

# PoliSim

## *Open Civic Intelligence*

Open-source · AGPL v3 · [github.com/AICap27/polisim](https://github.com/AICap27/polisim)

### Technical White Paper — Methodology and Validation

Version 1.2 — June 2026

[polisim.dev](https://polisim.dev)

#### Abstract

PoliSim is a three-stage open-source computational platform (AGPL v3 licence) for electoral simulation under the Italian mixed proportional-majoritarian system (Rosatellum bis). Stage 1 applies a uniform swing model to official Ministry of the Interior data (Eligendo) across 221 single-member constituencies. Stage 2 refines seat allocation through a Multilevel Regression and Poststratification (MRP) model implemented in PyMC 5.28.4 with Dirichlet-Multinomial likelihood, calibrated on 252 observations across 13 regions. Stage 3 generates demographically targeted communication strategies across three verticals — political consultant, NGO, and civic organisation — integrating psychographic profiles derived from ESS Round 11 (N=2,865), TRIPOL IT (N=1,231), and calibration data from the Meta Ad Library. The backtest on 14 Camera constituencies in Lazio (2022 general election) returns MAE 4.01 pp, RMSE 4.64 pp and winner accuracy 14/14. The empirical CI90 coverage is 39.2% (measured across the full validation dataset, 252 observations): a declared open structural problem. The MRP model outperforms an OLS baseline on all available validation datasets. A deterministic Cognitive Science Layer (five frameworks from the cognitive literature, latency <50 ms) and a factual-grounding verifier complete the message analysis pipeline. Known limitations — under-calibrated CI90, residual CDX bias of -3.33 pp, absence of proprietary polling data — are explicitly declared; LLM self-evaluation bias in the layer is architecturally mitigated through a separate opt-in evaluator.

# 1. Introduction and Motivation

The Italian electoral system introduced by Law 165/2017 — known as the Rosatellum bis — combines proportional and majoritarian seat allocation: 37% of seats in both chambers of Parliament are assigned through single-member constituencies, while the remaining 63% is distributed proportionally among closed lists. This hybrid structure generates non-linear interdependencies between national vote shares, local candidate strength, and coalition composition, making constituency-level projections methodologically complex.

Currently available forecasting tools for Italian analysts and political operators fall into two broad categories: (i) national-level polling aggregators that report voting intentions without translating them into constituency-level seat distributions, and (ii) proprietary models used by major broadcast networks whose methodological specifications are not publicly accessible. A gap therefore persists between freely available polling data and operationally useful seat projections.

PoliSim bridges this gap through a reproducible, open-source (AGPL v3) platform with a public methodology, articulated in three computational stages. The design philosophy rests on three principles: methodological transparency, with all modelling assumptions and known limitations explicitly declared; empirical grounding, using exclusively official institutional sources (Eligendo, ISTAT, ESS) as inputs; and progressive validation, requiring multi-regional backtests before any public claim.

The platform was developed within the Q-Italia (qitalia.org) civic-technology organisation, which uses PoliSim as a live case study. Every piece of content produced requires human approval before publication; autonomous publication is architecturally impossible.

## 1.1 Operational Verticals

---

PoliSim serves three user typologies with distinct functionalities:

- Political consultant / electoral analyst — access to the simulator (Stage 1–2): seat-distribution projection given a vote scenario, dual-majority check, swing analysis by macro-area and constituency.
- NGOs and non-profit organisations — access to the Message Optimizer (Stage 3): evaluation of message value-coherence against the organisation's charter, optimisation per demographic segment and communication channel.
- Civic organisations and labour bodies — access to the Message Optimizer with the civic vertical: communication anchored to a declared value corpus, deterministic charter→message flow, mandatory human approval.

## 1.2 Research Questions

---

The three operational questions of the system:

- Q1 — Seat allocation: Given an observed or hypothetical national vote distribution, what is the expected seat distribution in the Chamber of Deputies and the Senate of the Republic?
- Q2 — Dual-majority check: Does a given scenario produce a stable governing majority in both parliamentary chambers simultaneously?
- Q3 — Message–territory coherence: Is a given message coherent with the organisation's declared values and appropriate for the demographic reality of the target territory?

### 1.3 Scope and Licence

---

This white paper documents the methodology, data sources, and validation results for PoliSim version 2.1, in production in June 2026. The source code is publicly available on GitHub ([github.com/AlCap27/polisim](https://github.com/AlCap27/polisim)) under the AGPL v3 licence. The platform operates as a scenario-analysis tool, not a predictive polling instrument: it translates externally supplied vote distributions into seat distributions with associated uncertainty estimates; it does not autonomously generate vote-intention forecasts.

## 2. Data Sources

PoliSim draws on four primary data sources, all public and institutional. A cardinal principle of the data architecture is exclusive reliance on sources with documented provenance: no proprietary data, no commercial third-party data, no personal data collected or processed by the platform.

