

# Critical Averaging and the Politics of Statistical Representation

Tomas Lorenzo <sup>[0000-0002-3482-8389]</sup>

University of Colorado Boulder, Boulder, CO, USA  
tomas@laurenzo.net

**Abstract.** Responsible AI usually evaluates a system *after* its task has been accepted as legitimate: it asks whether an outcome is fair or transparent while treating the task itself as given. The consequential choices are made upstream of fairness, in the formulation and aggregation of the task; this *task-legitimacy gap* is poorly served by post-hoc correction. Rather than describing the gap again, we instantiate it: a generative system whose function is to make the upstream operations of statistical representation—selection, representation, weighting, normalisation, synthesis—manipulable and legible, a strategy we name *critical averaging*. Built entirely from established methods—PCA, a compact VAE, the Stable Diffusion autoencoder, CLIP—the system computes weighted averages of national flags across seven ordered representation spaces. The France–Uruguay–Palestine case shows population weighting arithmetically erasing the smaller polities, while CLIP retrieval names the erasure: France and its overseas departments and collectivities. The same inputs under the same weighting also produce a different average in every space: the choice of space is itself a political act. A second case, Spain–Venezuela–Western Sahara, shows the same rule redistributing visibility rather than simply erasing it. The contributions are *critical averaging* as a method, the *representation ladder* as an extensible comparative instrument, and a counter-apparatus whose machine-readable provenance trace surfaces task legitimacy before fairness.

**Keywords:** Responsible AI, Generative AI, Problem Formulation, Representation Learning, Commensuration, Critical Technical Practice, National Symbols

## 1. Introduction

The dominant grammar of responsible AI is corrective and downstream. Given a system that exists and a task that has been posed, it asks whether the system discharges that task fairly, explainably, robustly. These questions are necessary, and characteristically late. By the time fairness is measured, prior decisions have been naturalised: that this outcome is worth predicting, that these are the admissible categories, that these features may be weighed against one another, and—least examined of all—that there is a single representational space in which the comparison is meaningful. The literature on problem formulation has named several of these prior moves [8, 14, 23]. We take that

diagnosis as a point of departure and ask a constructive question: not how to describe the gap between upstream formulation and downstream fairness, but how to *build something that occupies it*.

This paper presents a generative system whose purpose is to expose, and place under the user’s hand, the operations by which a statistical representation is produced. Where a conventional system conceals aggregation behind a smooth and authoritative output, this system makes aggregation its surface—the criteria, weights, exclusions, normalisation, and, distinctively, the *representational space* become the manipulable substance of the work. We call the strategy *critical averaging* and develop it against a deliberately charged domain, the national symbol.

The domain is not incidental. Averaging incommensurable things into a single representative figure is a constitutive gesture of statistical and generative AI, performed in embeddings, fitted distributions, and “typical” outputs. National symbols are among the most intensely incommensurable objects a society produces—carriers of identity, sovereignty, and historical injury that are *meant* not to be exchangeable [2]. To compute an average flag is to stage, at maximum symbolic charge, the operation that AI naturalises. The estrangement is the method: the absurdity and violence of the averaged symbol make ordinary aggregation strange enough to be examined.

The system is built around a simple thesis: contemporary AI does not merely *represent* an ideology; it *executes* one. Aggregation rules, weighting schemes, category systems, and learned spaces are not descriptions of the world but operations performed upon it—procedural, infrastructural, largely unseen. The political content of a representation has migrated from what it depicts into the executable apparatus that produces it. *Critical averaging* does not add critique to a working system; it makes the working system *be* the critique.

We make three contributions. First, *critical averaging*: a method that treats aggregation as a political operation to be exposed rather than a neutral procedure to be optimised, decomposed into five sites of decision (Section 3). Second, the *representation ladder*: an extensible prototype that composes established methods as comparative rungs. It holds entities and weights fixed while representation changes, making a normally invisible decision inspectable as a difference in the result (Section 4). Third, an instrument of *task legitimacy*: a way of insisting, through the system and its machine-readable provenance trace, that what is averaged, in what space, by whom, and at whose expense be asked before whether the result is fair (Sections 5–6).

