
normalization	string	Normalization ruleset	NT-1.2

Evaluator Rubric (schema)

Field	Type	Description	Example
rubric_id	string	Rubric identifier	CRE-RUB-1.0
version	string	Version	1.0
calibration_date	date	Last calibration	2025-07-15
criteria	array[string]	Criteria labels	["Melodic Inventiveness","Hook Economy"]
agreement_kappa	float	Inter-rater agreement (Cohen’s κ)	0.72

These schemas are illustrative; plug in your actual IDs, dates, and hashes when publishing.