### 2.1 Eligendo — Electoral Archive, Ministry of the Interior

The primary Stage 1 input is the official electoral results database managed by the Ministry of the Interior through the Eligendo OpenData platform. For the 2022 general election, the data covers all 221 single-member constituencies (Chamber: 147; Senate: 74), including votes by party, coalition composition, candidate names, and official counts.

Data are downloaded as ZIP archives from the Eligendo portal and processed by `polisim_buildcollegi.py`. A persistent quality issue — documented and managed, not concealed — is the inconsistency in municipality name encoding across dataset versions (UTF-8/Latin-1 mojibake), particularly for bilingual municipalities in Alto Adige and Valle d’Aosta. Match rate achieved: 98.4%. Unmatched records are excluded from the geographic join but retained in aggregate totals.

Anomalies in local coalitions — SVP in Alto Adige, the Misto group in several constituencies — introduce approximately 17 seat discrepancies between calculated totals and officially documented CDX results. This discrepancy is documented and does not affect the validity of the swing model for 97% of constituencies with standard coalition compositions.

### 2.2 ISTAT 2021 Decennial Census

The demographic variables for the MRP model (Stage 2) derive from the ISTAT 2021 decennial census. Data are obtained through bulk download of census-section indicator files and aggregated to constituency level via geographic join with official shapefiles (Electoral Constituencies Geographic Bases, MIT).

Eight demographic variables are in use in the model: percentage of residents aged 15–34, 65+, with tertiary education, female; employment rate; percentage born abroad; percentage of extra-EU origin; plus a binary indicator for election type. A structural limitation: the immigration-related variables are populated with real data only for 14 Lazio constituencies; the remaining 238 constituencies receive imputation to zero. National extension is a priority milestone.

Variable	Description	Source	Coverage
<code>pct_giovani</code>	% residents 15–34 years	ISTAT 2021	All 252 constituencies
<code>pct_anziani</code>	% residents 65+	ISTAT 2021	All 252 constituencies
<code>pct_alta_istr</code>	% with tertiary education	ISTAT 2021	All 252 constituencies
<code>pct_femmine</code>	% female residents	ISTAT 2021	All 252 constituencies
<code>pct_occupati</code>	Employment rate	ISTAT 2021	All 252 constituencies
<code>pct_stranieri</code>	% born abroad	ISTAT 2021	14 Lazio only *

pct_extra_ue	% of extra-EU origin	ISTAT 2021	14 Lazio only *
is_politiche	Binary: general (1) vs regional (0)	Eligendo	All 252 constituencies

\* Non-Lazio constituencies receive an imputed value of zero. Extension to national coverage is planned.

## 2.3 Survey Data — ESS Round 11, TRIPOL IT, Meta Ad Library

Stage 3 draws on three complementary datasets to build psychographic profiles for the supported demographic segments:

- ESS Round 11 (2023–24): N=2,865 Italian respondents, face-to-face probability sample. Variables: institutional trust, left–right self-placement, immigration attitudes, democratic satisfaction. Used as a proxy for post-2022 structural values, not as vote intentions.
- TRIPOL IT (2021–22): N=1,231 Italian respondents. Variables: affective polarisation scores, WAPSV index. Used to calibrate the emotional valence of communication per segment.
- Meta Ad Library (May 2026): 30 NGO advertisements analysed. Used exclusively for tone calibration and hooks in the legacy and CSR segments. Not used for demographic inference.

Declared limitation: TRIPOL coverage for the TERZO\_POLO segment is based on N=44 ESS respondents — a statistically fragile sub-sample. Estimates for this segment must be interpreted with high uncertainty.

## 2.4 Validation Dataset — Regional Results

Model validation draws on official Eligendo results from 13 Italian regions across multiple electoral cycles (2020–2024), for a total of 252 constituency-election observations. The primary validation datasets for the OLS vs. MRP comparison are: Lombardia 2023 (N=23), Lazio 2023 (N=14), Emilia-Romagna 2024 (N=11), Liguria 2024 (N=4). The 14 Lazio constituencies (2022 general election) constitute the primary backtest dataset for point-accuracy metrics.

### 3. Stage 1 — Electoral Swing Model

Stage 1 implements a uniform swing model at constituency level applied to the 221 single-member seats under the Rosatellum. The model translates a national vote distribution — specified by the user or drawn from the integrated polling aggregator (PolitPro, daily automatic update) — into constituency-level outcomes.

#### 3.1 Model Specification

Let  $V^{2022}_c$  be the observed vote share of coalition  $c$  in the 2022 election, and  $V^{sc}_c$  the scenario share. The coalition-level swing is:

$$\Delta_c = V^{sc}_c - V^{2022}_c$$

For each single-member constituency  $i$ , the incumbent winner  $c^*_i$  retains the seat unless the cumulative lead of the strongest challenger exceeds half the observed 2022 victory margin:

$$\text{seat}_i \leftarrow c_{\text{challenger}} \text{ if } \max_{\{c \neq c^*\}} (\Delta_c - \Delta_{\{c^*\}}) > \text{margin}_i / 2$$

The victory margin is derived from official Eligendo 2022 results for the 14 Lazio constituencies and from macro-area averages for the remaining 207 constituencies. This asymmetry constitutes the main known approximation of Stage 1.

