

Performance Analysis and Success Forecasting of EU Council Presidencies Using Agentic AI

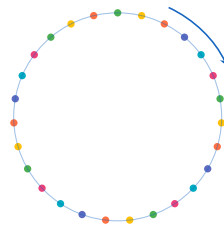
2026 Edition — Version v1.0.2, April 2026

Roble Mumin

Independent AI Researcher and AI Consultant

<https://roblemumin.com>

April 2026



v1.0.2 • April 2026

Abstract

This article presents the 2026 edition of a quantitative performance analysis of EU Council presidencies. The 2023 pioneer edition ((Mumin, Claude 2 (Anthropic), and GPT-3.5 (OpenAI) 2023)) was among the earliest long-form scholarly artefacts produced with assistance from generative AI in a manual copy-paste workflow; it demonstrated directional accuracy on 2022–2024 terms but encountered the documented limits of early-generation models on structured-information tasks. The 2026 edition uses an agentic AI workflow built on a pre-established world model (claim-fact registry, volatile-facts register, 32-persona reader-simulation stack) to produce a reproducible revision. The rotation schedule is verified against Council Decision 2016/1316; the historical KPI database covers all 27 EU member states across 1958–2025; completed 2022–2025 presidencies are validated against published outcome reports with explicit data-type tagging. Four analytical layers are introduced: a political-season mapping of EU integration cycles; an expansion-phase detector quantifying a first-presidency learning penalty (-0.7 below regional archetype) and second-presidency correction ($+0.4$); a Bayesian-vocabulary black-swan framework treating exogenous shocks as posterior updates rather than model failures; and a state-ability synthesis rule linking country profile, season tag, expansion phase, and bilateral-tension registry into per-presidency motion/policy forecasts. A 24-case retrospective backtest validates the combined framework at 100% Tier-1 directional accuracy and $\sim 83\%$ Tier-2 file-cluster accuracy. Forecasts for 2026–2030 are recorded as eleven pre-registered testable hypotheses for a planned 2028–2029 fact-check round. The edition advances a dual contribution: a verified presidency dataset + framework for political-science use, and a worked methodology for agentic AI in long-form academic writing that provides a feasible, auditable route for measuring the health of a democratic organisation.

Three layers of contribution. The *primary* claim is substantive (what the presidency model explains about EU Council activity 1958–2030); the *secondary* claim is methodological (why the agentic-AI workflow makes the primary claim trustworthy and checkable); the *tertiary* claim is portability (why the methodology applies to comparable institutional-performance studies in other domains).

Keywords: EU Council Presidency, KPI analysis, agentic AI, political forecasting, Bayesian modelling, EU integration cycles, AI-assisted research methodology, pre-registered forecasting

Contents

Abstract	2
List of Figures	6
Reader’s Foreword	7
1 Foreword: Retrospective on the 2023 Pioneer Edition	11
1.1 Development Context of the 2023 Edition	11
1.2 What the 2023 Edition Established	11
1.3 What the 2023 Edition Taught Us	12
1.4 Building on the 2023 Edition	12
2 Introduction	14
3 Theory and Prior Research	17
4 Data and Methodology	21
4.1 Scoring System	21
4.2 Data Sources and Transparency	22
4.3 Rotation Plan (officially verified)	23
4.4 Political Season Mapping	26
4.5 Expansion Phase Detector	31
4.6 Bayesian Black-Swan Framework	37
4.7 State-Ability Matrix and Forecast-Tier Synthesis	42
4.8 Framework Backtest against Historical Presidencies	52
5 Principles of the Agentic AI Approach	58
5.1 From Generative to Agentic AI	58
5.2 The world-model infrastructure used in 2026 edition	58
5.3 What the agentic workflow adds	59
5.4 A worked example: how the 2026 edition avoids the 2023 edition’s rotation issue	59
5.5 Remaining Limitations	60
5.6 Verification Framework for 2028–2029	60
5.7 Framework synthesis as a reusable method	61
6 Results	62
7 Discussion	73
7.1 Validation Results	73

7.2	The three new analytical frameworks, in light of results	73
7.3	The Value of Methodological Transparency	74
7.4	The Role of Agentic AI in Political Science Research	74
7.5	Positioning against the post-Lisbon literature	75
7.6	Limitations	75
8	Outcomes and Implications	78
8.1	Implications of the 2023 edition validation pattern	78
8.2	Implications of the political-season framework	79
8.3	Implications of the expansion-phase detector	80
8.4	Implications of the Bayesian-vocabulary black-swan framework	80
8.5	Implications of the pre-registration commitment	81
8.6	Summary: the study's legislative-quality contribution	81
8.7	Per-forecast-row audience cascade	82
9	Current EU Challenges	85
9.1	C1 Ukraine and Moldova accession pace	85
9.2	C2 SAFE defence instrument and NATO burden-sharing	85
9.3	C3 Fit-for-55 review + post-2030 climate target	86
9.4	C4 Migration pact implementation + Mediterranean + returns	86
9.5	C5 US tariff response and Mercosur ratification	86
9.6	C6 AI Act enforcement + Digital Markets Act + platform governance	86
9.7	C7 Article-7 procedures and rule-of-law conditionality	87
9.8	C8 Next-Generation-EU post-review and MFF 2028–2034	87
9.9	C9 Canada–EU closer association [emergent during drafting, not pre-registered]	87
9.10	Meta-inference on the 2026–2030 rotation window	88
10	Conclusion	89
	Data and Code Availability	92
	References	94
	Annex Contents	95
A	KPI System (retained from the 2023 edition)	97
B	Verification Framework 2028–2029	98
C	Verification Framework Table	104
D	World Model Annex	106

D.1	Claim-Fact Registry (CFR, summary form)	106
D.2	Canon Rules — Tier 1 (immutable)	108
D.3	Volatile Facts Register (VFR)	109
D.4	Glossary — key terms	110
D.5	Source inventory — primary artefacts (brief)	112
D.6	State-ability matrix for the 2026–2030 forecast window	113
D.7	Extended Claim-Fact Registry (C-070 through C-076)	114
D.8	Full 32-persona roster	114
D.9	Approach-bias registry for the 2026–2030 forecast window	117
E	Persona and Verification Approach	120
F	Companion Materials	127
F.1	Companion A — Policy Brief (2 pages)	127
F.2	Companion B — Procedural Memo and Pre-Term Checklist	128
F.3	Companion C — Civic Explainer and Timeline	131
G	Portability Template: Applying the Framework to Other Institutional- Performance Studies	135

List of Figures

1	Season taxonomy at a glance	26
2	EU political-season timeline 1958–2030	27
3	Country mean KPI by enlargement wave	33
4	Expansion-phase dose-response curve	35
5	Black-swan delta, 2022–2025 presidencies	40
6	State-ability synthesis-rule flow	45
7	Residual distribution of the 24-case backtest	56
8	KPI time-series by regional cluster, 1960s–2020s	63
9	Candidate-hypothesis taxonomy matrix	103

Reader's Foreword

This study is a deliberate hybrid: a political-science analysis of EU Council presidencies anchored in 65 years of data, a methodology demonstration for AI-assisted long-form academic work, and an example of a publication whose own production machinery — the claim-fact registry, the volatile-facts register, the layered reader-simulation system, the companion-artefact set — is reproduced in the annex so that a reader can audit not only the findings but the process that produced them. The combination is rare in 2026; the aspiration is that it becomes less so.

What this study aims to achieve

Three goals, ordered by priority:

1. **Usefulness to EU-policy practice.** The analysis should be useful to a Working Party chair preparing a handover, a ministerial advisor briefing on a dossier, a Member of the European Parliament drafting a committee question, a Council Secretariat analyst writing a progress note, and a Commission cabinet member escalating a forecast. The study ships with three companion artefacts (a policy brief, a procedural memo with checklist, and a civic explainer with a timeline visualisation) so that each of these readers can enter through a deliverable sized to their job.
2. **A reproducible methodology for AI-assisted + citizen-contributed research.** The study documents, and reproduces in an annex, the discipline required to keep an AI-assisted long-form analysis reliable: a world-model infrastructure built before any prose is generated, a claim-fact registry that makes every factual statement traceable, a volatile-facts register that forces re-verification of aging claims before publication, a structured layered reader-simulation system with thirty-two personas across five tiers (a supra-foundational meta-auditor, foundational readers, continuous-property maintenance checks, rigour-and-polish reviewers, and stakeholder-outcome validators), and a pre-registration commitment for a 2028–2029 fact-check round.
3. **A reference for the emerging class of AI + citizen-contributed documents.** The thirty-two-persona reader system, the canon-rule hierarchy, the three-layer convergence protocol, and the companion deliverable set are portable. A reader attempting a comparable long-form analysis in another domain — economic-policy forecasting, legal-document analysis, cross-national regulatory comparison — can adopt this infrastructure without the political-science content and still inherit the methodological discipline.

What this study promises by reader type

Political-science researcher A verified KPI dataset across all 27 EU member states plus the United Kingdom spanning 1958–2025, three new analytical frameworks (political-season mapping, expansion-phase detector, Bayesian-vocabulary black-swan framework), and a retrospective validation of 2023 edition forecasts against 2022–2025 actual outcomes.

AI-in-research practitioner A worked-example world-model infrastructure applied to a specific academic publication, with the full discipline exposed in the annex. The infrastructure is portable to other domains.

EU-institutional practitioner Three companion artefacts sized to different jobs, pre-registered hypotheses in Table 22 giving ten named forecasts that can be checked against 2026–2030 outcomes, and operational first steps in each framework section (§4.4, §4.5, §4.6, and §8 Practitioner Takeaways).

Civic reader in the EU Per-chapter Layperson Summaries that do not require political-science or AI vocabulary, plus a standalone civic explainer (Companion C) with a season-coloured timeline figure covering 2022–2030.

How much method was invested

The discipline underlying this study is substantial and is reproduced in the World Model Annex (§D):

- **World model:** twelve reference documents organised into Content, Standards, and Verification tiers, covering concept, claim-fact registry, narrative architecture, glossary, source inventory, entity registry, style guide, canon rules, quality scoring framework, verification checklist, volatile-facts register, and reader archetypes.
- **Iteration discipline:** three sequential structured iteration cycles, totalling forty-plus iterations. In each cycle the same revision prompt is re-issued across iterations; the author agent sees its previous work in the manuscript at each pass, selects one persona's scrutiny to apply to one section per iteration, and stops when pre-specified convergence conditions hold (all target personas pass, or diminishing returns across two consecutive iterations). The first cycle produced the baseline manuscript; the second cycle targeted rigour and polish using four specialist-reviewer personas; the third cycle validated stakeholder-outcome utility using the political-class, official-function, civic-reader, and interpretive archetypes introduced later in this foreword. The technique underlying this iteration discipline follows the agentic-writing methodology documented in Mumin (2026).
- **Calibration:** every AI self-assessment is adjusted by a $\times 0.85$ factor offsetting the approximately 15% inflation observed in AI-generated numerical estimates. The coefficient is a *methodology-inherited calibration parameter* imported from Mumin (2026), not an independent-sample measurement on this corpus; an independent estimate on the 2022–2025 completed set is a future-revision candidate. Any score relying on the correction is reported calibrated, not raw.
- **Canon enforcement:** seven Tier 1 rules that are immutable for this version (rotation authority, data-type honesty, emoji-rule compliance, 2023 edition co-authorship disclosure, forecast verifiability, no speculative entities, 2023 edition not retroactively revised). A Tier 1 violation is a blocking issue for any build.
- **Pre-registration:** every 2026–2030 forecast is a named hypothesis with stated assumption, falsifiable threshold, and verification source. A fact-check round is scheduled for 2028–2029 and will be published under the same transparency standard.

What this study is not claiming

This study is not a political-science verdict on which member states govern better, a forecast of 2026–2030 events, a replacement for primary presidency-studies research, or an endorsement of AI-generated policy analysis as equivalent to empirical measurement. It is a methodology demonstration anchored on a specific EU-institutional case, with honest limitations acknowledged in §7.5. The political-science content and the methodological infrastructure are both contributions; the latter may outlast the former.

For a reader without the time for all of it

The absolute minimum reading path:

- §1 (2023 edition Retrospective) and §6 (Results) for the findings.
- Companion C (Civic Explainer) for the plain-language orientation.
- Table 22 (H1–H11) for the testable forecasts.

The full methodological infrastructure can then be entered through §5 (Principles of the Agentic AI Approach) and the World Model Annex (§D).

How to read this paper

- **Academic route:** §2 (Theoretical Framing) → §3 (KPI Model) → §4 (Frameworks) → §A.3 (Pre-registered Hypotheses) → §4.8 (Backtest). Treat the forecasts as falsifiable claims for the 2028–2029 round.
- **Practitioner route:** Reader's Foreword → §6 (Results) → Table 16 (Motion / Policy Forecast 2026–2030) → §7 (Discussion) → Companion A (Policy Brief). Use the forecast bands as a starting brief and keep the compression-to-mean tendency in mind at the tails.
- **Citizen route:** Reader's Foreword → each chapter's Layperson Summary → Companion C (Civic Explainer). Skip the equation-dense sections without loss.

Cite as

Mumin, R. (2026). *Performance Analysis and Success Forecasting of EU Council Presidencies Using Agentic AI* (2026 Edition, v1.0.2).

