Below is the polished site copy plus a rigorous, reproducible blueprint for how **each score** can be measured and computed in practice — designed to meet the standard of a PhD-level implementation team (statistics, epidemiology, econometrics, ML). It includes formulas, data pipelines, normalization choices, validation checks, and notes on uncertainty, fairness, and privacy.

Overview — what HDI 2.0 measures

HDI 2.0 is a composite index of *health*, *education*, *happiness*, and *environmental impact*. Each dimension produces a score in $[0, 1]$. The overall HDI 2.0 is the geometric mean of the four scores:

$$\mathrm{HDI}_{2.0} = (S_H \times S_E \times S_{Hp} \times S_{Env})^{1/4}$$

We recommend implementing a fully reproducible pipeline: **Data collection → Preprocessing → Normalization → Indicator scoring → Domain aggregation → Composite aggregation → Validation & reporting.**

1) AarogyaShree Score (Health) — practical calculation

Purpose

Measure physical, mental, preventive and lifestyle health using objective diagnostics + validated subjective scales.

Inputs (examples & data sources)

- Biomarkers: HbA1c, fasting glucose, LDL, HDL, blood pressure, BMI, CRP, eGFR (lab data).
- Preventive checks: frequency of screenings (mammogram, BP checks), vaccination status (admin records).
- Psychosomatic: PHQ-9, GAD-7, Perceived Stress Scale (self-report).
- End-of-life quality expectancy: predicted QALYs from survival model (EHR + cohort life tables).
- Genetic risk: polygenic risk scores (if available), converted to relative risk.
- Lifestyle adherence: steps/day (wearable), minutes/week of moderate-vigorous activity, diet score (food frequency questionnaire), yoga/meditation minutes.

Step-by-step scoring

1. **Preprocess & transform each raw indicator**
   - Continuous biomarkers: winsorize at 1st/99th percentiles, log transform highly skewed variables (e.g., CRP).
   - Questionnaire scales: convert to 0–100 raw then invert where needed so higher = better.
   - Genetic risk: convert PRS into a percentile and then map to a protective score: $g = 1 - \text{scaled\_PRS}$ (so higher is better).
2. **Normalize to [0,1] (per indicator)**

   Two robust options — choose based on distribution & interpretability:
   - **Min–max (bounded contexts)**:

     $$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

     where $x_{\min}, x_{\max}$ are chosen from healthy/target bounds (not raw extremes).
   - **Sigmoid mapping (robust to outliers)**:

     $$x' = \frac{1}{1 + \exp(-\alpha(x - \mu))}$$

     Choose α s.t. clinically meaningful differences map sensibly.

   Use min–max for simple lab values with known clinical bounds; use logistic for skewed or non-linear relationships.
3. **Indicator sub-scores and weights**

   The AarogyaShree Score S_H is a weighted average of five indicator groups:

   $$S_H = \sum_{i=1}^{5} w_i \cdot I_i, \qquad \sum w_i = 1$$

   Using your design: $w_i = 0.20$ for each group. Each I_i is itself a weighted mean of normalized subindicators (e.g., biomarkers group may have HbA1c weight 0.3, BP 0.25, LDL 0.2, etc.)

4. **Handle zero/edge cases**

   Use a small epsilon to avoid zeroing the geometric mean later:

$$S_H := \max(\epsilon, S_H), \qquad \epsilon = 10^{-4}$$

5. **Uncertainty & CI**
   - Bootstrap individual-level data to compute 95% CI for S_H.
   - Propagate measurement error from lab tests where known.

Example (toy formula)

If biomarkers produce normalized composite I_{bio}, psychosomatic I_{psy}, end-of-life I_{eol}, genetics I_{gen}, lifestyle I_{life}:

$$S_H = 0.2 \cdot I_{bio} + 0.2 \cdot I_{psy} + 0.2 \cdot I_{eol} + 0.2 \cdot I_{gen} + 0.2 \cdot I_{life}$$

Notes (research tips)

- For I_{eol} use Cox or parametric survival models trained on representative cohorts to predict remaining QALYs, then scale predicted QALYs to [0,1] relative to local reference.
- Validate S_H against hard outcomes (hospitalization, mortality) via ROC/AUC and survival analysis.

2) Education Score — practical calculation

Purpose

Measure literacy, skills, critical thinking, domain mastery (ikigai), and entrepreneurial capacity — focus on *competencies* not just years of schooling.