## 2. Background and Related Work

**Problem formulation and the limits of post-hoc fairness.** Fairness cannot be secured by adjusting a model in isolation from its sociotechnical context. Selbst et al. identify the framing and solutionism traps that follow from such abstraction [23]; Passi and Barocas show that decisive fairness work happens in *problem formulation*, where translating an organisational goal into a prediction target fixes who is advantaged and harmed [14]; Kalluri asks not whether AI is fair but how it shifts power [8]; and Wang et al. argue that *predictive optimisation* may be illegitimate prior to any fairness

adjustment [26]. We operationalise this critique as a generative system, on the wager that an instrument performing upstream operations in the open advances the argument differently than another taxonomy.

**The average as a manufactured norm.** The intellectual history of the average is a history of norms presented as facts. Quetelet’s *homme moyen* made the average an ideal type against which individuals could be found deviant [15]; Hacking traces how this fiction reorganised normality [7]. The visual ancestor of our system is Galton’s composite portraiture: photographic superimposition of many faces into one “resultant figure,” devised to reveal a biologically determined criminal type [6]. The experiment failed instructively: Galton observed that the composites were “much better looking” than their components and that their “villainous irregularities” disappeared [6]. The composite became not a discovered type but a smoother norm belonging to no one. Sekula reads composite portraiture and the criminal archive as two poles of a positivist apparatus for regulating deviance [22]; the line from Galton’s optical averaging to the eigenface—and hence to every learned space in this paper’s ladder—is direct [24]. Espeland and Stevens generalise the operation as *commensuration*: transforming qualitatively different things into a common metric by discarding what cannot be measured [4]. Averaging is not the neutral recovery of a central tendency but the construction of a representative fiction and the erasure of differences for which it stands.

**Making aggregation visible in art and design.** Salavon’s mean-averaged image populations translate Galton’s composite into software [20], while Udzu’s averaged world flags include population- and area-weighted variants [25]. Onuoha’s *The Library of Missing Datasets* makes selection legible as absence [13], and Crawford and Paglen expose the taxonomies through which classifiers produce subjects [3]. *Critical averaging* does not join this lineage by producing a further average. The averaged flag already exists, including in population-weighted form [25]; our contribution is an apparatus in which entities and weights hold still while representational space varies, and in which every result carries a trace of the five decisions that made it.

**National symbols as incommensurable signs.** Cerulo shows that flags and anthems are structured symbolic systems whose syntax encodes claims about cohesion, sovereignty, and difference [2]. Their political function depends on non-substitutability, which makes them productive material here: averaging is most legible where it is most transgressive.

**Representation spaces and generative synthesis.** That the space in which data is represented is consequential is a technical fact, even where its political reading remains disputed. Linear subspace methods make any instance a coordinate and any average a reconstruction [24]; variational autoencoders learn nonlinear latents in which barycentric averaging is routine [9]; latent diffusion models synthesise inside the latent of an autoencoder trained at web scale [18]; contrastive encoders organise images by a learned semantic geometry [16]. Each embeds priors about what is salient and discardable, priors that descend from the politics of its training collections [3]. Here these methods become rungs: averaging the same symbols in each exposes those priors as differences in the result.

**Statistical sovereignty, seen from the Global South.** By statistical sovereignty we mean a state’s capacity to define, produce, and control the statistics by which it is known. Seers articulated an early development-economics form of the underlying concern, arguing that statistical priorities should follow a country’s own conditions and purposes rather than imported accounting conventions [21]. Ricaurte analyses data-centric epistemologies as an expression of coloniality [17], and Milan and Treré oppose data universalism that positions the South as data source rather than epistemic agent [12]. We extend the concern from data to apparatus: in a generative regime, sovereignty over symbols is exercised through criteria, weights, exclusions, and representational space. We return to this *symbolic-statistical sovereignty* in Section 6.