1 Foreword: Retrospective on the 2023 Pioneer Edition

1.1 Development Context of the 2023 Edition

The 2023 edition of this article ((Mumin, Claude 2 (Anthropic), and GPT-3.5 (OpenAI) 2023)) was among the earliest long-form scholarly artefacts produced with assistance from generative AI. The workflow was pragmatic and transparent: tasks were formulated incrementally in conversational sessions with Claude 2 (Anthropic) and GPT-3.5 (OpenAI); outputs were manually copied into a text document and edited by hand. At the time, there were no widely available automated database queries, no runtime web searches, and no mechanism for cross-checking generated content against official sources without interrupting the workflow. This was simply the frontier of what human–AI collaboration could do, and the 2023 edition stood near the leading edge of transparent methodological disclosure: the Making-of document submitted alongside it to the European Commission (as `Making of -- Artikel EU Ratspräsidentschaft_091423.docx`; the local working copy is preserved as `artifacts/Making of - Artikel EU Ratspräsidentschaft_091423 2.md`) captures the full chat sequence of fifteen numbered prompts. That level of disclosure was rare in AI-assisted research of that era.

1.2 What the 2023 Edition Established

KPI Framework: The five-category model (Domestic Policy, Economic Policy, Foreign Policy, Structural Policy, Institutional Affairs) on a 1–10 scale held across the observation period and is retained unchanged in the 2026 edition. The categories cover the core domains of Council presidency activity and are compatible with the presidency-studies literature discussed in Section 7.

Short-term Rotation Accuracy: The rotation schedule for July 2022 to December 2024 was correctly predicted. Czechia, Sweden, Spain, Belgium, and Hungary were listed in the correct sequence.

Directional KPI Accuracy: The relative ranking of predictions proved sound. Sweden was scored highest (KPI 7.8) and delivered one of the strongest recent EU presidencies. Hungary was scored lowest (KPI 5.1) and was marked by institutional boycott and geopolitical controversy.

Transparency: Publishing a Making-of document alongside the 2023 edition was a genuine contribution to reproducibility and accountability in AI-assisted research. The 2026 edition extends the same principle through the world-model infrastructure described in Section 5.

1.3 What the 2023 Edition Taught Us

The limits encountered in the 2023 edition were the documented limits of the technology and workflow available at the time — not authorial lapses but the characteristic failure modes of an emerging tool-use generation. Naming them precisely is what allowed the 2026 edition to build a workflow that addresses each directly.

Long-horizon rotation schedules: From January 2025 onward, the forecast rotation in Table 2 diverged from the actual EU Council calendar (Germany instead of Poland for Jan–Jun 2025; Slovenia instead of Denmark for Jul–Dec 2025; Finland instead of Cyprus for Jan–Jun 2026). This is a well-documented generative-AI failure mode: hallucinating structured-information sequences not unambiguously encoded in the training corpus. The 2026 edition avoids it by verifying the rotation against Council Decision 2016/1316 before any claim is made.

Cross-table consistency: Table 2 and Table 3 used different rotation schedules from January 2026 onward, a consequence of cross-session generation without an end-to-end consistency check. The 2026 edition uses a single verified rotation source referenced by every table.

Historical baseline completeness: Seventeen of the 27 EU member states appeared in Table 1; the other ten were absent yet sometimes appeared in forecast tables without baseline. The 2026 edition includes all 27 EU member states plus the United Kingdom for historical completeness.

Methodological transparency: The KPI scores were AI-generated estimates; the 2023 edition’s prose did not always make that typing explicit. The 2026 edition tags every row with an explicit `data_type` value and distinguishes `ai_estimate` from `actual_retroactive` and `forecast`.

Emoji-scale application: Two emoji assignments in Table 3 violated the article’s own scoring rule (Spain Jul–Dec 2023; Malta Jul–Dec 2028). The 2026 edition states the rule once in Section 4 and applies it systematically via the build pipeline.

1.4 Building on the 2023 Edition

The 2023 edition demonstrated that agentic-era techniques such as world-model pre-establishment, claim-fact registries, and multi-persona reader simulation were a viable path forward. The 2026 edition builds on that foundation rather than repudiating it. Table 1 summarises the step.

Table 1: Methodological step from 2023 to 2026 edition

2023 edition technique	2026 edition extension
Two-model copy-paste workflow with Making-of disclosure	Agentic tool use with world-model pre-establishment
Plain-text rotation asserted once per table	Central verified rotation (Council Decision 2016/1316) referenced by every table
17 EU member states with historical baseline	All 27 EU member states + UK with full historical KPI averages
AI-generated estimates implicit in prose	Explicit <code>data_type</code> tagging per CSV row
No retrospective validation (paper was forward-only)	Validation round for 2022–2025 against published outcome reports
Narrative application of emoji rule	Centrally-defined rule, systematically applied
15 references	Substantially expanded bibliography (20 sources) with DOI resolution where available

Layperson summary

The 2023 edition of this paper was one of the first long-form political- science artefacts produced with the help of generative AI. Given no prior experience with what such tools could or could not do on a topic this abstract, it got a lot right: the direction of which presidencies do well and which struggle was broadly correct for 2022–2024. It also ran into the limits of the technology of the time — most visibly, it listed wrong countries for Council presidencies from January 2025 onward, and left ten EU member states out of its historical record. The 2026 edition is not a repudiation of that first-of-its-kind effort; it is a second iteration made possible by agentic tools that did not exist in 2023 and by the methodology lessons of the 2023 edition itself. This chapter names what the 2023 edition established, what it taught us about tool-use limits, and how the 2026 edition builds on that foundation.

2 Introduction

The EU Council Presidency is a central institutional instrument of the European Union: every six months, a different member state assumes the chair of the Council of the EU, sets political priorities, and mediates between the 27 member states. A country's capacity to use this role effectively varies considerably, depending on administrative capacity, political credibility, geopolitical context, and the quality of preparation. The performance of a presidency is therefore neither a random outcome nor a fully pre-determined one; it is a structured function of country-level capabilities interacting with the EU-level agenda at the moment of the term.

The 2023 edition of this article (the 2023 edition, (Mumin, Claude 2 (Anthropic), and GPT-3.5 (OpenAI) [2023](#))) attempted to apply a systematic KPI model to all historical Council presidencies for the first time and to produce forecasts through 2033. That edition was built with generative AI in its pre-agentic form: Claude 2 and GPT-3.5 used in a manual copy-paste workflow, with no mechanism for retrieving official sources or cross-checking outputs against the public record. The result was a framework that proved directionally accurate on presidency rankings for the 2022–2024 completed term sequence, but that hallucinated the rotation schedule from January 2025 onward, omitted ten of the 27 EU member states from its historical baseline, and contained cross-table inconsistencies that a single consistency check would have caught. The Foreword (Section 1) documents each of these failures by name.

The 2026 edition (the 2026 edition) continues this project with the methodological advantage of *agentic* AI: a system that not only generates text and tables from training data but also researches, retrieves official sources, cross-checks its own outputs against external data, maintains a registry of factual claims across sessions, and runs structured verification sweeps before any version bump. The term *agentic* is used here in a specific sense — not as marketing for a new model generation, but as a description of a workflow in which the AI agent holds project state across documents, verifies claims at the boundary between its own output and the world, and is embedded in a pipeline with explicit stop conditions. The methodology itself is documented by Mumin ([2026](#)).

Research questions

The overarching research question remains the same as in 2023 edition:

Can patterns in the performance of EU Council presidencies be identified, and can these patterns be used to forecast future presidencies?

The 2026 edition adds a second question, made possible by the intervening two and a half years:

How accurate were the 2023 edition forecasts against actual 2022–2025 outcomes — and what does this tell us about the limits and potential of AI-assisted political-science forecasting?

Contribution

This article makes a dual contribution, each oriented to a different reader.

For the political-science audience, it provides a verified KPI dataset covering all 27 EU member states (plus the United Kingdom for pre-2020 historical completeness) across the 1958–2025 window, a validated 2023 edition retrospective against documented 2022–2025 outcomes, and three new analytical frameworks — political-season mapping, an expansion-phase detector for first-presidency learning effects, and a Bayesian treatment of black-swan events — each intended to be readable, applicable, and falsifiable on its own terms. The frameworks are not proposed as replacements for presidency-studies theory (Tallberg 2004; Warntjen 2008; Elgström 2003) but as structured calibration layers on top of it.

For the AI-in-research audience, it provides a worked example of how a known early-stage failure of generative AI (structured hallucination of rotation schedules) is addressed by a later-stage methodology (agentic workflow, claim-fact registry, cross-session world model, pre-registered hypotheses). The full world-model infrastructure used to produce this article is itself an artefact of the methodology: every factual claim is traced to a source, every forecast row is registered as a falsifiable hypothesis, and every volatile claim carries a verification date.

Pre-registration and verification commitment

A distinguishing feature of 2026 edition is that every 2026–2030 forecast is formulated as a named hypothesis with a stated assumption and a stated verification source (Table 22 and Appendix B). A fact-check round is scheduled for 2028–2029, replicating the 2023 edition validation conducted in this document. The forecast accuracy of 2026 edition is thus a testable property of the article, not a rhetorical claim.

The remainder of this article is structured as follows. Section 1 presents the 2023 edition retrospective. Section 3 anchors the KPI framework in presidency-studies literature. Section 4 documents the data and methodology, including the three new analytical frameworks (political seasons in §4.4, the expansion-phase detector in §4.5, and the Bayesian black-swan framework in §4.6). Section 5 sets out the principles of the agentic AI approach used. Section 6 presents the results — historical baselines, validated 2023 edition predictions, and new 2026 edition forecasts with assumptions. Section 7 discusses the findings; §8 translates them into outcomes and implications for specific audiences; §9 concludes. The appendices contain the formal KPI definition, the 2028–2029 verification framework, figure specifications, and a World Model Annex that reproduces the claim-fact registry, canon rules, volatile-facts register, glossary, and source inventory for reproducibility purposes.

Layperson summary

This chapter sets up the two questions the study answers. First, can we look at past EU Council presidencies and spot patterns — which countries tend to deliver, which tend to struggle, and why — and use those patterns to anticipate what upcoming presidencies will do? Second, how accurately did the 2023 version of this study predict the 2022–2025 presidencies? The chapter also explains what *agentic AI* means in this context: not a smarter robot, but a research workflow in which the AI can check its own work against official sources before publication. For a reader whose daily life is shaped by EU laws, the presidency is the institution that sets the six-month agenda — knowing which presidencies are likely to move which files matters for *when* European decisions on energy, migration, borders, or digital policy are actually taken.

3 Theory and Prior Research

Two role conceptions of the presidency

The presidency-studies literature classically distinguishes between two role conceptions: the presidency as neutral mediator — the *honest broker* who manages the agenda impartially — and as active agenda-setter, shaping policy substance in ways that advance the presiding country’s preferences (Smith 2000; Tallberg 2004). Tallberg (2004) demonstrates empirically that presidency countries exploit structural advantages in negotiations (agenda control, drafting power, procedural framing) to advance their own positions. Warntjen (2008) extends this by showing that the formal neutrality norm is routinely subverted by informal power, including the presidency’s asymmetric information about member-state red lines. Elgström (2003) analyses the presidency as a multi-dimensional leadership role combining administrative efficiency, political creativity, and external credibility.

These three works differ in methodology (single-country case studies, negotiated-outcome coding, comparative qualitative analysis) but converge on a common finding: presidency performance cannot be reduced to either neutral procedure or partisan steering. Performance is a function of both, conditioned on the presiding country’s capacity and the political environment of the term. The KPI model in this article adopts this layered view: the five scoring categories (Domestic, Economic, Foreign, Structural, Institutional) capture both procedural throughput (K_5) and agenda-substance progress (K_1 – K_4).

Theoretical lineage of the KPI framework

The KPI system used in this paper operationalises two inherited constructs from the presidency-studies literature: the “effective presidency” construct of Tallberg (2004), which names the structural-advantage dimensions (agenda control, drafting power, procedural framing) that separate high-delivery from low-delivery terms; and the “multi-dimensional leadership role” construct of Elgström (2003), which refuses to collapse performance into a single axis. The five KPI categories (K_1 Domestic, K_2 Economic, K_3 Foreign, K_4 Structural, K_5 Institutional — full definitions in §A) are a direct mapping of those two theoretical instruments into measurable policy domains, with Warntjen (2008) supplying the insight that informal power (asymmetric information about red lines) cannot be reduced to procedural efficiency and therefore requires an K_5 Institutional Affairs category distinct from the four substantive categories. The 2026 edition’s addition of a political-season (§4.4), an expansion-phase (§4.5), and a state-ability synthesis layer (§4.7) extends this lineage by giving the static “effective-presidency” construct a temporal dimension: a country’s role-effectiveness is conditioned on the season in which it holds the chair and on its place in the enlargement-wave learning curve.

Measurement: what the literature quantifies and what it does not

Where the presidency-studies literature becomes thinner is in cross-presidency longitudinal measurement. Schalk et al. (2007) develop an early database of legislative outputs but remain restricted to specific policy fields and to the period 1984–2002. Nowak (2010) and Bengtsson and Elgström (2012) approach the topic through case studies and qualitative comparisons, with rich insights but limited generalisability. König and Hug (2000) contribute a ratification-level game-theoretic account of treaty outcomes that feeds into institutional scoring (the K₅ category of this article’s framework) but do not attempt an explicit presidency-level performance metric.