#### 3.2 Proportional Allocation

The 63% of seats (245 Camera; 122 Senato) is distributed using the d’Hondt method with a national 3% threshold (Rosatellum Art. 83, c. 1(c)):

- $\text{admitted} = \{p : V_p \geq 3.0\%\}$
- $\text{seats}_p = \text{floor}(V_p / \text{sum}(V_{\text{admitted}}) \times N_{\text{seats}}) + \text{remainder correction}$

Residual seats are assigned by the highest fractional remainder (Hamilton/Hare method). Parties not assigned to any coalition are grouped under ALTRI.

#### 3.3 Macro-Area Structure

Macro-area	Camera seats	CDX 2022	CSX 2022	M5S 2022	Avg. margin	Marginal
North-West	38	28	5	0	12.5 pp	8
North-East	27	22	3	0	14.2 pp	5
Centre *	14	14	0	0	24.9 pp	4
South	36	22	4	8	8.1 pp	14
South-Islands	18	8	2	7	6.4 pp	9

\* Centre macro-area is partially replaced by constituency-level Lazio data in v2.1.

#### 3.4 Dual-Majority Check

Stage 1 simultaneously calculates seat totals for both chambers and explicitly signals three conditions:  $\text{dual\_majority}$  (Camera  $\geq 201$  + Senato  $\geq 104$ ),  $\text{partial\_majority}$  (one chamber only),  $\text{no\_majority}$ . This functionality addresses the distinctive constitutional constraint of the Italian system.

## 4. Stage 2 — Multilevel Regression and Poststratification

Stage 2 applies a Multilevel Regression and Poststratification (MRP) model to refine the seat distribution produced by Stage 1, explicitly accounting for demographic heterogeneity across constituencies. The MRP framework was developed by Gelman and Little (1997), formalised by Park, Gelman and Bafumi (2004), and evaluated for electoral forecasting by Wang et al. (2015). PoliSim adapts the framework to the Italian multiparty bicameral context.

### Technical note —

*Original contribution: to our knowledge, no published literature applies MRP with Dirichlet-Multinomial likelihood to Italian census data (71 million ISTAT 2021 observations) at single-member constituency level, with out-of-sample backtesting on real elections. Documented MRP applications in the literature concern predominantly the US context (states and electoral districts) and the UK (constituencies), typically with individual geo-located survey inputs. The ecological application on aggregated Italian census data constitutes, to date, an undocumented case.*

### 4.1 Likelihood Function

A Dirichlet-Multinomial likelihood is used to model the joint distribution of vote shares across four coalitions. This choice addresses two requirements: (i) vote shares must sum to one; (ii) the model must treat all coalitions symmetrically. Previous specifications with Normal likelihood for CDX margins introduced asymmetric distortions; the move to Dirichlet-Multinomial was motivated by CDX over-prediction observed in preliminary validation sessions.

For constituency  $i$  with total votes  $N_i$ :

$$y_i | \alpha_i, N_i \sim \text{DirichletMultinomial}(N_i, \alpha_i)$$
$$\alpha_{\{i,c\}} = \mu_c + \sum_k \beta_{\{k,c\}} \times X_{\{ik\}} + \gamma_{\{\text{region}[i],c\}}$$

where  $\mu_c$  is the coalition intercept,  $\beta_{\{k,c\}}$  the coefficient of the  $k$ -th demographic variable,  $X_{\{ik\}}$  the standardised ISTAT value,  $\gamma_{\{\text{region},c\}}$  the regional random effect.

### Technical note —

*Probability extraction: final probabilities are extracted as  $p = \alpha / \alpha.\text{sum}()$  (Dirichlet division), not via softmax. The softmax-vs-Dirichlet bug was verified and closed on 07/06/2026.*

### 4.2 Variable Set and Prior Distributions

Variable (z-score)	Description	Prior
pct_giovani_z	% residents 15–34 years	Normal(0, 1)
pct_anziani_z	% residents 65+	Normal(0, 1)
pct_alta_istr_z	% with tertiary education	Normal(0, 1)
pct_femmine_z	% female residents	Normal(0, 1)
pct_occupati_z	Employment rate	Normal(0, 1)
pct_stranieri_z	% born abroad (Lazio only)	Normal(0, 1)
pct_extra_ue_z	% of extra-EU origin (Lazio only)	Normal(0, 1)
is_politiche	Binary: general (1) vs regional (0)	Implicit Bernoulli

*All Normal(0,1) priors are weakly informative, centred on null effect. Dirichlet concentration uses Half-Normal(1) hyperprior. HalfStudentT(v=4) tested on 07/06/2026: no CI90 improvement; not promoted to production.*

The selection of `pct_stranieri_z` and `pct_extra_ue_z` (added in trace v4) was established through systematic backtest analysis: across 14 Lazio constituencies, these two variables show the strongest correlation with MRP prediction error ( $r = -0.800$  and  $r = -0.792$  respectively). Their inclusion reduced MAE by 0.66 pp (from 4.67 to 4.01 pp).