**Critical technical practice.** Methodologically, the system belongs to Agre’s *critical technical practice*: technical work in which reflection on the practice forms part of the practice [1]. It continues the author’s work on media appropriation and *explicitation*, turning concealed operations into visible material [10], and on AI-mediated systems for collective memory [11]. Flusser supplies the operative image: a technical image is the output of a black box whose coding must be appropriated if its products are to be read critically [5]. *Critical averaging* is an attempt at literacy by exposure.

### 3. Critical Averaging

We define *critical averaging* as a design and computational method in which aggregation is implemented so that its constitutive operations are explicit, manipulable, and legible, and in which the system surfaces the politics of aggregation rather than producing an authoritative result. It decomposes the seemingly atomic act of “taking an average” into five operations, each a site of decision and politics:

1. **Selection.** Which entities enter the average? Inclusion in the reference set is the first and often most consequential choice, usually presented as a precondition rather than a decision.
2. **Representation.** In what space are entities encoded before combination? A flag may be a raster, palette, structured vector, subspace coordinate, or latent code; each makes some differences averageable and others invisible.
3. **Weighting.** With what force does each entity contribute? Here the average acquires much of its political content, and that content most readily disguises itself as a neutral parameter.
4. **Normalisation.** Into what common frame are incommensurable entities coerced? Flags of different aspect ratios must be forced into a shared frame before averaging; the coercion is itself commensuration [4].
5. **Synthesis.** By what procedure are encoded, weighted, normalised entities combined, and how is the result rendered back into the symbolic register?

Conventional systems perform all five operations and expose none. *Critical averaging* inverts this: the synthesised output is secondary and deliberately allowed to remain unsatisfying, while the operations become the work’s primary material. The user does not consume an average; the user *operates* one and confronts the absence of a neutral

way to do so. Sharply: because each operation is a decision, there is no “correct” average to recover by better engineering, and no space innocent of priors in which to look for one. The familiar semantics of averaging—that it reveals a representative centre latent in a population—is false here: the centre is manufactured by the apparatus, and its claim to representation can enact domination precisely because it appears merely descriptive. Purpose is another contestable input, recorded in the trace; Section 6 takes up what it does and does not license. *Critical averaging* therefore relocates the political question from calibrating an output to legitimising the operations that produce it.

#### 4. The Representation Ladder

The conceptual core is that averaging is one operation performed in many spaces. Each backend computes

$$\mathbf{z}_i = E_b(s_i), \quad \mathbf{z}^* = \sum_i \mathbf{w}_i \mathbf{z}_i, \quad \sum_i \mathbf{w}_i = \mathbf{1}, \quad s^* = D_b(\mathbf{z}^*) \quad (1)$$

where  $s_i$  is an input symbol,  $E_b$  is the backend-specific encoder, and  $\mathbf{z}_i$  may be a vector, latent code, or structured specification. The weighted barycentre  $\mathbf{z}^*$  is computed in the corresponding representation space. The operation  $D_b$  denotes either decoding back into an image or score, or, in non-decodable spaces such as CLIP, retrieving the nearest corpus item in embedding space.

Deterministic projections or posterior means are used rather than samples. The averaging function is shared code, deliberately boring: for a fixed tuple of entities, weights, backend, and seed, if an output is surprising, the explanation must be the space.

The seven rungs are ordered by the provenance and degree of learnedness of their coordinate systems: direct appearance (pixel); hand-specified reductions (palette, structured vector); spaces fitted to the flag corpus (PCA, VAE); and externally learned foundation spaces (Stable Diffusion VAE, CLIP). Any backend satisfying encode, average, and decode-or-retrieve can extend the ladder.

1. **Pixel (per-pixel RGB).** The transparent control condition: a weighted mean of pixel values. Nothing is learned; everything is directly inspectable.
2. **Palette (dominant-colour vector).** Each flag becomes dominant colours with area shares, decoded as proportional bands—discarding spatial structure and exposing a representational choice.
3. **Structured vector (SVG specification).** Palette plus a coarse layout descriptor, averaged and re-emitted as a vector image. How one averages layouts is exactly the commensuration problem; the difficulty is the finding.
4. **Linear learned subspace (eigenflags).** PCA over the corpus yields “eigenflags,” Galton’s composite reborn as linear algebra [6, 24]. A flag becomes a coordinate; the average, a reconstruction.
5. **Nonlinear learned latent (VAE).** A compact convolutional VAE trained on the corpus; the average decodes averaged posterior means [9]. Its soft, grainy output—

the detail the space learned to discard—is the visible signature of a learned prior, not a defect.