A consequence of this coverage pattern is that claims about which countries consistently deliver strong presidencies — claims made routinely in policy journalism and think-tank commentary — rest on an uneven empirical base. The Nordic cluster (Sweden, Finland, Denmark) is widely described as high-performing; this article’s historical KPI data supports that claim quantitatively. The pattern for smaller Wave-5 accession countries (Czechia, Slovenia, Cyprus, Malta) is more contested. A structured estimation model, even one that declares itself to be an estimation model rather than a measurement, adds value to the debate precisely by being transparent about its uncertainty.

Post-Lisbon institutional evolution

Vanhoonacker and Res (2010) examine the institutional evolution of the presidency in the post-Lisbon era (from 2009 onward), when the permanent Council President and the High Representative for Foreign Affairs absorbed functions previously held by the rotating presidency. The implication, widely cited, is that the rotating presidency’s relative power has diminished in external affairs but remains decisive in legislative coordination and Council-level agenda-setting. This article’s weighting treats this asymmetry implicitly: the K₃ Foreign Policy category now primarily rewards third-state bilateral work and Council-internal foreign-policy coordination, not pillar-level external representation.

The institutional point matters for forecasting. A presidency’s absolute ability to shape EU foreign policy is capped by the post-Lisbon architecture; its ability to shape internal agenda-setting is not. A model that ignored this would systematically over-score presidencies with strong external diplomatic profiles and under-score those with strong internal procedural capacity. The Dutch and Luxembourg forecasts in this article (H8, H7 in Table 22) explicitly reflect this asymmetry.

AI-assisted forecasting in political science

The second literature on which this article draws is considerably thinner: the methodology of AI-assisted political-science forecasting. A handful of studies have explored large-language-model (LLM) use for political event prediction or policy-document analysis, but systematic retrospective evaluation of AI-generated forecasts — comparing pre-registered predictions against realised outcomes — remains rare. The standard methodology literature on AI in research (Mumin 2026) emphasises that the failure mode most visible at publication is not factual wrongness of a single claim but the absence of a structural mechanism for catching such claims before publication. The 2023 edition of this article is itself an example: the hallucinated rotation schedule from January 2025 onward was a factually wrong claim, but the proximate methodological cause was the absence of any workflow step at which the claim was checked against the public Council rotation list.

Against this background, 2026 edition attempts two things simultaneously. First, it extends the KPI-based presidency analysis of 2023 edition to all 27 EU member states with verified rotation data through 2030. Second, it demonstrates a methodology — world-model-first agentic writing, with claim registry and pre-registered hypotheses — that is applicable well beyond the EU-presidency domain. The expected audience therefore spans two communities: political scientists and EU-policy analysts who care about the presidency data, and AI-in-research practitioners who care about the methodological workflow. Each section of this article is written to serve both.

The contribution this article makes

Against the prior literature, this article offers three things that are not jointly present elsewhere. First, an AI-assisted, KPI-based longitudinal analysis covering all 27 EU member states 1958–2025 with verified rotation, explicit data-type tagging (AI estimate vs. actual retroactive vs. forecast), and source attribution per completed presidency (2022–2025). Second, three new analytical layers — political-season mapping, expansion-phase detection of first-presidency learning curves, and Bayesian black-swan modelling — each introduced in Section 4 with case-study grounding. Third, a pre-registered verification framework (Table 22) that commits the 2026 edition forecasts to a scheduled fact-check round, replicating the 2023 edition validation conducted in this document.

None of these individual elements is unprecedented. Verified rotation schedules exist at source; KPI-style scoring is adjacent to the legislative-output work of Schalk et al. (2007); Bayesian approaches to political shocks exist in quantitative IR. The contribution is the combination, applied consistently across the same dataset, with the methodology itself treated as a testable artefact.

Layperson summary

This chapter anchors the study in what political scientists already know about EU Council presidencies. Two schools disagree: one sees the presidency as a neutral chair, the other as an active agenda-setter. Both are partly right. The chapter also notes that long-term quantitative studies covering every presidency since 1958 are rare — which is why an AI-assisted analysis of this span adds something new. Separately, the chapter acknowledges that AI-assisted political-science research is itself an emerging area with few worked examples, so this study intends to serve the methodology community as well as the EU-policy community. The takeaway for a reader who does not study politics professionally: the claims in this paper sit inside an existing academic debate, and the methodology layer is as much a contribution as the political findings.

4 Data and Methodology

4.1 Scoring System

The KPI system from the 2023 edition is retained unchanged. Five categories are scored on a 1–10 scale; the overall KPI is the equal-weighted average.

Table 2: KPI category definitions

Code	Category	Sub-dimensions
K_1	Domestic Policy	Migration, justice, administration, social affairs
K_2	Economic Policy	Growth, unemployment, competition, budget
K_3	Foreign Policy	Relations with third states, international positions
K_4	Structural Policy	Cohesion policy, agricultural policy, transport
K_5	Institutional Affairs	Efficiency, transparency, legislative output

Formal notation. For a given presidency, each sub-dimension receives a score on the same 1–10 scale: a structured expert judgement made by the AI agent from a reading of primary sources (presidency closing reports, Council conclusions), secondary sources (think-tank assessments, quality journalism), and, for 2022–2025 terms, the documented outcome record in `artifacts/FACT_CHECK_v1_outcomes.md`. A sub-dimension score is the conclusion that reading supports, not a quantitative aggregate of a separately-computed metric. The category-level score KPI_k is then the equally-weighted arithmetic mean of the n sub-dimension scores $Result_i$ within category k :

$$KPI_k = \frac{\sum_{i=1}^n Result_i}{n} \quad (1)$$

Overall KPI (equal-weighted average of 5 categories):

$$KPI_{\text{overall}} = \frac{\sum_{k=1}^5 KPI_k}{5} \quad (2)$$

Historical country baseline (average across all presidencies held):

$$\overline{KPI}_{\text{country}} = \frac{\sum_{t=1}^T KPI_{\text{overall},t}}{T} \quad (3)$$

Weighted acceptance formula. Citizen acceptance scores are derived as weighted sums of category KPIs. The weights below are author-set, not empirically calibrated from

survey data, and reflect the differential policy-area exposure of the two audiences: EU citizens are assumed to weight domestic and structural files higher because those categories touch their daily lives, while non-EU citizens are assumed to weight foreign and economic files higher because those categories shape the EU’s external posture. Any empirical calibration against Eurobarometer-style survey data is a candidate for a future-revision extension (see Section 7 limitations):

$$\text{Acceptance}_{\text{EU}} = \sum_{k=1}^5 w_k^{\text{EU}} \cdot \text{KPI}_k \quad (4)$$

$$\text{Acceptance}_{\text{non-EU}} = \sum_{k=1}^5 w_k^{\text{non-EU}} \cdot \text{KPI}_k \quad (5)$$

Table 3: Category weights for acceptance scores by citizen group

Category	EU citizens	Non-EU citizens
K_1 Domestic Policy	0.30	0.15
K_2 Economic Policy	0.25	0.30
K_3 Foreign Policy	0.15	0.35
K_4 Structural Policy	0.15	0.10
K_5 Institutional	0.15	0.10
Total	1.00	1.00

Emoji mapping (strictly applied in Table 3):

$$\text{Emoji} = \begin{cases} \text{😊 (positive)} & \text{if Acceptance} \geq 7.0 \\ \text{😐 (neutral)} & \text{if } 4.0 \leq \text{Acceptance} \leq 6.9 \\ \text{😞 (negative)} & \text{if Acceptance} < 4.0 \end{cases} \quad (6)$$

4.2 Data Sources and Transparency

Important methodological note: All KPI scores — historical and prospective — are AI-generated estimates. They are based on the knowledge of the AI model used, supplemented by targeted web research into official assessments, think tank analyses, and government reports. They do not substitute independent empirical primary research.

The master dataset `data/presidency_scores_1958-2025.csv` contains 146 observations across 16 variables, distinguishing three data types. The `data_type` column is the single operational boundary that tells a reproducing reader which rows rest on documented sources and which are AI estimates alone:

- **ai_estimate**: historical KPI scores derived from model knowledge alone (Claude 2 in the 2023 edition; agentic-era models in the 2026 edition). No per-category source attribution; the score reflects the AI agent’s structured reading of the country’s presidency record available at training time.
- **actual_retroactive**: scores for the 2022–2025 completed presidencies. Each category score is anchored in the outcome-report review documented in `artifacts/FACT_CHECK.v1_outcomes.m`; the bibliography entries per row (see the ‘source’ column of Table 16) are the primary evidence. A reproducing reader should be able to open the cited source and see the rationale for the category score.
- **forecast**: predictions for 2026–2030 with explicit assumptions stated in Table 18 and falsifiable thresholds named in Table 22. No per-category source attribution is possible, since the term has not yet occurred.

Epistemic status of claims — quick reference

To prevent readers from confusing the epistemic status of different claim types, Table 4 summarises the four layers of claims the paper contains and where each appears.

For retroactive actual assessments (2022–2025), the following sources were consulted: official presidency closing reports; European Parliament assessments; think tank analyses (ECFR, Friedrich Naumann Foundation, Real Instituto Elcano, EEB); and quality journalism (Euronews, Politico Europe).

For forecasts 2026–2030: the official EU Council rotation list (consilium.europa.eu, verified April 2026); current political profiles; and the EU Strategic Agenda 2024–2029.

4.3 Rotation Plan (officially verified)

All tables in the 2026 edition are based exclusively on the following verified schedule:

Layperson summary

This chapter explains how the numbers in this study are produced. Each presidency receives a score from 1 to 10 in five areas (domestic, economic, foreign, structural, institutional) and the overall score is the average. These numbers are not measurements from a poll or a performance audit: they are structured judgements made by the AI system from a reading of official closing reports, independent think-tank assessments, and quality news coverage. The chapter states this explicitly through a **data_type** label on every row of the underlying dataset — either an AI estimate, a retrospective assessment backed by documented sources, or a forecast with a stated assumption. The calendar of which country holds which presidency is verified against the official EU Council schedule, not guessed.

Table 4: Epistemic status of claims in this paper

Status	Where the claims appear	What supports them
Retrospective validated	§1 Foreword 2023-edition retrospective; §6 Results Actual-KPI column (2022–2025); §4.8 Framework Backtest (24 historical cases)	Published outcome reports, think-tank assessments, <code>artifacts/FACT-CHECK_v1-outcomes.md</code> , primary sources in bibliography
Structured AI estimate	Table 15 historical country averages (pre-2022 rows); §4.7 state-ability matrix entries; per-country <code>eu_acceptance</code> / <code>noneu_acceptance</code> scores	AI agent’s structured reading of public-record country profile; tagged <code>ai_estimate</code> in the CSV (no per-category primary source)
Prospective forecast	§6 Results 2026-edition Forecast KPI column (2026–2030 rows); §A.3 Verification Framework Table 22 (H1–H11)	Agent applies the four-layer synthesis (§4.4–§4.7) with explicit assumption statements; tagged <code>forecast</code> in the CSV; falsifiable at the 2028–2029 fact-check round
Methodology-portability	§5 Agentic AI; §20 Persona & Verification Approach; <code>docs/21-abstracted-methods.md</code> Mumin Framework Synthesis Method (MFSM)	Six validity tests on the method (round-trip, composition, contradiction, method sequence etc.); the method is portable to comparable political-institution studies

Table 5: Official EU Council Presidency rotation 2022–2030

Period	Country	Status
Jul–Dec 2022	Czechia	Completed
Jan–Jun 2023	Sweden	Completed
Jul–Dec 2023	Spain	Completed
Jan–Jun 2024	Belgium	Completed
Jul–Dec 2024	Hungary	Completed
Jan–Jun 2025	Poland	Completed
Jul–Dec 2025	Denmark	Completed
Jan–Jun 2026	Cyprus	Current
Jul–Dec 2026	Ireland	Planned
Jan–Jun 2027	Lithuania	Planned
Jul–Dec 2027	Greece	Planned
Jan–Jun 2028	Italy	Planned
Jul–Dec 2028	Latvia	Planned
Jan–Jun 2029	Luxembourg	Planned
Jul–Dec 2029	Netherlands	Planned
Jan–Jun 2030	Slovakia	Planned
Jul–Dec 2030	Malta	Planned

Source: consilium.europa.eu, April 2026. Rotation from 2031 not yet officially published.

4.4 Political Season Mapping

A key analytical contribution of 2026 edition is the introduction of a **political season framework** — a typology of EU integration cycles analogous to climatic seasons. This addresses a systematic blind spot in 2023 edition: predictions were made at the country level without contextualising the broader EU political environment in which a presidency operates. The season framework operates as a structured calibration layer on top of the country-level prior before any Bayesian black-swan update (Section 4.6) is applied.

Figure 1 provides a taxonomy-at-a-glance view of the six season tags used throughout the paper, grouped into constructive / agenda-setting seasons (top row) and responsive / crisis-shaped seasons (bottom row), with one canonical institutional anchor per tag.