### 4.3 Hierarchical Structure

---

Regional random effects  $\gamma_{\{region,c\}}$  capture structural political differences across macro-regions:

$$\gamma_{\{region,c\}} \sim \text{Normal}(0, \sigma_{region}) ; \quad \sigma_{region} \sim \text{HalfCauchy}(1)$$

The `is_politiche` variable accounts for the systematic difference between national general elections and regional elections (different candidate structures, different turnout, higher M5S shares in some regions). Including regional elections in the training set expands N from 14 to 252, substantially improving the statistical basis of estimates.

### 4.4 Sampling Configuration

---

`draws=2000`, `tune=1000`, `chains=4`, `target_accept=0.90` (default). Convergence diagnostics: maximum  $\hat{r} = 1.000$  across all parameters; 125 divergent transitions (1.6% of post-warmup samples). Divergences increased from 10 (v3) to 125 (v4) with the addition of two new variables — acceptable for applied work, to be monitored. Tuning with `target_accept=0.99` is planned as a future improvement.

## 5. Validation and Backtesting

The validation protocol operates at two levels: (i) point backtest on 14 Camera constituencies in Lazio (2022 general election), and (ii) OLS vs. MRP comparison on out-of-sample regional datasets. Both protocols use exclusively out-of-sample predictions.

### 5.1 Lazio Backtest — 14 Constituencies (General Election 2022)

Lazio was selected as the primary validation dataset because ISTAT 2021 data are available at full census-section resolution, enabling genuine MRP demographic prediction. Results report MRP v4 model predictions against official Eligendo 2022 counts.

Metric	MRP v3	MRP v4	Interpretation
MAE vote share CDX	4.67 pp	4.01 pp	↓ 14.1% improvement
RMSE vote share CDX	6.08 pp	4.64 pp	↓ 23.7% improvement
Bias (mean error) CDX	-3.69 pp	-3.33 pp	Systematic underestimate, reduced
Winner accuracy	14/14	14/14	✓ 100% constituencies correctly called
Empirical CI90 coverage	n.a.	△ 39.2%	Structural problem (see §6.1)

CDX = centre-right coalition (Fdl, Lega, FI). CI90 coverage of 39.2% is measured across the full validation dataset (252 observations, general + regional) with a script verified on 07/06/2026. The 29% figure reported in earlier versions was measured on the Lazio 2022 subset alone using a different method.

Residual analysis by constituency type reveals a systematic geographic pattern: the five Lazio 2 constituencies (predominantly provincial) show higher absolute errors ( $|e|$  mean = 3.1 pp) compared to the nine Lazio 1 Rome-centred constituencies ( $|e|$  = 0.97 pp for central Rome). This heterogeneity motivates the inclusion of sectoral employment variables (DCSC\_CONDPROFOCCUP) in the planned v5 specification.

### 5.2 Comparison with OLS Baseline — Regional Datasets

An OLS regression with the same 8-variable design matrix was estimated on the same training data. Out-of-sample RMSE is computed for available regional datasets:

Dataset	N constituencies	RMSE OLS	RMSE MRP	MRP advantage
Lombardia 2023	23	4.65 pp	4.28 pp	+0.37 pp
Lazio 2023	14	5.36 pp	5.08 pp	+0.28 pp
Emilia-Romagna 2024 △	11	12.09 pp	10.36 pp	+1.73 pp
Liguria 2024 †	4	9.75 pp	8.45 pp	+1.30 pp
Average (4 datasets)	—	7.96 pp	7.04 pp	+0.92 pp

△ Emilia-Romagna outlier: RMSE CSX 15.7 pp — probable cause: local competitive structure diverges from training set (strong civic list, near-absent M5S). † Liguria N=4: indicative result, not statistically robust. OLS baseline uses identical variable set and training data.

The MRP model outperforms the OLS baseline on all available datasets. The average delta of +0.92 pp is lower than that reported in white paper v1.0 (+1.14 pp) because it is calculated on more recent datasets with different composition. For reference, Wang et al. (2015) report MAE ~2–3 pp for MRP models trained on individually geo-located survey data: direct comparison is not appropriate given the absence of such input in PoliSim.

## 6. Declared Limitations

Scientific credibility requires that limitations be declared with the same prominence as results. The following are architectural properties of the current system — not temporary gaps, but structural characteristics that condition the interpretation of all outputs.