6. **Foundation-model latent (Stable Diffusion VAE).** The same averaging in the Stable Diffusion autoencoder latent [18]. Its priors come from outside the corpus entirely: the averaged flag passes through assumptions about what an image is that the user did not author and cannot inspect from the output alone.
7. **Semantic-embedding retrieval (CLIP).** CLIP organises images by learned semantic geometry [16]. Lacking a decoder, the rung averages unit-normalised embeddings and retrieves the corpus flag nearest the computed average. The “average” is therefore answered by naming an existing flag, a decision made inside priors the user did not set. Retrieval is included because it exposes another failure mode: an average collapsing into an existing named entity.

Two additional rungs—concept-token embeddings and a learned symbolic-music latent—remain scaffolds and are excluded from reported results.

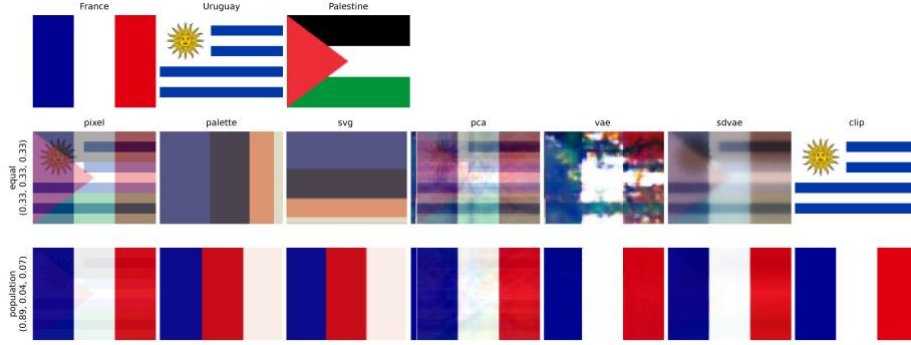
**Implementation.** The backends share an interface and registry. The corpus is a versioned set of 251 country- and territory-flag entries rasterised to a common 4:3 canvas (96×72; 256×192 for the foundation rung), with transparent margins flattened to white. The filter removes supranational bodies and administrative subdivisions while retaining dependent and non-self-governing territories carrying distinct flag entries. This taxonomy is a recorded selection, not an equation of corpus membership with sovereign status. The eigenflag rung uses 32 SVD components; the local VAE, a 64-dimensional latent trained for 1,000 full-batch epochs at seed 0; the foundation rungs use `sd-vae-ft-mse` and CLIP ViT-B/32, run on CPU. Weighting includes uniform, manual, criterion-based, and the continuous family  $w_i \propto P_i^\alpha$ , where  $P_i$  is the population of entity  $i$ . Criterion weighting refuses entities whose metadata remains unvetted, rather than producing impressive but meaningless outputs. The implementation and replication scripts are maintained in the project repository at <https://github.com/krahd/The-National-Average>.

**Provenance.** Every artefact carries a machine-readable trace recording selected entities, representation, raw and normalised weights, normalisation, synthesis backend, and—for learned rungs—checkpoint, latent dimension, and seed; retrieval records the ranked neighbourhood. The trace is the counter-black-box made concrete: it externalises, for inspection and contestation, exactly the decisions a conventional system conceals. The default shareable export is a per-output bundle containing the JSON trace with the generated or retrieved artefact; an image alone is not a complete export. A secondary music pipeline applies the same five operations to hand-authored anthem-like symbolic profiles.

## 5. Worked Examples and Observations

We begin with France, Uruguay, and Palestine: a former colonial metropole, a small Ibero-American republic, and a contested polity, with 2024 population inputs of approximately 68.55, 3.39, and 5.29 million, respectively [27], an order of magnitude apart and sufficient to make weighting legible. The selection is itself operation 1,

exercised in the open and recorded in the trace. We distinguish computed observations from their critical reading: the trace fixes the former; the argument defends the latter.