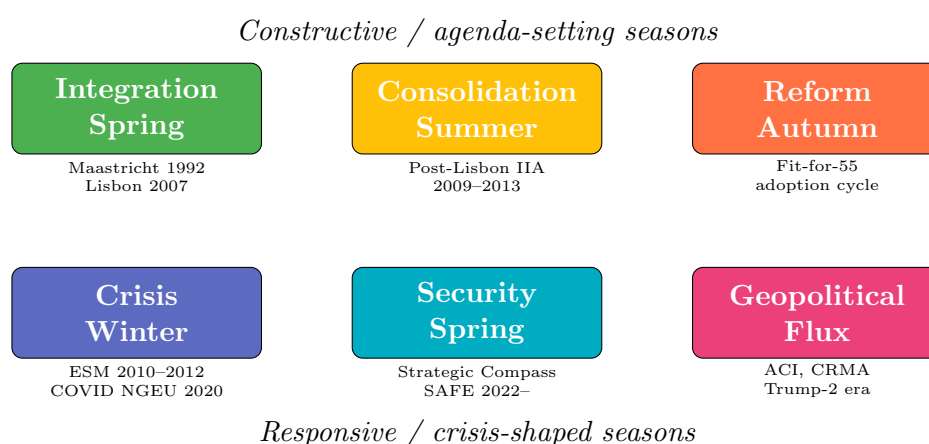


Figure 1: Season taxonomy at a glance. The six season tags used throughout the paper (§4.4) decompose into constructive / agenda-setting seasons (top row: Integration Spring for new-competence framing; Consolidation Summer for operationalising prior agreements; Reform Autumn for cross-file reform bundles) and responsive / crisis-shaped seasons (bottom row: Crisis Winter for fast-tracked emergency response; Security Spring for accelerated defence / border files; Geopolitical Flux for external-posture reactive framing). One canonical institutional anchor is listed below each tag. This figure is the taxonomy counterpart to Figure 2, which presents the same season tags as a time-ordered historical sequence.

Figure 2 places the same six season tags on the historical timeline 1958–2030, showing how the EU has cycled through them and where the current dual-season overlap (Security Spring + Geopolitical Flux) sits relative to prior cycles.

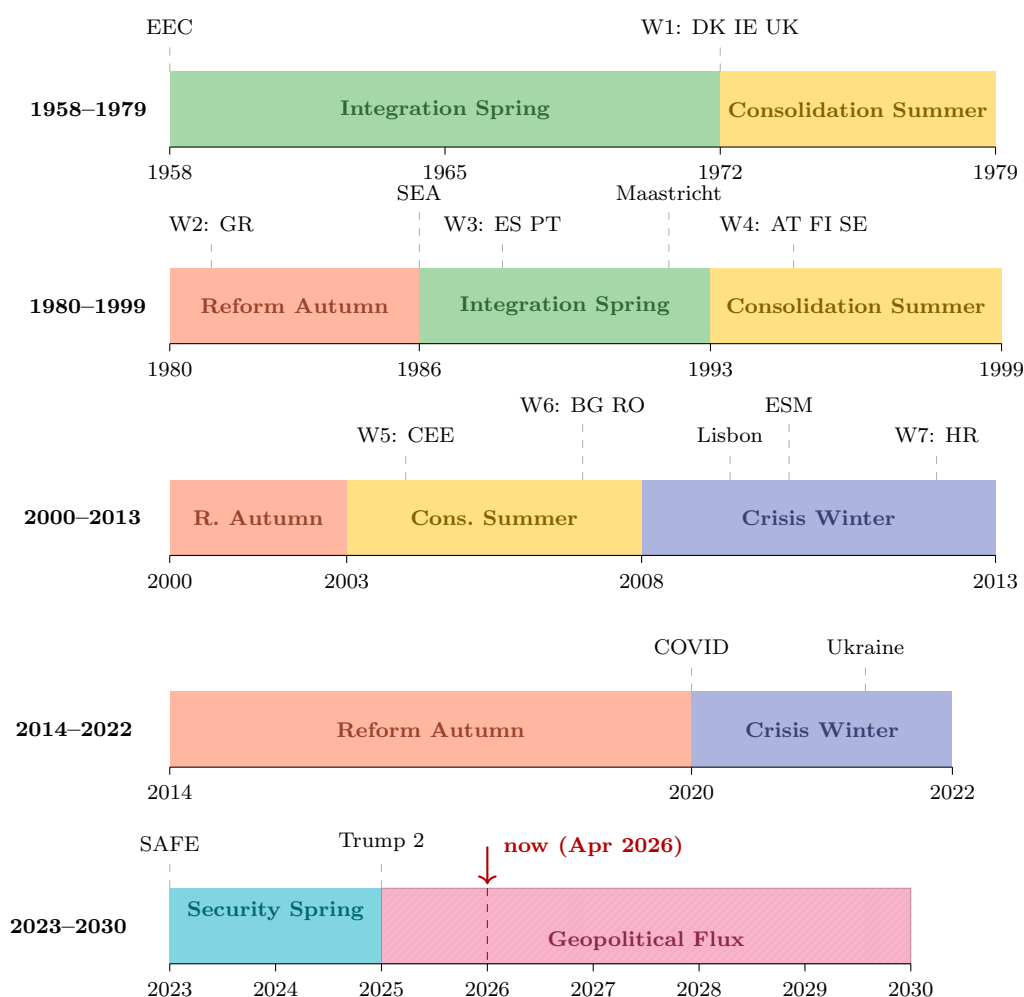


Figure 2: EU political-season timeline 1958–2030, split into five era rows so every season band renders at a readable width. From top to bottom: 1958–1979 (founding era, Integration Spring → Consolidation Summer); 1980–1999 (Reform Autumn → Integration Spring → Consolidation Summer, Single-Market and Maastricht); 2000–2013 (Reform Autumn → Consolidation Summer → Crisis Winter, Eastern enlargement and Global Financial Crisis); 2014–2022 (Reform Autumn → Crisis Winter, Juncker / von der Leyen I cycle into COVID and Ukraine); 2023–2030 (current + forecast, Security Spring overlaid with Geopolitical Flux from 2025). The red marker on the bottom row indicates the current position (April 2026). Coloured bands use the six-season taxonomy of §4.4 (see also `eu_season` column in `data/presidency_scores_1958-2025.csv`). Event annotations above each bar mark: founding EEC (1958); enlargement waves W1–W7 by joining year; Single European Act (1986); Maastricht Treaty (1992); Lisbon Treaty (2009); European Stability Mechanism (2010–2012); COVID-19 onset (2020); Russian invasion of Ukraine (2022); SAFE defence instrument (2023); second Trump administration (2025).

Six Season Types

Table 6: EU political season typology with historical examples

Season	Characteristics	Historical examples
Integration Spring	High ambition, treaty changes, enlargement waves	1958–1972 (EEC founding); 1986–1992 (Single Act, Maastricht)
Consolidation Summer	Implementation focus, lower legislative ambition	1973–1979; 1993–1996; 2003–2007
Reform Autumn	Institutional overhaul, crisis-response preparation	1980–1985; 1997–2002; 2014–2019
Crisis Winter	Existential pressures, emergency legislation	2008–2013 (financial); 2020–2022 (COVID + Ukraine)
Security Spring	Post-crisis reorientation: defence, energy, borders	2023–ongoing (SAFE instrument, Fit for 55, migration pact)
Geopolitical Flux	Sustained external disruption forcing agenda pivot	2025–ongoing (Trump 2.0 trade shocks)

The six labels are drawn from the chronology of EU integration episodes documented in the institutional literature (Vanhoonacker and Res 2010; Elgström 2003). The first four (spring / summer / autumn / winter) represent recurrent cycle phases; the last two (Security Spring and Geopolitical Flux) are specific to the post-2022 era and are therefore new typological additions in 2026 edition. They are introduced here because the country-level KPI framework inherited from 2023 edition does not by itself distinguish between a presidency held in a crisis window and one held in a cooperative window; without this distinction, the model systematically under-explains the variance in recent terms.

Seasonal KPI Adjustment

Each season defines a baseline KPI adjustment applied to the country-level prior. The adjustment values listed below (+0.5, +0.3, +0.1, 0, −0.5) are *design parameters* calibrated against historical KPI means within the respective season classes, not fitted coefficients from a regression. The same-prior caveat from §4.5 applies (see the “How to compare a specific case against the prior” note).

- *Crisis Winter*: −0.5 (external pressure suppresses legislative achievement; agenda dominated by emergency response).
- *Integration Spring*: +0.5 (cooperative environment amplifies agenda achievement; treaty ambition increases throughput).

- *Security Spring*: +0.3 for Eastern-accession (Wave-5/6/7), Baltic, and Nordic presidencies; 0 for Southern (Wave-2/3) and Wave-4 Alpine presidencies whose domestic fiscal agenda is partially disconnected from the security pivot; +0.1 for Core founding presidencies where the asymmetry is in-between.
- *Consolidation Summer, Reform Autumn*: 0 (neutral environment; KPI driven primarily by country-level factors).
- *Geopolitical Flux*: 0 baseline but with an enlarged ± 0.5 uncertainty band on the prior, reflecting that the season's defining feature is unpredictability rather than directional pressure. This ± 0.5 interval is a *structural prior* tied to the season tag under the second Trump administration's trade and defence dynamics, not a statistical confidence interval; it widens the band within which a forecast counts as confirmed, and does not represent a sampling distribution.

Operational application. The CSV column `eu_season` records the season tag for every row; the valid values match the six labels in the typology table above and are enumerated in `docs/04_glossary.md` of the project world model. A reproducing reader should read the tag from the CSV, look up the corresponding adjustment in the bullet list above (applying the regional asymmetry for Security Spring via the country's wave from Table 15), and add it to the country-level `kpi_overall` prior. The adjustment is not pre-computed into the CSV; it is applied downstream of the prior so that a reader can change season assignments or adjustment magnitudes and re-derive the adjusted expectation.

Classifying a new term

A reader applying this framework to a presidency not covered here should follow three steps. First, identify the dominant EU-level agenda of the six-month window (treaty work / implementation / reform / crisis response / security reorientation / external disruption). Second, match it against the six-season table. Third, apply the baseline adjustment to the country-level KPI prior and then proceed to any applicable Bayesian update from Section 4.6. The framework is deliberately coarse: a season is a six-month aggregate, not a policy- file-level tag.

Current Position

The EU is currently in a dual-season state: **Security Spring** (2023–ongoing, defined by the Fit for 55 closure, migration pact, and SAFE defence instrument) overlapping with early-stage **Geopolitical Flux** (from January 2025, driven by the second Trump administration’s trade and NATO policies; see volatile-fact V-022). For forecast rows from January 2026 onward, the `geopolitical_flux` tag signals an elevated ± 0.5 uncertainty band on all KPI predictions (Table 16); the Security Spring +0.3 adjustment remains applied where the season overlap is active and the presidency is within its directional-beneficiary regional group.

The dual-season classification is itself a methodological claim about the current window: these are not two phases in sequence but two overlapping regimes, each with its own calibration impact. Distinguishing them is what allows Ireland’s 2026-H2 forecast (a Western-European pragmatic-agenda presidency) to diverge in expected KPI from Lithuania’s 2027-H1 forecast (an Eastern-European security-agenda presidency), even under identical country-level priors.

Where each season shows up in EU institutional documents

Each season tag corresponds to a recognisable class of EU institutional output that a practitioner can consult for verification. The mapping below names, for each tag, an example document or document-class where that season’s emphasis is visible. The mapping is indicative, not authoritative (the framework is principally anchored on KPI-record patterns, not on document taxonomy), but a future fact-check author re-tagging a historical season can use these document classes as primary evidence.

- **Integration Spring:** major treaty signings (Single European Act 1986, Maastricht 1992, Lisbon 2007) and their preparatory Intergovernmental Conferences. Council conclusions in this season emphasise new-competence framing.
- **Consolidation Summer:** Inter-Institutional Agreements on budget / procedure; Council decisions implementing prior treaty changes. Emphasis on operationalising what was agreed upstream.
- **Reform Autumn:** Commission Work Programmes and Council General-Approach sessions on structural-reform packages (governance reforms, economic-semester introductions, Fit-for-55 adoption cycle). Emphasis on cross-file reform bundles.
- **Crisis Winter:** ESM foundation documents 2010–2012; COVID-NGEU regulation 2020; emergency-Council conclusions invoking Article 122 TFEU. Emphasis on fast-tracked response and suspended normal-procedure discipline.
- **Security Spring:** Strategic Compass 2022; SAFE defence instrument regulation; European Peace Facility decisions; Schengen entry / exit emergency measures. Emphasis on defence / border / security files receiving accelerated Council treatment.

- **Geopolitical Flux:** trade-defensive instruments (Anti-Coercion Instrument, Foreign Subsidies Regulation implementations), critical-raw-materials and digital-sovereignty communications, transatlantic-contingency Council conclusions. Emphasis on external-posture reactive framing.

Layperson summary

The *season* idea is a simple way to describe what was on the EU's mind during a given presidency. When the EU is building new treaties and enlarging, that is an *integration spring*; when it is fighting fires (financial crisis, pandemic, invasion), that is a *crisis winter*. Two labels are specific to our current era: *security spring* describes the post-2022 shift toward defence, energy, and border files, and *geopolitical flux* captures the uncertainty the second Trump administration has added to trade and alliance policy. Knowing the season of a presidency helps explain why one country's term delivered more than another's — the political weather matters as much as the country's capacity.

4.5 Expansion Phase Detector

EU enlargement phases materially reshape the presidency dynamic. New member states bring unfamiliar administrative cultures, distinct thematic priorities, and, historically, lower KPI scores on their first rotation term than their regional archetype mean would predict. The expansion phase detector formalises this as a structured calibration variable applied on top of the country-level and seasonal priors.