Inputs

- Standardized tests: reading comprehension in mother tongue & English, numeracy.
- Cognitive assessments: Raven's matrices, reasoning tests.
- Growth-mindset survey items (validated scales).
- Design thinking & critical thinking assessment (project rubric scores, capstone evaluation).
- Practical experience: documented apprenticeships, verified certificates, hours of supervised practice.
- Income proxy / earning potential: current income percentile within cohort, probability of sustained income from labor-market models.

Step-by-step scoring

1. **Create validated instruments**
   - Short adaptive tests for literacy and numeracy (e.g., computer adaptive).
   - Rubrics for project-based assessments (0–100).
2. **Normalize individual indicators to [0,1]** (same mapping rules as health). For aptitude and attitude, combine cognitive test percentile with validated attitude index:

$$I_{\text{Apt}\times\text{Att}} = 0.6 \cdot \text{cog\_pctile} + 0.4 \cdot \text{attitude\_score}$$

3. **Weighting & aggregation**
   Education score S_E:

$$S_E = \sum_{j=1}^{5} v_j \cdot J_j, \quad v_j = 0.2$$

   where each J_j is normalized.
4. **Earning potential modelling**
   - Train a local wage prediction model (e.g., gradient boosting + local features). Convert predicted income to percentile and map to 0–1 with logistic scaling to reduce influence of extreme incomes.
5. **Robustness & fairness**
   - Adjust for socio-economic inputs (SES) when comparing individuals across vastly different environments. Report both raw S_E and SES-adjusted S_E^{adj}.
6. **Validation**
   - Predictive validity: S_E should predict employment outcomes, job retention, or entrepreneurship survival using out-of-sample tests.

3) Happiness Score — practical calculation

Purpose

Capture subjective well-being across 9 domains (33 indicators) using a mixed-methods approach: surveys, time-use diaries, administrative proxies, and community metrics.

Inputs

- Psychological well-being: Life satisfaction (Cantril ladder), PANAS, scaled to 0–100.
- Health & education: from prior domain scores or mapped subcomponents.
- Time use: time diaries or smartphone sensor estimates of activity balance.
- Cultural participation: event attendance, survey items.
- Governance & community vitality: civic participation rates, perceived trust scales.
- Ecological connection: self-report + local green space per capita.
- Living standards: income, housing quality metrics.

Step-by-step scoring

1. **Domain construction**

   For domain d, compute normalized domain score D_d as the weighted average of its indicators. Use psychometric techniques (factor analysis / IRT) to check that grouped indicators measure a single latent construct.

2. **Weights inside domain**

   Use expert elicitation or data-driven weights (principal component analysis) — but keep interpretability. Example: Psychological (15%), Living standards (15%), others as specified.

3. **Construct the Happiness Score**

$$S_{Hp} = \sum_{d=1}^{9} \omega_d \cdot D_d$$

   with $\sum \omega_d = 1$ and domain weights per your spec.

4. **Survey design & sampling**
   - Use stratified sampling within populations for representativeness.
   - Use short cognitive interviews to validate translated instruments.

5. **Handling subjectivity**
   - Anchor vignettes can be used to adjust for cross-cultural differences in response scales.
   - Time stability: measure at multiple time points; report rolling averages.

6. **Psychometric validation**
   - Cronbach's alpha / McDonald's omega for internal consistency.
   - Confirmatory factor analysis to validate the 9 domain structure.

4) Net Impact on Environment — practical calculation

Purpose

Measure environmental footprint (negative) and positive contributions (restorative actions) at personal and group levels.

Inputs

- Personal carbon footprint estimate (travel, energy use, diet) — computed from activity logs or modeled from consumption data.
- Personal water/land use proxies.
- Corporate/group emissions: workplace or community scope 1–3 data.
- Positive actions: verified tree planting, renewable energy installations, recycling rates, contributions to local restoration.