**Fig. 1.** Same three flags, two weightings, seven representation spaces. Top: sources. Middle: uniform weights (0.33 each). Bottom: population weights (0.89, 0.04, 0.07). Columns, left to right: pixel, palette, structured vector (SVG), eigenflags (PCA), VAE, Stable Diffusion VAE, CLIP; the CLIP column is a retrieval, not a synthesis.

**Weighting: there is no neutral average.** Under uniform weighting the pixel-space average is a muddy composite dominated by a desaturated mauve-grey (mean #8f7a8f) belonging to no flag and no nation—Galton’s norm of no one [6]. Under population-proportional weighting—at least as defensible a notion of fairness, and the principle behind much democratic representation—with normalised weights 0.888, 0.044, and 0.068, the result is almost indistinguishable from the flag of France: Uruguay and Palestine are not under-represented in some correctable sense; they are arithmetically erased. Identical data, identical code; only weighting intent differs. Both rules are reasonable, both objective-sounding, and they encode incompatible politics: one manufactures a non-existent norm and presents it as balance; the other reproduces, under the sign of mathematical fairness, the effacement of small and contested polities by large ones. The population-temperature sweep  $w_i \propto P_i^\alpha$  makes this a continuum (Fig. 2): at  $\alpha = 0$  the average is uniform; by  $\alpha = 2$ , Uruguay and Palestine together hold under one per cent. The slider is not a hyperparameter to tune for fairness; it is the political act itself, exposed as a manipulable surface. That the population rule effaces precisely a contested and occupied territory is not an artefact of the example; it is a small, legible instance of how aggregation under a neutral-seeming rule can perform a politics while disclaiming one.



**Fig. 2.** Population-temperature morph in the Stable Diffusion VAE latent,  $\alpha \in \{0, 0.25, 0.5, 1, 2\}$ . As  $\alpha$  rises, the average resolves from a composite belonging to no one into the French tricolour.

**Representation: the space is political too.** Holding entities and weights fixed and varying only the space, pixel and eigenflag spaces yield continuous blends; palette and structured-vector spaces yield banded reconstructions; the corpus-fitted VAE yields a soft, grainy rendering belonging to none of the sources; the foundation latent resolves high frequencies through external priors into a generic flag-shaped field. Across the six synthesised rungs, none of the 15 pairwise comparisons is pixel-identical. Under uniform weighting, mean pixelwise  $\Delta E_\infty$  is 24.07 (range 6.73–34.46); under population weighting it is 19.08 (0.40–33.87). For calibration, France–Uruguay source distance is 46.79. Divergence is therefore not constant. It is widest under uniform weighting and compresses under population weighting, whose minimum of 0.40 is below the conventional  $\Delta E_\infty \approx 1$  perceptibility threshold, and that compression is itself a finding: when one entity dominates the weights, the spaces converge on it and the choice of space stops mattering. The representational question is loudest precisely where the weighting is least settled. Under uniform weighting: same inputs, same weights, six synthesised averages, no two alike. These measures establish image-level divergence, not its political meaning. That meaning follows from the controlled comparison itself: the “national average” is determined as much by the space as by the weighting, and the choice of space, like the choice of weights, is a decision made on someone’s behalf.

The retrieval rung delivers the starkest version because it answers with a name (Table 1). Under uniform weighting the nearest corpus entry is Uruguay (cosine similarity 0.9203), although France and French overseas entries already crowd the neighbourhood. Under population weighting the nearest entry is France (0.9979), followed at indistinguishable similarity by Saint-Barthélemy, French Guiana, Guadeloupe, and Wallis and Futuna (WLF): French overseas departments and collectivities. The erasure rendered elsewhere as colour acquires a proper name: inside this semantic geometry, the population-weighted “average” of France, Uruguay, and Palestine returns France together with French overseas jurisdictions—the map of an empire returned as a nearest-neighbour list. The ranking is the measurement. Our reading is stronger: the space did not soften the politics of weighting; it ratified and concretised it.