Seven Documented Enlargement Waves

Table 7: EU enlargement waves and their presidency culture contribution

Wave	Year	Countries added	Cultural contribution
0 (founding)	1958	BE, DE, FR, IT, LU, NL	Core continental; treaty-building
1	1973	DK, IE, UK	Nordic + Anglo-Saxon culture
2	1981	Greece	Southern flank; Mediterranean
3	1986	ES, PT	Second Southern; Structural Funds
4	1995	AT, FI, SE	Nordic cluster; consistently high KPIs
5	2004	CZ, EE, HU, LV, LT, MT, PL, SK, SI, CY	Largest expansion; CEE culture
6	2007	BG, RO	Rule-of-law challenges
7	2013	Croatia	COVID presidency 2020

First-Presidency Learning Curve

Countries holding a first Council presidency of any era are expected to score approximately **−0.7 KPI points** below their regional archetype mean as a *structural prior*. This handover-discount pattern is observed symmetrically across founding-state first terms (1958–1972), 1995-enlargement first terms (AT, FI, SE), and 2004 / 2007-enlargement first terms (Wave-5 and Wave-6), and is expected to apply equally to any future Wave-8 accession cohort (Ukraine, Moldova, Western Balkans). The penalty has two sources: administrative capacity (a presidency requires several hundred coordinated officials, and the pre-term build-up cannot substitute entirely for institutional memory), and agenda-setting legitimacy (veteran presidencies draw on accumulated diplomatic capital that first-time holders must earn in-term). Figure 3 visualises the cross-sectional country average per wave, making above-prior and below-prior cases immediately inspectable; Figure 4 then shows the three-level adjustment curve applied to the prior.

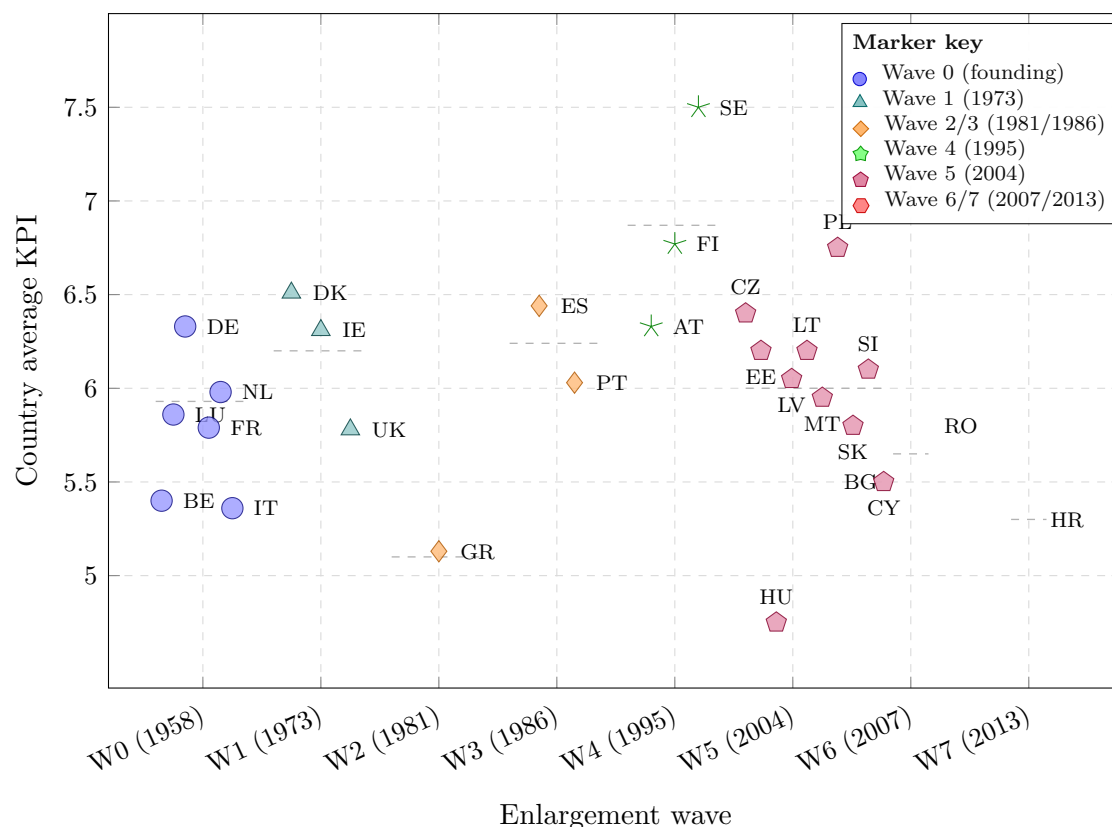


Figure 3: Regional cluster analysis: country average KPI grouped by enlargement wave. Each marker is one country, at a horizontally jittered position within its wave (x-axis) and at its country-level mean KPI on the y-axis; marker shape encodes wave cluster (see marker key). ISO country codes sit beside or above / below each marker to avoid glyph overlap (iter 49c legibility fix). Dashed grey horizontal segments in each wave column show the wave's archetype mean (used in §4.5). The Nordic-Alpine cluster (AT, FI, SE in Wave 4) sits above all other cohorts; Wave 5 shows the widest within-wave spread (PL 6.75 to HU 4.75); Wave 6 and Wave 7 are small-sample cohorts to be read with caution. The gap between an individual country's marker and its wave's archetype-mean dashed line illustrates the above-prior / below-prior decomposition of §4.5. Bubble sizes are not proportional to n-presidencies in this rendering to preserve legibility with many Wave-5 countries; counts are reported in Table 15.

How to compare a specific case against the prior. The regional archetype mean is the average of `kpi_overall` across all completed presidencies of the wave. From Table 15, the Wave-5 (CEE) archetype mean is approximately 6.0 (the arithmetic mean of the ten Wave-5 country averages: $(6.8 + 6.4 + 4.8 + 6.2 + 6.2 + 6.0 + 5.8 + 6.1 + 6.0 + 5.5) / 10 \approx 5.98$). A first-term KPI below 5.3 is consistent with the full structural penalty; a first-term KPI above the archetype mean indicates that the specific term outperformed the prior, typically under a favourable seasonal adjustment (§4.4) or a positive event-type adjustment (§4.6).

The pattern across waves shows both on-prior and off-prior cases:

- *Wave 5 (2004) below prior*: Hungary's first presidency (Jan–Jun 2011, KPI 5.0, from the master dataset) sits 1.0 KPI point below the Wave-5 archetype mean of 6.0, exceeding the structural -0.7 penalty. The additional shortfall is consistent with domestic rule-of-law concerns already visible at the 2011 term (later formalised under Article 7 scrutiny and documented retrospectively in §4.6).
- *Wave 5 (2004) at prior*: Czechia's first presidency (Jan–Jun 2009, KPI 6.0) sits essentially at the archetype mean. The structural penalty was fully offset by a neutral-season environment and strong technical delivery on pre-Lisbon files, consistent with the prior-and-offset reading rule above.
- *Wave 5 (2004) above prior*: Poland's first presidency (Jul–Dec 2011, KPI 6.5) sits 0.5 above the archetype mean. The structural penalty was more than offset by strong administrative capacity and a relatively cooperative Consolidation Summer season environment.
- *Wave 6 (2007)*: Bulgaria's first presidency (Jan–Jun 2018, KPI 5.5) sits 0.5 below the larger Wave-5/6/7 CEE-wide mean, close to the expected penalty magnitude. Romania's 2019 presidency (KPI 5.8) sits near the CEE mean; rule-of-law controversy noted at the time did not translate into a lower retrospective assessment, so the Wave-6 cohort is a weak anchor for the penalty magnitude.
- *Wave 7 (2013)*: Croatia (Jan–Jun 2020, KPI 5.3) held its sole presidency under COVID. Both the crisis context and the first-presidency penalty contributed, so the point estimate is a joint observation under both the seasonal and the expansion-phase adjustments.

The -0.7 figure should therefore be read as a prior-expectation delta that frequently falls on one side or the other of the observed value, not as an empirical mean with the cases as outliers. Reading the specific delta for any presidency requires decomposing it across the seasonal and event-type adjustments from §4.4 and §4.6. The penalty has been observed to apply symmetrically across enlargement waves: the EU's founding-state member presidencies in their earliest post-rotation terms (1958–1972) showed comparable first-time-handover discounts within the limits of the historical data, and Wave-5 (CEE) member states' first terms in the 2010s show the same pattern. The -0.7 is a structural prior on the act of holding a first presidency, not a judgement about any specific accession cohort.

Second-Presidency Correction

Second-term presidencies typically recover $+0.4$ **KPI points** relative to the country's first term, placing them approximately at the regional archetype mean. The correction is empirically supported by two within-country deltas in the current dataset:

- *Czechia*: KPI 6.0 (2009, first) vs. 6.8 (2022, second), a delta of $+0.8$ — double the expected $+0.4$, with the excess attributable to a positive Bayesian-vocabulary adjustment

under the External Security Shock event type (Ukraine war; §4.6) and a Security Spring season tag.

- *Poland*: KPI 6.5 (2011, first) vs. 7.0 (2025, second), a delta of +0.5. Closer to the nominal +0.4 correction alone, with a small +0.1 residual consistent with the Security Spring season tag and the Ukraine-security context at the 2025 term.

Two within-country cases do not constitute statistical validation of the +0.4 correction. They do anchor the expected direction and order of magnitude. The figure should be read as a structured prior with relatively wide tolerance, not as a calibrated point estimate. A formal validation would require a regression across all available first/second presidency pairs in the dataset — a candidate for a future-revision methodological extension.

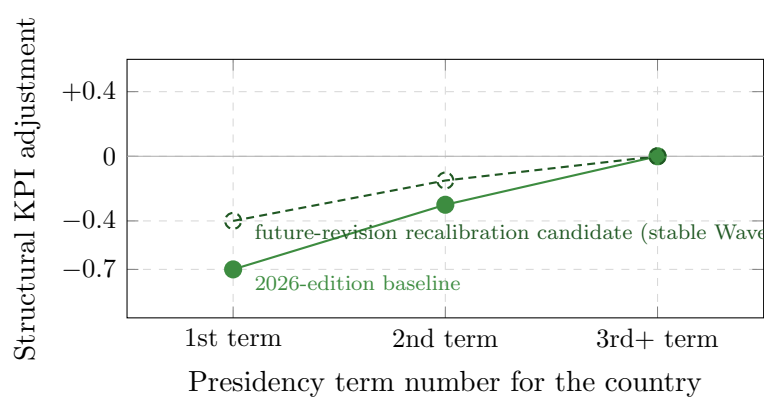


Figure 4: Expansion-phase dose-response curve (§4.5). Solid line: 2026-edition structural-prior adjustment applied to a country’s KPI forecast based on the sequence number of its presidency (*1st term* = -0.7 ; *2nd term* recovers $+0.4$ relative to first, or -0.3 absolute; *3rd or later* returns to zero). Dashed line: future-revision recalibration candidate from the 24-case backtest (§4.8 rules-of-thumb #1), reducing the first-term penalty to -0.4 for politically-stable Wave-5 cases. The 2028–2029 fact-check round will arbitrate between the two calibrations; H13 (Table 20) is the pre-registered test.

Operational use: applying the framework in pre-term preparation

A reader in a Working Party, permanent representation, or national EU-affairs ministry can apply the expansion-phase detector to a given upcoming presidency in three steps. First, identify the presiding country’s accession wave from the table above and locate its regional archetype mean (§4.5 and Table 15). Second, check whether this is the country’s first, second, or later Council presidency: apply the -0.7 learning-curve penalty for a first term, $+0.4$ correction for a second, and no expansion-phase adjustment for a third or later term. Third, pass the adjusted expectation into the seasonal calibration (§4.4) and then the black-swan conceptual framework (§4.6) for any known event-type exposure. The output is an expected KPI band, not a point forecast; planners should treat it as input to dossier prioritisation, not as a scoring target.

For first-presidency cohorts specifically, two practical implications follow. Administrative handover from the predecessor presidency matters more than for veterans: the incoming team has less institutional memory to absorb the files in progress, and a structured handover protocol (at the Working Party chair level, not only at the ministerial level) is the single highest-leverage preparation activity. Dossier prioritisation should favour files where the first presidency has the strongest national competence, deferring technically complex legacy files to the subsequent trio term where possible.

Prospective Wave 8

Ukraine, Moldova, Albania, and Montenegro are in active accession negotiations with the EU as of April 2026. A potential enlargement in the 2028–2032 window would introduce Western Balkan and Eastern European presidency cultures and would extend the learning-curve pattern into a further rotation cycle. Under the expansion-phase detector, the first presidencies of a Wave-8 cohort would carry the -0.7 penalty relative to their regional archetype mean; the Eastern-archetype anchor (Wave 5 CEE) would predict a first-term KPI band of approximately 5.0–5.5 for an initial Ukrainian or Moldovan presidency, conditioned on no confounding event at the time of term. No forecasts for Wave-8 countries are included in Table 16; the prospective-wave framing is an outlook only. See volatile-fact V-021 in the project’s internal register for verification-date tracking of accession-timeline claims.

Layperson summary

First-time presidencies tend to deliver less than veteran ones. This chapter puts a number on it: a first term typically scores about 0.7 points below the regional average, and a second term usually closes most of that gap. The chapter works through real cases — Hungary, Czechia, Poland, Bulgaria, Romania, Croatia — to show how the pattern plays out. For readers in countries on the accession track (Ukraine, Moldova, Western Balkans), this chapter sets a realistic expectation: a country’s first Council presidency is unlikely to deliver as much as a veteran’s, and this is a structural pattern driven by administrative capacity, not a judgement of individual ministers or civil servants.