Step-by-step scoring

1. **Estimate emissions & impacts**
   - Convert activities to CO_2e using standard conversion tables (transport: gCO_2/passenger-km; energy: kWh × grid factor).
   - For diet: use average emission factors per food type.
2. **Normalize & invert footprint**
   - Personal Net Emissions indicator E_p: map emissions per capita to [0,1] where **lower emissions** → higher E_p.
   - Use a transformation such as:

$$E_p = 1 - \frac{\log(1 + \mathrm{CO2e}) - \log(1 + L)}{\log(1 + U) - \log(1 + L)}$$

     where L and U are chosen lower/upper bounds (local targets and 99th percentile).
3. **Group/corporate emissions**
   - Allocate workplace emissions to individuals via full-time equivalent (FTE) or activity share. Normalize similarly.
4. **Positive impact**
   - Translate actions (trees planted, kWh renewable used, tonnes CO_2 sequestered) into CO_2e reductions or ecosystem restoration units and normalize to [0,1].
5. **Aggregate Net Impact score**

$$S_{Env} = 0.3 \cdot PBA + 0.2 \cdot E_p + 0.2 \cdot E_{group} + 0.3 \cdot I_{pos}$$

   where PBA = planetary boundaries adherence score (composite of multiple metrics; see Rockström framework mapping).
6. **Counterfactual & double-counting**
   - Require verifiable evidence for positive claims (satellite, receipts, third-party verification). Avoid double counting (e.g., tree plantings claimed by both employer and individual).

Aggregation details — how to compute the final HDI robustly

Geometric mean with floor

Compute domain scores S_1, S_2, S_3, S_4 with each in $[0, 1]$. To avoid zeros:

$$\mathrm{HDI}_{2.0} = \left(\prod_{k=1}^{4} \max(\epsilon, S_k) \right)^{1/4}, \qquad \epsilon = 10^{-4}$$

Sensitivity analysis

- Vary domain weights and compute elasticity: percentage change in HDI given a 1% change in a domain score.
- Report sensitivity plots and Tornado charts.

Uncertainty propagation

- Use bootstrap sampling of individuals to estimate variance and 95% CI of HDI.
- If indicators have measurement variance, propagate via Monte Carlo simulation.

Practical implementation pipeline (engineering + analytics)

1. **Data ingestion**
   - APIs for wearables, labs, administrative records, surveys (mobile/web).
2. **ETL & cleaning**
   - Standardize units, handle missingness, apply winsorization/log transforms.
3. **Normalization module**
   - Implement selectable mapping functions (min–max, logistic, percentiles).
4. **Indicator engine**
   - Compute normalized indicators, subindices, domain scores.
5. **Aggregation & reporting**
   - Compute CI, sensitivity, and produce dashboards.
6. **Storage & reproducibility**
   - Versioned datasets, schema registry, unit tests for scoring logic.
7. **Auth & privacy**
   - Strong encryption in transit/at rest, differential access, consent management, GDPR/India PDP alignment.

Suggested tech stack: PostgreSQL / data lake, Python (pandas, scikit-learn, lifelines), R for psychometrics, Airflow for pipelines, Metabase / Superset / Tableau for dashboards.

Statistical & ethical safeguards

- **Missing data:** multiple imputation (MICE) or Bayesian hierarchical imputation. Report share of imputed values.
- **Bias & fairness:** test for systematic bias by gender, caste, income; provide SES-adjusted scores and raw scores.
- **Transparency:** publish codebook, scoring algorithms, thresholds and clinical references.
- **Privacy:** apply deidentification and limit raw data retention; require user consent for genomic data; use aggregation for public reporting.
- **Third-party verification:** allow external audits, especially for environmental claims.

Validation & calibration (research protocol)

1. **Internal psychometrics:** factor analysis, item response theory for survey-based indicators.
2. **External validity:** correlate domain scores with objective outcomes (mortality, employment, ecosystem health).
3. **Predictive tests:** train models to predict downstream outcomes (1–5 year horizons).
4. **Cross-country comparability:** use anchoring vignettes and local calibration sets.
5. **Pilot & refine:** run pilot studies (N≥1,000) in diverse settings, recalibrate min/max bounds, retrain any predictive models.

Reporting & user experience (website / dashboard copy)

On the site, show:

- **Hero metric:** current HDI 2.0 value with 95% CI.
- **Domain cards:** each shows domain score, top 3 strengths, top 3 improvement actions.
- **Drilldowns:** indicator-level numbers, raw units, and how the normalized score was computed (transparency modal).
- **Policy levers:** what changes (e.g., +0.05 in AarogyaShree via improved preventive screening) would do to the composite HDI — show counterfactual simulations.

Short summary (for website readers)

HDI 2.0 is a scientifically rigorous, transparent, and actionable index that measures well-being across health, education, happiness and environment. Each dimension is computed from validated indicators, normalized to a 0–1 scale using robust transforms, aggregated with carefully justified weights, and combined via a geometric mean to produce a single, interpretable number. We provide uncertainty estimates, fairness adjustments, and a reproducible pipeline so individuals, communities and policymakers can trust and act on the result.