The music pipeline makes the same structure audible: under population weighting the anthem profile collapses note-for-note onto the dominant polity’s motif (0.93 major / 0.07 minor), so the listener hears in eight seconds what the weighted sum performs.

**A second triad.** Spain, Venezuela, and Western Sahara change the geometry: an Iberian metropole, a populous former colony, and a contested non-self-governing territory. Population weighting assigns approximately 0.628, 0.365, and 0.008, respectively. Uniform weighting retrieves Venezuela (0.9346), with Spain second (0.9223); population weighting reverses the pair, retrieving Spain (0.9783). Venezuela is not displaced by becoming less similar—its similarity rises slightly, to 0.9353—but by being overtaken. The case shows that population weighting does not erase every secondary polity equally; it distributes disappearance according to scale, while Western

Sahara falls below one per cent. Its population value and source are themselves recorded inputs, exposing how a weighting parameter imports another representational decision.

**Table 1.** CLIP retrieval: five nearest corpus flags (seed 0).  
F/U/P = France/Uruguay/Palestine; S/V/WS = Spain/Venezuela/Western Sahara.

| Case;<br>weighting    | 1st                 | 2nd                 | 3rd                  | 4th                  | 5th               |
|-----------------------|---------------------|---------------------|----------------------|----------------------|-------------------|
| F/U/P;<br>uniform     | Uruguay<br>0.9203   | France<br>0.9149    | Fr. Guiana<br>0.9149 | St-Barth.<br>0.9149  | Réunion<br>0.9149 |
| F/U/P;<br>population  | France<br>0.9979    | St-Barth.<br>0.9979 | Fr. Guiana<br>0.9979 | Guadeloupe<br>0.9979 | WLF<br>0.9979     |
| S/V/WS;<br>uniform    | Venezuela<br>0.9346 | Spain<br>0.9223     | Bolivia<br>0.9207    | Colombia<br>0.9019   | Andorra<br>0.8973 |
| S/V/WS;<br>population | Spain<br>0.9783     | Venezuela<br>0.9353 | Bolivia<br>0.9273    | Colombia<br>0.9089   | Andorra<br>0.9005 |

Several French tricolour entries tie at four decimal places; their internal order is therefore not substantively meaningful.

## 6. Discussion: From Bias to Task Legitimacy

The examples are trivial as engineering and pointed as argument; the distance between those facts is the contribution. The same instrument applies wherever a system manufactures a representative object from a population: the average customer of a segmentation model, the standard body of a sizing system, the modal voice of a synthetic speaker, or the reference population of a synthetic dataset. Each selects, encodes, weights, normalises, and synthesises a result presented as a finding about the world rather than the product of contestable decisions. Flags supply not generality but legibility: a domain where the cost of commensuration is already understood as a cost. *Critical averaging* is a magnifying instrument for an operation normally too smooth and too fast to see.

This reframes the standard order of evaluation. Post-hoc fairness asks whether a system, having performed its five operations, treats groups equitably in its output. *Critical averaging* insists on the prior questions of *task legitimacy*. Should this aggregation be performed at all? Who selected the reference set, and who was excluded? In whose space are entities encoded, and what differences does that space render unrepresentable? Who set the weights, under what description of fairness? What was forced into commensuration, and at what cost? Whose norm is the resulting average, given that it may be the norm of no one at all? These are not refinements of a fairness metric; they are conditions of legitimacy that a fairness metric presupposes and cannot itself interrogate [26]. The system does not answer them but makes them unavoidable, the most an instrument can do.

Purpose-relative justification does not rescue the correct average. Population weighting may be appropriate to a question about persons and area weighting to one about territory; that justification remains external to the arithmetic. It must be declared, attributed, and disputed, because the resulting average carries no evidence of the purpose that authorised it. The trace makes purpose and rule stated objects rather than inferences from an image. It constrains justification without restoring neutrality.