### 6.1 Under-Calibrated Confidence Intervals

---

#### **CI90 coverage at 39.2% instead of the nominal 90%.**

Empirical CI90 coverage is 39.2% on the full validation dataset (252 observations), against the nominal expectation of 90%. This means that in approximately 60% of cases the true outcome falls outside the 90% posterior credible interval — severe under-calibration. Confirmed cause: training dataset too small (252 observations), not prior misspecification. Test with HalfStudentT( $\nu=4$ ) prior on 07/06/2026: identical CI90 at 39.2%. The structural solution is dataset expansion with further regional election cycles, not prior modification. Operational guidance: Stage 2 CI outputs are directional uncertainty indicators, not calibrated probabilistic statements. Point estimates (MAP) are the operationally valid outputs.

### 6.2 Systematic CDX Bias

---

#### **Systematic forecast error of -3.33 pp for CDX vote share.**

Persists in trace v4. Concentrated in provincial Lazio 2 constituencies (mean error 4.8 pp) versus urban Lazio 1 (mean error 2.1 pp). Root cause identified in agricultural employment share and small-municipality voting patterns not captured by ISTAT 2021; the DCSC\_CONDPROFOCCUP dataset is a priority for the v5 specification.

### 6.3 Ecological Model — No Proprietary Polling Data

---

PoliSim is an ecological model: it estimates vote shares from demographic and historical variables, not from individually geo-located survey responses. The model cannot autonomously detect shifts in opinion unless they are already visible in the Eligendo historical record or the ISTAT demographic profile. It is a hypothetical-scenario calculator, not a predictive tool. Integration with PolitPro polling aggregates (active daily scraper) allows answering ‘given today’s polls, what would the seat distribution be?’ but remains contingent on the accuracy of those polls.

### 6.4 Partial ISTAT Coverage

---

pct\_stranieri\_z and pct\_extra\_ue\_z are populated with real data for only 14 of 252 constituencies. The full 8-variable model is effectively a 6-variable model for 94% of constituencies. RMSE improvements from v4 variables are Lazio-specific and may not generalise nationally. Extension requires downloading ISTAT files for all 20 regions — engineering work estimated at 5–8 sessions.

### 6.5 Stage 3 — Unvalidated Calibration Weights

---

Stage 3 weights were calibrated by domain experts on the basis of ESS/TRIPOL microdata and Meta Ad Library analysis, but have not yet been validated through controlled A/B tests. Outputs must be treated as structured hypotheses to be tested. A validation protocol with organisational partners and standardised outcome measurement is being actively recruited.

### 6.6 LLM Self-Evaluation Bias in the Language Layer

---

The Message Optimizer generates message variants and annotates them across the four Entman dimensions (problem, cause, moral evaluation, solution) in the same LLM call. Ziems et al. (2023, Computational Linguistics, doi:10.1162/coli\_a00502) document that language models exhibit systematic confirmation bias in self-annotation of framing: they tend to confirm the frame just produced.

Implemented mitigation (June 2026): a separate AI evaluator (`cognitive_evaluator.py`, cf. §8) performs framing evaluation in a distinct call. The evaluator receives the text as if written by a third party, does not know it was generated, and receives deterministic measurements as refutable features — never as conclusions to confirm. The evaluator is opt-in for latency and cost reasons; when active, the `framing_eval_method` field in the log changes from 'self' to 'independent'. Self-evaluation remains the declared default and the residual limitation is documented for every output that does not activate the evaluator.

## 7. Stage 3 — Message Optimizer

Stage 3 generates message strategies for civic and non-profit organisations, calibrated on the demographic and psychographic profile of the target segment and anchored to the organisation’s declared value corpus. The component operationalises Entman’s (1993) framing theory through a prompt-engineering pipeline built on the Anthropic Claude API. The AI layer is replaceable by design: any instruction-following LLM (Mistral, LLaMA, Phi, local models via Ollama) can replace the current provider without modifications to the surrounding infrastructure.

### 7.1 Three Operational Verticals

- Political consultant — access to the electoral simulator (Stage 1–2). The Message Optimizer is not the primary component for this vertical.
- NGOs and non-profit organisations — Message Optimizer with the ong vertical: value-coherence evaluation, optimisation for 6 channel types, 5 engagement levels (prospect → legacy\_prospect), deterministic engagement\_level × domain → communicative\_function mapping.
- Civic organisations and labour bodies — Message Optimizer with the civic vertical: communication anchored to the organisation’s charter, 5 thematic sub-verticals (collective bargaining, assembly, institutional communication, membership, welfare/services). Every output requires human approval before publication — architecturally non-bypassable.

### 7.2 Domain-Agnostic Architecture

Stage 3 contains no hardcoded values. The reference value corpus is supplied by the user organisation through the org\_corpus field (charter, principles document, manifesto). The system evaluates message coherence against that specific corpus — not against values pre-defined by the platform. This is the architectural differentiator relative to commercial tools: the platform is a mirror of the organisation’s values, not an alternative value system.