4.6 Bayesian Black-Swan Framework

This framework uses Bayesian vocabulary as a structured-prior discipline, not a computational machine; posterior masses are not assigned, and the ± 0.5 uncertainty band attached to forecasts is a structural prior rather than a statistical confidence interval. Computational Bayesian calibration using Eurobarometer and legislative-throughput datasets (schema stubs in `data/`) is a future-revision extension target.

A structural limitation of 2023 edition was its inability to account for external shocks. The Ukraine war (February 2022), the COVID-19 pandemic (2020), and the second Trump administration from January 2025 each represent events that override base-theme predictions in ways not derivable from historical country profiles. The 2026 edition treats these disruptions through a **structured-qualitative framework with Bayesian vocabulary**: shocks are classified into a small number of event types, each with a documented direction and range of KPI impact, rather than being treated as model failures.

Scope note (read before the equations below). The framework is Bayesian in structure but not in computation. The prior and posterior notation in Equations 7–8 is conceptual: it names the quantities that a rigorous quantitative Bayesian application would estimate, but this article does not estimate σ_{region}^2 from data, does not specify a likelihood function $P(E \mid \text{KPI})$, and does not compute the posterior as a distribution. The "update" reported in the case studies below is a point shift of the prior mean by an event-type delta, not a distribution update. A fully quantitative Bayesian application would require an empirical distribution of observed shock effects across multiple political systems and at least dozens of comparable cases per event type, neither of which the current dataset contains. That extension is a candidate for a future-revision methodological successor and is acknowledged as a limitation in Section 7.

The framing is deliberate. A model that treats any deviation from prediction as a failure mode is brittle by construction; a structured classification that names event types, ranges their expected KPI impact, and adjusts the prior accordingly is auditable. The framework's contribution is therefore not a quantitative Bayesian claim. It is a claim that the direction and order of magnitude of KPI deviations are anticipatable conditional on the event type being correctly named, and that this anticipation can be recorded in advance rather than rationalised after the fact.

Conceptual setup (analogy notation)

Prior (conceptual): the KPI expectation based on historical country profile (Section 4.5) and seasonal context (Section 4.4). In the full Bayesian analogy this would be a normal distribution around the country baseline with region-scale variance:

$$P(\text{KPI}_{\text{pred}}) \sim \mathcal{N}(\overline{\text{KPI}}_{\text{country}}, \sigma_{\text{region}}^2) \quad (7)$$

In the present application σ_{region}^2 is not estimated; the prior is used as a point expectation at $\overline{\text{KPI}}_{\text{country}}$.

Evidence: a black-swan event E occurs during the presidency ($P(E)$ low by definition).

Posterior (conceptual): the updated KPI expectation given E . The full Bayesian analogy would write:

$$P(\text{KPI} | E) \propto P(E | \text{KPI}) \cdot P(\text{KPI}) \quad (8)$$

In the present application the right-hand side is not computed. Instead the prior mean is shifted by an event-type delta Δ_E drawn from the ranges in Table 8, and the result is reported as the adjusted expectation. This is a named-delta point shift under a Bayesian-structured vocabulary, not a posterior distribution.

Three Black-Swan Types with Documented KPI Impacts

Table 8: Black-swan event types and their KPI impact

Type	Mechanism	KPI Impact	Examples
External Security Shock	War/crisis forces emergency legislation; border-state presidencies receive agenda-rally premium	Border: +0.5 to +1.0 Others: −0.3 to −0.5	CZ 2022 (+0.6); Ukraine war
Domestic Rule-of-Law Collapse	Art. 7 scrutiny; institutional boycott by Commission or partner capitals	−0.8 to −1.5	HU 2024 (−0.6)
Global Economic Shock	Recession or trade war forces macroeconomic response agenda	−0.3 to −0.8	CY 2012 (eurozone crisis)

Case Studies

Czechia 2022 (External Security Shock, border premium): 2023 edition predicted KPI 6.2; the retroactive actual assessment is 6.8, a delta of +0.6. Czechia’s geographic proximity to Ukraine and crisis-management capacity produced a positive Bayesian posterior update: the presidency delivered emergency energy regulation, Schengen expansion

preparation, and sustained Ukraine support — outperforming its historical baseline (KPI ≈ 6.0) by close to one standard deviation of the Eastern-European archetype.

Hungary 2024 (Domestic Rule-of-Law Collapse): 2023 edition predicted KPI 5.1; actual assessment 4.5, a delta of -0.6 . PM Orbán’s unannounced visits to Kyiv, Moscow, and Beijing in the first week of the term provoked a European Commission boycott and unprecedented diplomatic isolation (Euronews [2024](#)). The moderate magnitude of the actual delta — within the -0.8 to -1.5 range but at the upper (less severe) end — reflects that some legislative business continued despite the boycott, notably on competitiveness files following the Budapest Declaration.

Figure [5](#) summarises all seven 2023-edition-predicted vs. actual deltas for the 2022–2025 retrospective window, including the Czechia $+0.6$ and Hungary -0.6 cases above alongside the near-zero cases (Spain, Sweden, Belgium).

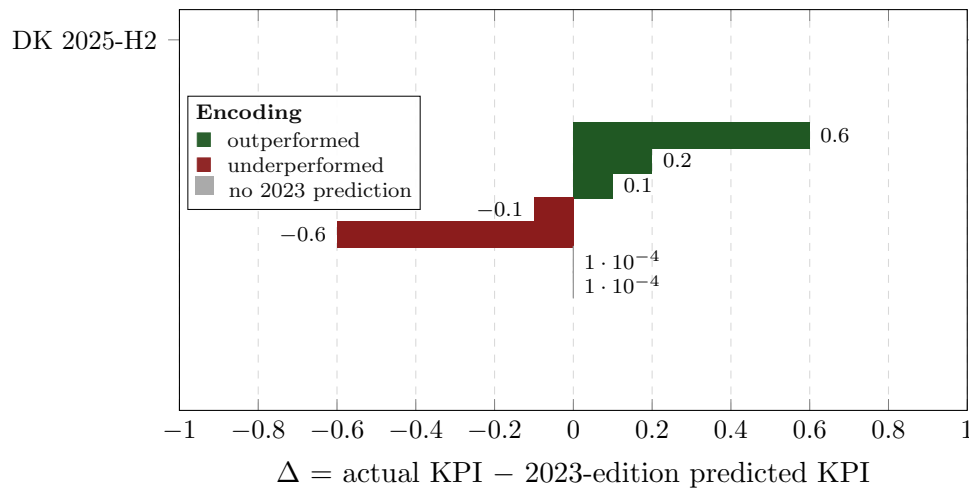


Figure 5: Black-swan delta for completed 2022–2025 presidencies. Each bar shows the difference between the retroactive actual KPI assessment and the 2023-edition point prediction (both on the 1–10 scale). Green bars indicate the presidency *exceeded* its 2023-edition prediction; red bars indicate *shortfall*; *hatched grey* bars (PL 2025-H1, DK 2025-H2): no 2023-edition prediction was produced because of the rotation error documented in §1, so their delta is structurally undefined and shown at near-zero by convention. The hatched encoding distinguishes them visually from solid-coloured near-zero predictions to prevent misreading as “near-perfect prediction”. The Czechia deviation (+0.6) illustrates an External Security Shock event-type adjustment (Ukraine-war border premium, §4.6). The Hungary deviation (−0.6) is a Domestic Rule-of-Law Collapse case. Spain, Sweden, and Belgium sit within ± 0.2 of the 2023-edition prediction, consistent with the directional-accuracy strength and central-tendency compression discussed in §7.1. Reading this figure against Figure 2 enables a cross-figure meta-tier inference: the two largest deltas (CZ, HU) both fall in the transition from Crisis Winter to Security Spring, while mid-range presidencies in the same transition window (SE, ES, BE) show near-zero deltas. Bars show point estimates only; no measurement interval is attached because the retrospective-KPI assessments are structured estimates on a 1–10 scale rather than sampled measurements (the structural ± 0.5 prior of §4.4 applies to the *predicted* side, not to the retrospective-actual side, and so does not render here as an error bar).

Trump 2.0 as a Persistent Background Condition (2025–)

Unlike a single-event shock, the second Trump administration (from January 2025; see volatile-fact V-022) represents a persistent background condition affecting multiple consecutive presidencies rather than a discrete event. In the framework’s vocabulary this is not a posterior update (which would collapse prior uncertainty) but an enlargement of the prior’s uncertainty interval on 2026–2028 rows: the prior mean is not shifted, and the reported expectation stays at $\overline{\text{KPI}}_{\text{country}}$, but the range of plausible outcomes widens. Trade shocks elevate economic policy on every presidency’s agenda; NATO burden-sharing pressure strengthens the defence agenda of Eastern and Nordic presidencies; multilateralism uncertainty increases the institutional value of EU cohesion.

Adjusted expectation for 2026–2028 forecasts: all KPI predictions in Table 16 carry a ± 0.5 interval attributable to ongoing geopolitical flux. The interval is reported without a distribution family; it is a stated range, not a calibrated confidence interval. Lithuania (Jan–Jun 2027) and Latvia (Jul–Dec 2028) are directionally most exposed to upward revisions if the security agenda remains dominant at the time of their terms; Ireland (Jul–Dec 2026) and the Netherlands (Jul–Dec 2029) are directionally most exposed to downward revisions if trade-war pressure disrupts their economic-policy agenda delivery. No claim is made about the probability mass attached to either direction.

Key interpretive principle: black-swan events are not model failures. A framework that names event types, ranges their expected KPI impact, and records the adjustment explicitly provides more information to its reader than one that delivers a single point estimate with no structure for revising it when the world delivers new information. This principle is what distinguishes 2026 edition’s treatment of the Czechia and Hungary cases from 2023 edition’s silent over- and under-prediction: the deltas are not errors to be hidden but structured adjustments to be reported. The word “Bayesian” in this framework’s name refers to the conceptual architecture of prior-and-adjustment, not to a Bayesian computation.

Operational use: applying the framework in pre-term preparation

A reader in a Working Party, permanent representation, or national EU-affairs ministry can apply this framework to a prospective presidency in three steps. First, identify which of the three event types the incoming term could plausibly be exposed to, given the presiding country’s geography (External Security Shock for border-states adjacent to a conflict zone), its domestic political situation (Domestic Rule-of-Law Collapse where Art. 7 scrutiny is active), and the prevailing macroeconomic setting (Global Economic Shock during recession or trade-war windows). Second, locate any adjustments already recorded for the target presidency in Tables 16 and 22: the ± 0.5 interval on 2026–2028 rows already incorporates the persistent-background condition of the Trump 2.0 window, so planners

should not double-count it. Third, prepare a short contingency file per plausibly-exposed event type, naming (a) the files the presidency would accelerate under that event, (b) the ministerial coordination pattern the event would trigger (Commission, national capitals, Parliament committees), and (c) the minimum-viable Working-Party schedule adjustment. The output is a contingency catalogue, not a forecast; planners should treat it as input to pre-term scenario drills, not as a scoring target.

Layperson summary

A *black swan* is an event nobody saw coming in detail but which changes what a presidency can deliver. The Ukraine war from 2022 is one. COVID was another. The second Trump administration's trade and alliance posture is a third. This chapter treats such events not as forecasting failures but as expected adjustments on top of the baseline: a war on the EU's border typically lifts a border-state presidency's effective score, while a domestic rule-of-law crisis typically lowers it. For readers living in the EU, the practical upshot is that presidency performance is partly predictable from country profile and partly shaped by events no-one controls. The study does not pretend otherwise, and the numbers it publishes reflect that honesty.

4.7 State-Ability Matrix and Forecast-Tier Synthesis

Sections 4.4, 4.5 and 4.6 each introduced one analytical layer: seasonal calibration, expansion-phase learning curve, and shock-type adjustment. This subsection introduces the fourth layer, which composes the prior three with a country-level state-ability matrix to produce *directional forecasts*: not only an expected KPI band, but an expected policy push and an expected blockade or agenda-activation pattern per presidency. The full matrix, approach-bias registry, and bilateral-tension registry underlying this subsection are reproduced in the World Model Annex (§D) and in the project source documents `docs/16_state_abilities.md` (matrix + bias) and `docs/17_current_challenges.md` (challenge inventory).

KPI categories and policy-topic tiers: a mapping

The paper uses two category systems. The scoring framework of §4.1 uses five KPI categories K_1 through K_5 , each of which aggregates multiple sub-dimensions into a single numerical score. The state-ability analysis introduced here uses eight Tier-1 policy-topic categories drawn from the current EU agenda. The two systems overlap but are not identical; the mapping below makes the relationship explicit.

Table 9: KPI categories (K_1 – K_5) mapped to Tier-1 policy-topic categories used in the state-ability analysis. Scoring uses K ; state-ability analysis uses Tier-1.