Here “visibility” names operational rather than psychological properties. *Trace completeness* means that each reported result serialises the five operations, entities, weights, backend, dimensions, checkpoint, and seed. *Controlled attribution* means that within a ladder comparison entities and weights remain fixed while one named representation changes. Whether operating the instrument changes how people reason about aggregation is an empirical question: a future evaluation can compare output-only, trace-only, and operable-ladder conditions on participants’ ability to identify weighting rationales and representation-space effects.

The ladder sharpens this in a direction the fairness literature does not often reach. Much fairness work treats the feature space as fixed; the ladder shows that the space is itself a choice with political consequences, and that as representation becomes learned and foundation-scale, decision migrates from the user into infrastructure—the movement Rouvroy and Berns call *algorithmic governmentality*, government through aggregates that bypasses the subjects it organises [19]. This is the empirical face of *executed ideology*. To critique such a system from outside, after the fact, is to arrive where Flusser warns we are illiterate before the black box [5]. *Critical averaging* answers at the level of the apparatus, in the lineage of critical technical practice [1] and explicitation [10]: executable politics becomes manipulable surface, the user appropriates the operation rather than merely consuming its product, and the trace carries legibility where the image alone no longer suffices.

This is *symbolic-statistical sovereignty*. Statistical sovereignty concerns who defines and controls the statistics by which a nation is known; Seers argued that those statistics should answer to a country’s own conditions rather than to imported conventions [21], and analyses of data coloniality sharpen the point for present infrastructure [12, 17]. The extension we propose is to the apparatus: in a generative regime, control of the averaging apparatus—weights, exclusions, representational space—is control of which symbolic nation can be produced. The examples make the Ibero-American stake concrete: under a population rule executed inside a web-scale embedding, a South American republic is absorbed into the neighbourhood of a European metropole and its overseas jurisdictions, and an Iberian metropole overtakes the republic it once colonised. The instrument does not resolve the question of sovereignty; it relocates it from the hidden interior of a model to a surface that can be argued over.

## 7. Limitations

The system is a research prototype. The corpus-fitted models are deliberately small and overfit; the VAE’s blur is analytically useful, but no generative-quality claim follows. Foundation rungs are run at modest resolution on CPU, and CLIP retrieval is reported

for two cases rather than as a corpus-wide benchmark. Pairwise  $\Delta E_\infty$  establishes image divergence, not semantic or political difference. The corpus taxonomy is described in Section 4, and the corpus is not claimed as a complete geopolitical dataset. Palette and structured-vector rungs assume slot correspondence that does not strictly hold; the music profiles are hand-authored. No user study is reported, and the evaluation question is specified in Section 6. Finally, the artefacts remain vulnerable to aestheticisation and recuperation. Criterion weighting refuses unvetted metadata, and default export couples each result to its trace, but neither mechanism abolishes that risk. They make separation of output from procedure explicit; they cannot prevent it.

## 8. Conclusion

Responsible AI begins, too often, after the decisive choices have been made and naturalised. We have argued that the legitimacy of a statistical or generative system is settled upstream—in selection, representation, weighting, normalisation, and synthesis—and built a system that occupies that gap rather than describing it again. *Critical averaging* makes aggregation’s politics the manipulable surface of the work; the *representation ladder* shows that the “national average” depends as much on space as on weights; and a machine-readable trace carries legibility into the most opaque rungs. Applied to national symbols, the instrument makes the manufacture and erasure performed by an average visible at a glance and, in retrieval, legible as a proper name. AI critique need not remain external commentary on a black box. It can be implemented as a system that exposes the conditions of its own representation—and insists that we ask not only whether a system solves its problem fairly, but in whose space, by whose weights, and at whose expense its problem became a legitimate thing to solve.