### 7.3 Engagement × Domain → Communicative Function Mapping

Domain	prospect	supporter	recurring_donor	legacy_prospect	at_risk
ong_lasciti	educate_cause	approach_theme	cta_legacy	conv_legacy	reactivation
ong_raccolta	educate_cause	cta_donation	recognition	recognition	reactivation
ong_fidelizzazione	recognition	recognition	recognition	celebrate_rel	reactivation
ong_advocacy	mobilisation	mobilisation	mobilisation	mobilisation	mobilisation
ong_corporate	institutional	institutional	institutional	institutional	institutional
ong_volontariato	educate_cause	direct_engage	direct_engage	direct_engage	direct_engage

### 7.4 Supported Channels

Channel	Output structure	Key constraints
email	SUBJECT: + BODY:	subject max 8 words, body 3–4 sentences, CTA at close
social	single text	hook in first 5 words, max 280 characters

flyer	HEADLINE: + SUPPORT:	headline max 6 words, support max 2 sentences
speech	OPENING: + BODY: + CLOSING:	spoken language, collective appeal at close
video	HOOK: + TEXT:	hook max 8 words ×3s, text max 25 words
press_release	TITLE: + TEXT: + QUOTE:	inverted pyramid, institutional tone

## 7.5 Psychographic Profiles — Supported Segments

Segment	N (survey base)	Primary source	Key attitudinal axes
Young precarious South	N≈380 ESS R11	ESS R11 + TRIPOL	Institutional distrust, emigration intent
Unemployed housewives South	N≈290 ESS R11	ESS R11	Recognition of care work, health access
Workers/craftsmen North	N≈310 ESS R11	ESS R11 + TRIPOL	Automation risk, tax burden
Urban progressive graduates	N≈420 ESS R11	ESS R11	Democratic quality, climate, housing costs
Value abstentionists	N≈180 ESS R11	ESS R11 + TRIPOL	Systemic cynicism, radical discontinuity
Pensioners Centre-North	N≈340 ESS R11	ESS R11	Healthcare, security, party loyalty
Legacy donors	N≈85 ESS R11	ESS R11 + Meta Ads	Heritage, institutional trust, death taboo
CSR managers	Expert-calibrated	Meta Ads library	ESG accountability, reputational ROI
Third Pole moderates	N≈44 ESS R11	ESS R11 $\triangle$	Small sub-sample — high uncertainty
SME entrepreneurs	N≈95 ESS R11	ESS R11 + TRIPOL	Regulatory burden, credit access

ESS R11 = European Social Survey Round 11 (2023–24), Italian sub-sample N=2,865. TRIPOL = Italian affective polarisation study (2021–22), N=1,231.

## 7.6 Empirical Integration and Prompt Structure

The segment’s psychographic profile (`dati_empirici`) and the organisation’s value corpus (`org_corpus`) are injected directly into the prompt at runtime, loaded from `itanes_profilo_enriched.json`. For the legacy and CSR segments, an additional `calibrazione_meta_ads` block is injected.

Every output includes: (i) primary message frame with explicit citation of the demographic trigger; (ii) secondary frame addressing the likely objection; (iii) channel recommendation; (iv) explicit list of elements to avoid. All outputs include the declaration that weights are expert-calibrated and not yet validated on A/B tests (cf. §6.5).

## 7.7 Three-Body Architecture and Factual Grounding

The prompt architecture follows a three-body separation: the value corpus (`org_corpus` — charter, principles) defines what coherence is evaluated against; the user’s message defines what to communicate; the optional factual corpus (`dati_progetto` — the project’s numbers, facts, dates) defines the permitted evidence base. The factual corpus is injected into prompts as an absolute constraint: variants may cite only data present in the source.

A deterministic grounding verifier (latency <5 ms, run at the pipeline convergence point) checks that every number appearing in generated variants exists in the supplied factual corpus. The verifier normalises separators (19,000 ↔ 19000), numeric word forms ('two million' →  $2 \times 10^6$ ), percentages, and multipliers. Each variant receives a grounding outcome and an explanatory note; an empty source renders the check non-applicable, never blocking.

At the first test on real cases, the verifier caught three numerical hallucinations generated by the language model (figures not present in the source). This empirically confirms the necessity of the control: LLMs introduce plausible but unsupplied numerical facts, and a system intended for civic organisations cannot delegate factual verification to trust in the model.

## 8. Cognitive Science Layer

The Cognitive Science Layer is an analysis component that evaluates how a message is cognitively processed by the declared audience — distinct from the Message Optimizer, which evaluates what to communicate and with what value coherence. The layer is atomic and composable: it is integrated as an autonomous endpoint (`/api/cognitive-profile`), as automatic enrichment of every Message Optimizer variant, and as a client-side module.

The layer never alters the communicative intent and does not suggest techniques for circumventing the recipient's critical thinking — an ethical constraint encoded in the recommendation synthesiser: suggested interventions concern exclusively clarity, evidence, coherence, and balance; never emotional escalation or bias exploitation.