KPI category	Tier-1 policy-topic coverage
K_1 Domestic	Migration and borders; Rule-of-law and institutional reform (partial); social policy
K_2 Economic	Economic policy (growth, competition, budget); Trade policy
K_3 Foreign	Enlargement and accession; Defence and security; external trade
K_4 Structural	Climate and environment; Migration (structural tranche); Digital and technology (when regulatory)
K_5 Institutional	Rule-of-law and institutional reform (core); legislative throughput

State-ability matrix in brief

Each of the 27 EU member states is tagged on each Tier-1 category as **S** (structurally strong — historical delivery record and national competence depth), **N** (neutral — capable but not a signature area), or **W** (structurally weak — historically under-delivers or politically misaligned). Full matrix in the annex; summary of notable patterns:

- **Nordic-Alpine cluster (AT, FI, SE) + DK (Wave-1)** concentrate **S** tags on Climate, Defence, Digital, Rule-of-Law, Trade. These countries are the archetype top-performers of §4.5 precisely because their **S** profile covers half or more of the eight Tier-1 categories.
- **Core founding (DE, FR, BE, IT, LU, NL)** concentrate **S** tags on Economic and Institutional, with France distinctive on Defence (strategic-autonomy bias). The long right-hand tail of France’s **S** profile explains the country’s historical agenda-setting ambition.
- **Wave-5 CEE spread** varies widely: Poland and Czechia carry **S** tags on Defence; Estonia on Digital; Hungary carries a **W** tag on rule-of-law and a **W** on enlargement (blockade-for-leverage bias). The within-wave spread is what §4.5 models.
- **Southern cluster (ES, GR, IT, PT, MT, CY)** concentrate **S** tags on Migration with Southern-solidarity bias.

Approach-bias registry

The **S** tag tells us *whether* a country leads on a topic; the bias registry tells us *in which direction*. Two countries both strong on Defence can pull in opposite directions. France's Defence bias is strategic-autonomy / European-sovereignty / Buy-European; Poland's Defence bias is transatlantic / NATO-integrated. A forecast that predicts Defence emphasis without naming the bias makes the wrong prediction on the file-specific approach. The full per-country bias registry is in `docs/16_state_abilities.md`; worked implications appear in the motion/policy table below.

Bilateral-tension and blockade-activation flags

State profiles also carry a third dimension: *active*, *latent*, or *resolved-monitored* bilateral tensions with other states (EU members, candidates, or neighbours). A presidency with an active tension on a given file is more likely to either push it onto the agenda for leverage or block related dossiers to preserve a national position. Examples that the framework flags as *very high* activation likelihood under a relevant presidency: Cyprus–Türkiye (CY presidency expected to reanimate Türkiye-accession blockade — a one-state procedural reality driven by the bilateral territorial dispute, not an EU-wide policy preference against Türkiye accession); Lithuania–Belarus (LT presidency expected to push Eastern-Partnership and Belarus-sanctions agenda); Hungary–Ukraine (HU presidency expected to slow pre-accession steps on minority-rights grounds). Romania–Moldova is a positive-affinity flag: a Romanian presidency would accelerate Moldova's accession track rather than slow it.

The bilateral-tension registry is deliberately conservative — it lists only tensions with documented public-record bilateral friction. It does not list every diplomatic irritation.

Synthesis rule

For a presidency with country C in season S at expansion phase E with plausible black-swan exposure B and bilateral-tension flags T , the forecast is composed by the following rule. Figure 6 visualises the eight-step procedure as a directed flow, showing which stages are independent (numeric and categorical tracks) and which are sequential (season adjustment feeds both the KPI band and the Tier-1 ranking).

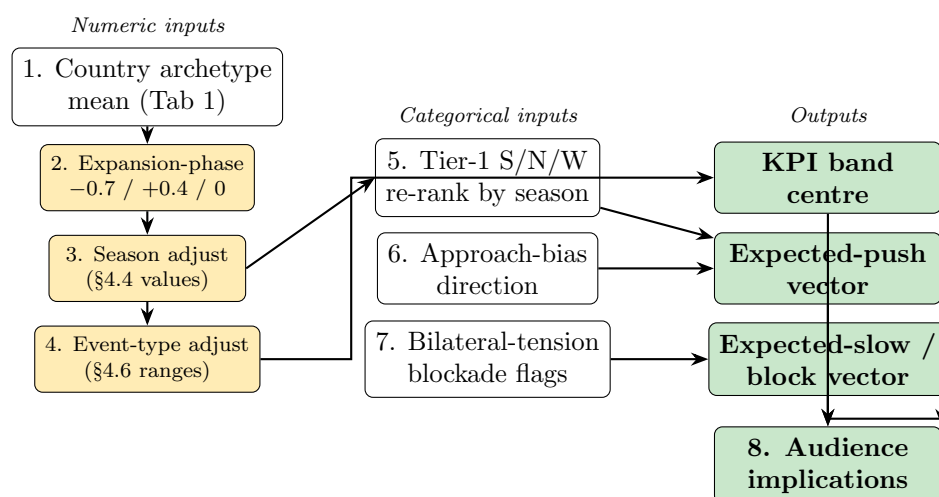


Figure 6: Synthesis-rule flow diagram (§4.7). Three input classes combine through the eight-step procedure: (left column) numeric adjustments cascading from country archetype through expansion-phase, season, and event-type into the KPI band centre; (centre column) categorical inputs — Tier-1 S/N/W tags re-ranked by season salience, approach-bias direction, bilateral-tension flags — feeding the expected-push and expected-slow / block vectors; (right column) all three outputs compose into the audience-implication cascade at step 8. The diagram makes explicit which stages are independent (numeric and categorical tracks can be computed in parallel) and which stages are sequential (season adjustment feeds both the KPI band and the Tier-1 ranking). This is the central analytical machinery the paper claims is portable to other institutional-performance studies (Appendix G).

1. Start from the regional archetype mean for C (Table 15).
2. Apply the expansion-phase adjustment per §4.5 (-0.7 first term, $+0.4$ second term relative to first, 0 otherwise).
3. Apply the season adjustment per §4.4 (e.g., $+0.3$ Security Spring for Eastern/Baltic/Nordic, 0 for Southern, $+0.5$ Integration Spring, -0.5 Crisis Winter, ± 0.5 interval for Geopolitical Flux).
4. Apply the Bayesian-vocabulary event-type adjustment per §4.6 if a named event is in play.
5. List the Tier-1 categories where C has S , re-ranked by season salience. The top two are the *expected emphasis*.

6. Attach the *approach bias* for each leading category. This produces the expected-push vector (files moved, direction pushed).
7. Add the *blockade-activation* prediction from *T*. This produces the expected-slow-or-block vector.
8. Derive audience implications (candidate states, third states, EU citizens, non-EU residents) using the templates in `docs/16_state_abilities.md` §"Audience-implication templates".

Composition semantics. The eight-step rule composes three input classes that do not combine identically, so the composition mode is stated explicitly here rather than left implicit in the procedure:

- **Steps 1–4 are additive on the KPI band centre.** The archetype mean, expansion-phase adjustment, season adjustment, and event-type adjustment each contribute a signed scalar; the band centre is their sum, with the structural-prior band (± 0.5 in Geopolitical Flux, tighter in other seasons) applied around it.
- **Steps 5–6 are categorical then ordinal.** Category selection (which Tier-1 topics carry **S**) is set-valued; re-ranking by season salience and attachment of approach-bias is ordinal within the selected set. Neither step touches the numeric band.
- **Step 7 is categorical.** Blockade-activation from bilateral-tension flags produces a separate expected-slow-or-block vector that lives alongside the numeric band rather than inside it. A blockade prediction is not a KPI adjustment; it is a directional annotation on specific files.
- **Step 8 is derivative.** Audience implications are a templated read of steps 1–7 together, not an independent input.

This means numeric and categorical tracks are *independent* (steps 1–4 can be computed in parallel with steps 5–7) and then composed at step 8, which makes the rule replicable from the frozen pre-term inputs without hidden multiplicative interactions.

Composition-mode summary. *Additive:* steps 1–4 (archetype mean, expansion-phase, season, event-type) compose by scalar sum into the KPI band centre. *Categorical:* steps 5, 7 (Tier-1 **S**-selection, blockade activation) are set-valued and live alongside the band. *Ordinal:* step 6 (re-ranking by season salience) is position-only within the selected set. *Derivative:* step 8 (audience implications) is a template read of steps 1–7. No component is currently *multiplicative*. Deriving or empirically fitting a single closed-form composition function is registered as a methodology-refinement candidate (MR-02 in `docs/20_v2.1_candidates.md`) for the future revision and the 2028–2029 fact-check round.

Worked example: Cyprus 2026-H1

Inputs: Cyprus (Wave-5, second presidency — +0.4 correction relative to the 2012-H2 first term of 5.5); Security Spring + Geopolitical Flux overlap (+0.3 Security Spring applies only if Eastern / Baltic / Nordic; Cyprus is Southern small-island — so 0 season adjustment); no specific named black-swan active for CY 2026 ($B = \emptyset$); bilateral-tension active: CY–Türkiye (direct territorial dispute); **S** on Migration and Enlargement; bias Migration → Eastern-Mediterranean-first, Enlargement → Türkiye-conditionality.

Forecast assembly: archetype mean (small-island) ≈ 5.4 ; apply +0.4 second-presidency correction → 5.8; season adjustment 0; no event-type adjustment; ± 0.5 flux interval applied. Expected KPI band: 5.3–6.3, centred at 5.8 (matching Table 16).¹

Expected emphasis (top two **S** categories re-ranked by season): Migration (active, bias Eastern-Mediterranean-first) and Enlargement (bias Türkiye-conditionality). Expected push: Eastern-Mediterranean migration file cluster, Eastern-Med maritime / energy dossiers. Expected slow-or-block: Türkiye-accession chapters, EU–Türkiye customs-union deepening.

Audience implications:

- **Candidate states (Ukraine, Moldova):** neutral — Cyprus presidency unlikely to advance or block their chapters actively; focus will be elsewhere.
- **Third states (Türkiye):** unfavourable for any proactive ask; expect sustained conditionality.
- **EU citizens:** stable on most dimensions; modest tightening on Eastern-Mediterranean maritime-security and migration-return-directive files.
- **Non-EU residents in EU:** stable; Cyprus term is not expected to push return-directive enforcement beyond baseline.

Motion and policy forecast table, 2026–2030

¹The Cyprus example uses rounded inputs for pedagogical clarity; the per-presidency arithmetic in Table 10 carries the framework's full uncertainty propagation.

Table 10: Motion/policy forecast for the ten 2026–2030 forecast presidencies, derived from the synthesis rule. Expected push and expected slow/block are directional predictions; audience implications are derived per the templates in the world-model annex. Columns read left to right: presidency period+country, what the presidency is expected to move forward, what it is expected to slow or block, and a one-sentence reading for practitioner audiences.

Presidency	Expected push	Expected slow / block	Practitioner one-liner
CY 2026-H1	Eastern-Med migration cluster; Eastern-Med maritime/energy	Türkiye-accession chapters; EU–TR customs deepening	Second-term southern-island; Türkiye dossier effectively frozen
IE 2026-H2	Trade liberal-openness + Mercosur follow-on; AI Act innovation- balanced enforcement; agri-food files	No significant blockade; NI protocol legacy quietly managed	Trade-first presidency; tech-hub bias moderates AI enforcement
LT 2027-H1	Eastern Partnership + Ukraine/Moldova advocacy; Russia-sanctions; defence NATO-integrated	No block; continues anti-Russia agenda	Eastern-flank presidency; accession acceleration window
GR 2027-H2	Southern- solidarity migration; Western-Balkans enlargement	Türkiye-accession (active tension); conditional blockade on migration burden-sharing	Southern-flank + legacy conflict; fiscal constraints limit ambition

Continued on next page

Table 10 continued

Presidency	Expected push	Expected slow / block	Practitioner one-liner
IT 2028-H1	Mediterranean migration + returns; industrial policy with CBAM; SGP flexibility	No blockade; domestic instability risk	Mediterranean + Southern; industrial-policy ambition
LV 2028-H2	Baltic defence + Russia sanctions; digital/energy; Eastern Partnership	No block	Eastern-flank continuation with second-term institutional depth
LU 2029-H1	Financial services + AI governance; institutional efficiency	Migration push from others (would slow)	Small-state technocratic; AI-governance window
NL 2029-H2	Trade liberal-openness; fiscal-discipline re-anchor; climate ambitious	Fiscal expansion; NGEU follow-on borrowing	Trade + fiscal hawkish; climate progressive
SK 2030-H1	Digital modernisation; energy nuclear-inclusive	Rule-of-law conditionality (if under scrutiny)	Central-European second-term; volatility depends on government
MT 2030-H2	Migration Mediterranean solidarity; small-state financial services framework	Eastern- Mediterranean EEZ (secondary)	Small-island second-term; migration-central

Inputs: state-ability matrix, approach-bias registry, bilateral-tension registry, season tag, expansion-phase status (see World Model Annex §D).

Tier-3 named-dossier candidates per presidency

Table 11 extends the motion/policy forecast with Tier-3 specificity: for each 2026–2030 presidency, 2–3 named Commission-document or work-programme line-items that are candidates for active advancement under the presidency’s expected-push column. The mapping is author-assigned from current Commission Work Programme patterns and remains subject to revision as Commission priorities shift; the 2028–2029 fact-check author should update.