**Disclosure of Interests.** The author has no competing interests to declare.

## References

1. Agre, P.E.: Toward a critical technical practice: lessons learned in trying to reform AI. In: Bowker, G.C., Star, S.L., Turner, W., Gasser, L. (eds.) *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*, pp. 131–157. Erlbaum, Mahwah (1997)
2. Cerulo, K.A.: *Identity Designs: The Sights and Sounds of a Nation*. Rutgers University Press, New Brunswick (1995)
3. Crawford, K., Paglen, T.: Excavating AI: the politics of images in machine learning training sets. *AI & Society* **36**, 1105–1116 (2021)
4. Espeland, W.N., Stevens, M.L.: Commensuration as a social process. *Annual Review of Sociology* **24**, 313–343 (1998)
5. Flusser, V.: *Towards a Philosophy of Photography*. Reaktion Books, London (2000)
6. Galton, F.: Composite portraits, made by combining those of many different persons into a single resultant figure. *Journal of the Anthropological Institute of Great Britain and Ireland* **8**, 132–144 (1879)
7. Hacking, I.: *The Taming of Chance*. Cambridge University Press, Cambridge (1990)

8. Kalluri, P.: Don't ask if artificial intelligence is good or fair, ask how it shifts power. *Nature* **583**, 169 (2020)
9. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. In: *Proceedings of ICLR 2014* (2014)
10. Lorenzo, T.: Media appropriation and explicitation. *Journal of Science and Technology of the Arts* **8**(2), 27–36 (2016)
11. Lorenzo, T.: Capturing collective memories of the disappeared with artificial intelligence. In: Correia, L., Rosá, A., Garijo, F. (eds.) *Advances in Artificial Intelligence – IBERAMIA 2024*. LNCS, vol. 15277, pp. 501–504. Springer, Cham (2025)
12. Milan, S., Treré, E.: Big Data from the South(s): beyond data universalism. *Television & New Media* **20**(4), 319–335 (2019)
13. Onuoha, M.: *The Library of Missing Datasets*. Mixed-media installation (2016), <https://www.mimionuoha.com/the-library-of-missing-datasets>, accessed 7 Aug 2026
14. Passi, S., Barocas, S.: Problem formulation and fairness. In: *Proceedings of FAT\* 2019*, pp. 39–48. ACM, New York (2019)
15. Quetelet, A.: *A Treatise on Man and the Development of His Faculties*. W. & R. Chambers, Edinburgh (1842)
16. Radford, A., Kim, J.W., Hallacy, C., et al.: Learning transferable visual models from natural language supervision. In: *Proceedings of ICML 2021*. PMLR 139, 8748–8763 (2021)
17. Ricaurte, P.: Data epistemologies, the coloniality of power, and resistance. *Television & New Media* **20**(4), 350–365 (2019)
18. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: *Proceedings of CVPR 2022*, pp. 10684–10695 (2022)
19. Rouvroy, A., Berns, T.: Gouvernamentalité algorithmique et perspectives d'émancipation. *Réseaux* **177**(1), 163–196 (2013)
20. Salavon, J.: *Every Playboy Centerfold, The Decades (normalized)*. Digital C-prints (2002), broader series begun 1997, <https://salavon.com/work/EveryPlayboyCenterfoldDecades/>, accessed 7 Aug 2026
21. Seers, D.: *Statistical Needs for Development—with Special Reference to National Accounting*. IDS Communication 120. Institute of Development Studies, Brighton (1977)
22. Sekula, A.: The body and the archive. *October* **39**, 3–64 (1986)
23. Selbst, A.D., boyd, d., Friedler, S.A., Venkatasubramanian, S., Vertesi, J.: Fairness and abstraction in sociotechnical systems. In: *Proceedings of FAT\* 2019*, pp. 59–68. ACM, New York (2019)
24. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* **3**(1), 71–86 (1991)
25. Udzu: *Average Flags of the World #2: Mean and Mode Flags Weighted by Population and Area*. Data visualisation (2018), documented at <https://www.visualcapitalist.com/visualizing-average-flag-world/>, accessed 7 Aug 2026
26. Wang, A., Kapoor, S., Barocas, S., Narayanan, A.: Against predictive optimization: on the legitimacy of decision-making algorithms that optimize predictive accuracy. *ACM Journal on Responsible Computing* **1**(1), Article 9 (2024)
27. World Bank: Population, total (SP.POP.TOTL). World Development Indicators, France, Uruguay, West Bank and Gaza (2024), <https://data.worldbank.org/indicator/SP.POP.TOTL>, accessed 7 Aug 2026