### 8.1 Five Frameworks from the Cognitive Literature

Framework	Foundational reference	What it measures
Cognitive Load	Sweller (1988)	Cognitive load: readability (Gulpease), sentence complexity, lexical density, abstractness
Dual Process	Kahneman (2011)	Balance System 1 (emotional-intuitive) vs System 2 (rational-deliberative)
Elaboration Likelihood	Petty & Cacioppo (1986)	Expected persuasion route: central (arguments) vs peripheral (surface cues)
Cognitive Fluency	Reber et al. (2004)	Perceived processing ease and its effect on truth judgement
Framing	Tversky & Kahneman (1981); Entman (1993)	Gain/loss frame; episodic vs thematic structure

### 8.2 Two-Level Architecture

**Deterministic level (always active, synchronous, <50 ms):** Implemented in Python (`cognitive_layer.py`) with a JavaScript port at verified parity for client-side use. Calculates Gulpease index, sentence statistics, lexical density, S1/S2 ratio via dedicated Italian wordlist (~600 terms, single shared file), claim/evidence balance, provisional ELM signals, composite fluency, gain/loss and episodic/thematic frame. A language guard rejects non-Italian text. The primary output is a synthetic recommendation.

**AI-assisted level (opt-in, separate call, 1.5–4 s):** The evaluator (`cognitive_evaluator.py`) determines the definitive ELM route, confirms or refutes the S1/S2 classification, and — architecturally critical — performs Entman annotation independently from generation, mitigating the self-evaluation bias documented in §6.6. The provider is replaceable (Claude, Mistral, local models via Ollama).

### 8.3 Input Requirements and Declared Constraints

- Audience never inferred: the backend rejects the request (HTTP 422) if the audience is not explicitly declared. The system never guesses the recipient.
- Gulpease unstable below 10 words: explicit flag `gulpease_reliable=false` — the system never blindly assigns scores to texts that are too short.
- Composite load weights declared, not calibrated: the 40/25/15/20 weighting (readability/sentences/density/abstractness) is expert-calibrated. Empirical

calibration requires data from the 20-campaign validation protocol — the same declared constraint as for Stage 3 weights.

- Deterministic ELM is signal only: the field is named `provisional_route`; definitive classification belongs to the AI evaluator when active.

## 8.4 Preliminary Validation

---

The layer was tested on eight real messages published by the validation organisation over three months. All tests pass: emotionally and episodically dominant messages are correctly classified S1/pure\_assertion; argumentative messages are classified S2/claim\_evidence; non-Italian text is correctly rejected by the language guard. Python–JavaScript parity is verified to the decimal on the Gulpease index. These results constitute functional validation, not empirical validation of predictive effectiveness — a distinction maintained throughout the documentation.

## 9. Propensity Module — RFML

The Propensity Module is a scoring engine for donors and members that calculates individual propensity across five organisational objectives: testamentary bequest, donation upgrade, transition to recurring support, reactivation, one-off emergency donation. The module operates entirely in client-side processing (browser): no personal data transits or is retained on the server — a GDPR-by-design property, not a subsequent addition.

### 9.1 RFML Signals and Redefinition for the Civic Context

The RFML framework (Recency, Frequency, Monetary, Longevity) is adapted to the civic organisation context, where standard signals require substantial redefinition:

Signal	Standard definition	Civic redefinition
R — Recency	Date of last transaction	Last donation / last service desk visit / last case handled
F — Frequency	Number of transactions in period	Service use (patronato, CAF, legal advocacy) + assembly participation. NOT the membership fee: in union contexts the fee is deducted automatically from payroll — administrative data, not an engagement signal
M — Intensity	Average donation amount	Average donation amount (NGO) / number of service accesses (union). Not the fee amount (fixed by contract)
L — Longevity	Relationship seniority	Length of membership or relationship with the organisation — historically the most stable signal

*The F/M distinction for the union context is architecturally critical: using the fee as an engagement proxy would produce identical scores for all active members, rendering the model useless. Service usage signals are the functionally correct substitute.*

### 9.2 Five Objectives and Scoring Structure

For each record in the membership database, the module calculates five independent scores (0–100):

- `prop_legacy` — testamentary bequest propensity: highest weight on L (longevity), `M_level`, `legacy_communications_received`, senior age bracket.
- `prop_upgrade` — donation upgrade propensity: highest weight on `M_trend` (amount growth), ISTAT territorial income, seniority.
- `prop_recurring` — transition to recurring donation propensity: weight on F (payment consistency) and current payment method.
- `prop_reactivation` — inactive donor return propensity: weight on R (inverted recency), historical amount, acquisition channel.
- `prop_one_off_emergency` — single emergency donation propensity: weight on previous campaign response, average age bracket.

Weights per objective were calibrated from Sargeant (2008) on donor lifecycle management literature and adapted to the Italian context. Not yet validated on real campaigns — cf. §9.4.

### 9.3 ISTAT Territorial Layer

The CAP field in the membership database (optional) enables a territorial multiplier that corrects the base score according to the demographic and economic profile of the province. The first two digits of the postcode map the region; ISTAT 2021 data used: percentage over-65, average household income, Gini index.

Objective	Over65	Avg. income	Gini index	Correction range
prop_legacy	+0.50	+0.30	-0.20	±15%
prop_upgrade	0.00	+0.60	-0.40	±10%
prop_recurring	-0.10	+0.60	-0.30	±10%
prop_reactivation	+0.05	+0.60	-0.35	±5%
prop_one_off_emergency	+0.15	+0.50	-0.35	±12%

*Territorial correction examples for prop\_legacy: Campania -6%, Calabria -3%, Tuscany +4%, Lombardia +3%. Territorial weights are reasoned hypotheses based on fundraising literature — not validated on real Italian data.*

## 9.4 Validation Metrics — Synthetic Dataset

Metric	Value	Notes
Legacy lift (top 20% vs rest)	1.65×	Top 20% by score → 25.8% declared legacies vs 15.6% in the rest
Legacy prospects identified	53/600	8.8% of the donor base
Base dataset	Synthetic ⚠	Not validated on real campaigns
Territorial calibration	Active	Based on ISTAT 2021 regional data

**⚠ All metrics are calculated on a synthetic dataset.**

Real lift on Italian campaigns may differ significantly. Validation on real data is a necessary condition before any operational use of metrics as benchmarks.

## 9.5 Data Shield — Integrated GDPR Anonymisation

The Propensity Module integrates an anonymisation tool (Data Shield) that mandatorily precedes data loading. Data Shield operates entirely in the browser (zero server transmission) and offers four operations per column: REMOVE, HASH (SHA-256 one-way), GENERALISE (age bracket, geographic area), KEEP. It includes auto-detection of sensitive patterns (fiscal code, IBAN, email, phone number) and produces a compliance report referencing relevant GDPR articles (Art. 4.5 pseudonymisation, Art. 89 archiving, Art. 25 privacy by design).

## 9.6 Declared Limitations

- Weights not validated on real campaigns: RFML coefficients and territorial multipliers are expert-calibrated from literature (Sargeant 2008; Sargeant & Shang, 2011). Not validated on A/B tests or real Italian campaigns.
- Lift on synthetic dataset: the 1.65× lift is calculated on artificially generated data. The real value on Italian donor bases is unknown until partner validation.
- No feedback loop: the module does not automatically update weights based on campaign outcomes. Recalibration is manual.
- Client-side RFML only: processing is client-side for GDPR reasons. For donor bases exceeding ~50,000 records, browser performance may degrade. A server-side version with extended GDPR documentation is planned.

## References

- Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282), 220–224. doi:10.1126/science.aac9213
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x
- Entman, R. M. (2003). Cascading activation: Contesting the White House’s frame after 9/11. *Political Communication*, 20(4), 415–432. doi:10.1080/10584600390244176
- Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23(2), 127–135.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. doi:10.1214/ss/1177011136
- Ghitza, Y., & Gelman, A. (2013). Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups. *Political Analysis*, 21(1), 98–116. doi:10.1093/pan/mps025
- ISTAT (2021). 15° Censimento generale della popolazione e delle abitazioni. <https://www.istat.it>
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Lax, J. R., & Phillips, J. H. (2009). How should we estimate public opinion in the states? *American Journal of Political Science*, 53(1), 107–121. doi:10.1111/j.1540-5907.2008.00360.x
- Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, 58(2), 258–279. doi:10.1111/j.1460-2466.2008.00384.x
- Ministero dell’Interno (2022). Eligendo — Archivio storico delle elezioni. <https://elezioni.interno.gov.it/opendata>
- Nickerson, D. W., & Rogers, T. (2010). Do you have a voting plan? *Psychological Science*, 21(2), 194–199. doi:10.1177/0956797609359326
- Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification. *Political Analysis*, 12(4), 375–385. doi:10.1093/pan/mp04
- Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. *Advances in Experimental Social Psychology*, 19, 123–205. doi:10.1016/S0065-2601(08)60214-2
- Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure. *Personality and Social Psychology Review*, 8(4), 364–382. doi:10.1207/s15327957pspr0804\_3
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2, e55. doi:10.7717/peerj-cs.55
- Sargeant, A. (2008). Donor retention: What do we know and what can we do about it? *Nonprofit and Voluntary Sector Quarterly*, 37(4), 624–643. doi:10.1177/0899764008316101
- Sargeant, A., & Shang, J. (2011). *Growing philanthropy in the United States: A summary report*. Philanthropy Leadership Council.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. doi:10.1207/s15516709cog1202\_4

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458. doi:10.1126/science.7455683

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980–991. doi:10.1016/j.ijforecast.2014.06.001

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2023). Can large language models transform computational social science? *Computational Linguistics*, 50(1), 237–291. doi:10.1162/coli\_a00502