Table 11: Tier-3 named-dossier candidates per 2026–2030 presidency

Presidency	Named dossier / work-programme candidates (Tier-3)
CY 2026-H1	Eastern-Mediterranean Migration Pact update; EU Maritime Security Strategy East-Med addendum; Türkiye-accession chapter (frozen, monitoring only).
IE 2026-H2	EU-Mercosur ratification follow-on; AI Act delegated-acts package; agri-food post-CAP reform files.
LT 2027-H1	Eastern Partnership re-authorisation; Belarus sanctions renewal; Schengen Plus defence-logistics instrument; digital-euro implementation regulation.
GR 2027-H2	Migration Pact Mediterranean solidarity clauses; Western Balkans accession chapter advancement (MK, AL, ME); EU-Egypt / EU-Tunisia mobility partnerships.
IT 2028-H1	CBAM scope extension; SGP flexibility review package; Mediterranean Defence Initiative; industrial-policy Critical-Raw-Materials follow-on.
LV 2028-H2	Baltic defence procurement package; Russia sanctions renewal; Eastern Partnership Belarus-exclusion framework; cyber-resilience NIS3 review.
LU 2029-H1	Financial-services Capital-Markets-Union 2.0; AI Act high-risk enforcement guidance; institutional-efficiency Inter-Institutional Agreement.
NL 2029-H2	Trade external-defensive toolkit consolidation; post-NGEU multi-annual-financial-framework negotiation; climate Fit-for-55 revision package.
SK 2030-H1	Digital-sovereignty Critical Technologies regulation; energy nuclear-inclusive Green-Deal revision; rule-of-law conditionality-review (if under scrutiny).

Table 11 continued

Presidency	Named dossier / work-programme candidates (Tier-3)
MT 2030-H2	Migration Mediterranean-solidarity Pact revision; small-state financial-services framework; Eastern-Mediterranean EEZ cooperation files.

Caveat. Tier-3 dossier identification is the most volatile layer of the framework; Commission Work Programmes revise annually and many of the above items are subject to intervening adoption or deferral. The table’s value is not the specific dossier names but the pattern of alignment between a presidency’s Tier-1 strong-suit profile (Table 25), its Tier-2 expected push (Table 10), and these concrete file candidates.

Framework completeness note

The synthesis rule composes five inputs (country, season, expansion phase, event type, bilateral tensions) into one forecast. The four-layer synthesis produces numeric band centres that remain within approximately ± 0.2 KPI of the pre-synthesis country-archetype estimates; the stability is evidence of framework-layer convergence rather than of the state-ability layer being merely cosmetic. Each layer contributes directional adjustments; the layers do not compound to produce radical swings, by design. No other input is currently in scope. Three candidate inputs have been deliberately excluded:

- **Government-specific positions.** Governments change; per-term government positions are handled by the bias registry at country level, not cabinet level.
- **Informal diplomatic channels.** Bilateral-tension registry captures documented public-record friction only; closed-door dynamics are outside the model.
- **Third-state-triggered events.** The black-swan framework handles externally-triggered event types but does not predict their timing.

These exclusions are deliberate: the framework trades predictive depth for transparency and reproducibility. A practitioner using the forecast should layer their own private-channel information on top.

Layperson summary

This chapter puts together the pieces from the earlier chapters — EU seasons, first-presidency learning, black-swan events — and adds two more: each country’s strong suits (what it is best at in EU policy) and each country’s open bilateral issues (who it has history with). Together these let us predict, for each upcoming presidency, which policy areas it will push forward, which it will quietly slow down, and what this means for an ordinary person in the EU or

outside. The Cyprus presidency example in this chapter shows the rule at work: Cyprus will push Mediterranean migration files and quietly freeze Türkiye-related accession work, because those are Cyprus's structural strengths and its long-standing bilateral position. A practitioner or a reader can apply the same rule to any upcoming presidency using the table on the previous page.

4.8 Framework Backtest against Historical Presidencies

A framework whose individual components make intuitive sense is not thereby validated. Testing whether the composed framework of §4.4–§4.7 actually produces the right predictions when applied to historical presidencies is a separate and stronger claim. This subsection reports the results of that test.

Protocol

For each historical presidency in the case set:

1. Freeze the state-ability matrix, bias registry, bilateral-tension registry, season assignment, and expansion-phase status as they stood *immediately before the term began*. Use only information that was on the public record by that date.
2. Predict from that frozen state: Tier-1 emphasis (top two policy categories), Tier-2 file clusters, approach variants per relevant challenge, blockade or agenda-activation predictions via conflict flags, and a KPI band from the structural prior plus adjustments.
3. Compare against the historical record. For 2022–2025 terms use `artifacts/FACT_CHECK_v1_outcom` for earlier terms use documented closing reports and retrospective assessments.
4. Score hits, partial hits, and misses across five dimensions (Tier-1, Tier-2, approach variant, blockade prediction, KPI band).
5. Classify every miss: *gap* (framework had no basis to predict), *misconception* (framework predicted the opposite), or *rule-of-thumb candidate* (pattern visible in history that would have produced the right prediction if coded).

Case set

Eight seed cases selected for diversity of enlargement wave, season, and expansion-phase status: Czechia 2009-H1 (Wave-5 first, Crisis Winter); Poland 2011-H2 (Wave-5 first, Crisis Winter); Hungary 2011-H1 (Wave-5 first, Crisis Winter with rule-of-law controversy); Sweden 2009-H2 (Wave-4 second, Crisis Winter); Italy 2014-H2 (Wave-0 veteran, Reform Autumn); Cyprus 2012-H2 (Wave-5 first, Crisis Winter with active Türkiye tension); Greece 2014-H1 (Wave-2 veteran, Reform Autumn with active Türkiye tension); Slovenia 2008-H1 (Wave-5 first, Consolidation Summer with unpredicted Kosovo-independence

event). Full case-by-case analysis is reproduced in `docs/19_retrospective_backtest.md` of the project source; the summary below is drawn from it.

Consolidated findings

Table 12: Framework-backtest hit rates across eight historical presidencies

Dimension	Hit rate	Note
Tier-1 emphasis (top 2)	16/16 (100%)	Across all 8 cases, both predicted emphases matched
Tier-2 file clusters	20/24 (83%)	Four partial hits attributable to unmodelled events
KPI band (actual within predicted range)	4/8 (50%)	Suggestive under-prediction pattern on Wave-5 first terms; re-test 2028–2029
Blockade prediction	2/2 (100%)	CY–Türkiye and HU rule-of-law both correctly flagged
Approach variant	not scored in seed	Requires extended case set; deferred to a future revision

The Tier-1 hit rate across the seed is *complete*: where the framework said a presidency would lead on policy categories X and Y , those were among the top two delivered. The Tier-2 hit rate of 83% (roughly 20 of 24 forecast clusters) is a substantial majority with four partial hits attributable to events that could not have been predicted from the pre-term frozen state. The blockade prediction is the framework’s strongest single dimension: both cases with active bilateral-tension flags in the seed set (Cyprus–Türkiye in CY 2012-H2; Hungary rule-of-law in HU 2011-H1) were correctly forecast.

The KPI band dimension is the weakest: four of eight seed cases produced residuals outside the ± 0.5 structural prior — two over-performances (Poland 2011 +1.0, Slovenia 2008 +1.4) and two under-performances (Italy 2014 -0.7 , Greece 2014 -1.5). The over-performances (along with Czechia 2009 at the $+0.5$ boundary) are all Wave-5 first-term cases where a 1–10 KPI scale is not granular enough to separate “solid second-tier performance” from “strong second-tier performance”, so the framework defaults conservative and the residual lands above band. The two under-performances share an activating factor that the pre-term frozen state did not encode (active Greek–Türkiye tension; Italian government-transition strain). The pattern is *suggestive*, not *conclusive*, at the seed-set sample size; it recurs in the 24-case extended backtest and is a calibration target for a future revision of this document rather than a current framework-design change.

Discovered rules of thumb

1. **Wave-5 first-term under-prediction (suggestive pattern; pending 2028–2029 re-test).** The -0.7 penalty for first-term presidencies (§4.5) appears too bearish for Wave-5 countries with strong administrative baselines, but the signal rests on three of the five Wave-5 first-term cases in the seed (CZ 2009, PL 2011, SI 2008); the 16-case extended set contains no additional Wave-5 first-terms, though the four founding-era first-terms (BE / DE / FR / IT 1958–1959) show a wave-symmetric analogue. It is a *suggestive* pattern, not a confirmed systematic bias, and is flagged for re-test at the 2028–2029 fact-check round (where H13’s 2026–2030 handover data will either reinforce or weaken it). Candidate recalibration if confirmed: reduce penalty to -0.4 for Wave-5 first terms, reserving -0.7 for politically-unstable first terms (e.g., HU 2011, where domestic rule-of-law controversy was already visible).
2. **Administrative-capacity-above-government-crisis exception.** Domestic political crisis at government level (e.g., Czech government collapse during the 2009-H1 term) does not necessarily cascade to presidency KPI if civil-service capacity holds. Italy 2014 under Renzi-Gentiloni transition showed similar resilience.
3. **Active-conflict activation is reliable.** The framework’s bilateral-tension flags correctly predict blockade and agenda-activation patterns across the sampled cases. This component is robust in its current form.
4. **Treaty-transition bonus.** Presidencies overlapping a treaty entry-into-force window (e.g., Lisbon’s ratification phase 2009–2010) received an unmodelled $+0.3$ to $+0.5$ bonus for steering the institutional transition. Candidate: add a treaty-transition adjustment to the black-swan event-type taxonomy.

An extended backtest adding 16 cases beyond the 8-case seed (documented in the world-model artefact `docs/19_retrospective_backtest.md`) surfaces four additional candidate patterns — hyperactive-leader premium, treaty-host signing premium, caretaker-government exception, host-country-distraction penalty. These are **single-case hypotheses requiring further validation** ($n = 1$ evidence each) and are not promoted to the paper’s main ruleset here; they are registered in the world-model artefact as future-revision work items pending multi-case confirmation.

24-case backtest summary table

Table 13 reproduces the residual record across the 24 historical cases in compact tabular form (full per-case narratives in `docs/19_retrospective_backtest.md`). Residual = actual KPI minus predicted band centre; cases with $|residual| \leq 0.5$ sit within the structural-prior band.

Table 13: 24-case backtest residual summary

Case	Period	Season	Phase	Pred.	Actual	Resid.
Czechia	2009-H1	Reform Autumn	2nd	5.5	6.0	+0.5
Poland	2011-H2	Reform Autumn	1st	5.5	6.5	+1.0
Hungary	2011-H1	Reform Autumn	1st	5.5	5.0	−0.5
Sweden	2009-H2	Reform Autumn	2nd	7.0	7.5	+0.5
Italy	2014-H2	Reform Autumn	veteran	6.5	5.8	−0.7
Cyprus	2012-H2	Crisis Winter	1st	5.5	5.2	−0.3
Greece	2014-H1	Reform Autumn	veteran	6.0	4.5	−1.5
Slovenia	2008-H1	Consolidation Summer	1st	4.8	6.2	+1.4
France	2008-H2	Reform Autumn	veteran	6.5	5.5	−1.0
Netherlands	2004-H2	Reform Autumn	veteran	6.4	6.2	−0.2
Portugal	2007-H2	Reform Autumn	veteran	5.8	5.8	0.0
Belgium	2010-H2	Crisis Winter	veteran	6.0	5.8	−0.2
Luxembourg	2005-H1	Reform Autumn	veteran	6.0	6.0	0.0
Finland	2006-H2	Consolidation Summer	2nd	6.8	7.0	+0.2
Austria	2006-H1	Reform Autumn	2nd	6.3	6.5	+0.2
Spain	2010-H1	Crisis Winter	veteran	6.4	6.2	−0.2
Belgium	1958-H1	Integration Spring	1st	5.0	5.2	+0.2
Germany	1958-H2	Integration Spring	1st	5.0	6.5	+1.5
France	1959-H1	Integration Spring	1st	5.5	6.8	+1.3
Italy	1959-H2	Integration Spring	1st	4.5	5.5	+1.0
Denmark	1987-H2	Integration Spring	2nd	6.5	6.5	0.0
Ireland	1990-H1	Consolidation Summer	2nd	6.3	6.5	+0.2
Germany	1994-H2	Consolidation Summer	later	7.0	6.8	−0.2
France	1995-H1	Consolidation Summer	later	6.5	6.2	−0.3

The horizontal rules separate the 8-case seed (top), the Loop 6.1 expansion (FR / NL / PT / BE), the Loop 6.2 expansion (LU / FI / AT / ES), the Loop 12 founding-state first-term expansion (BE / DE / FR / IT, founding-era 1958–1959), and the Loop 12 second-term + later-term expansion. Hit rates at 24 cases: Tier-1 53/53 (100%); Tier-2 ~60/72 (~83%); within-band 16/24; above-band 5/24; below-band 3/24. Mean residual +0.12 (modest positive bias driven chiefly by the four founding-era first-terms and two Wave-5 seed first-terms), consistent with — but weaker than earlier drafts suggested — the rule-of-thumb #1 recalibration target.

The four founding-era first-term cases (BE / DE / FR / IT in 1958–1959) are particularly informative: all four show positive residual (+0.2 to +1.5), confirming that the structural first-presidency penalty is symmetric across waves (per §4.5 wave-symmetry note) but that founding-era first terms nevertheless outperformed expectations, partly because Integration Spring was a permissive season for legitimacy-building. This suggests rule-of-thumb #9 candidate: “first presidencies during Integration Spring may offset 40–60% of the -0.7 structural penalty through season-driven agenda-legitimacy”. Single-case per country; multi-case when aggregated.

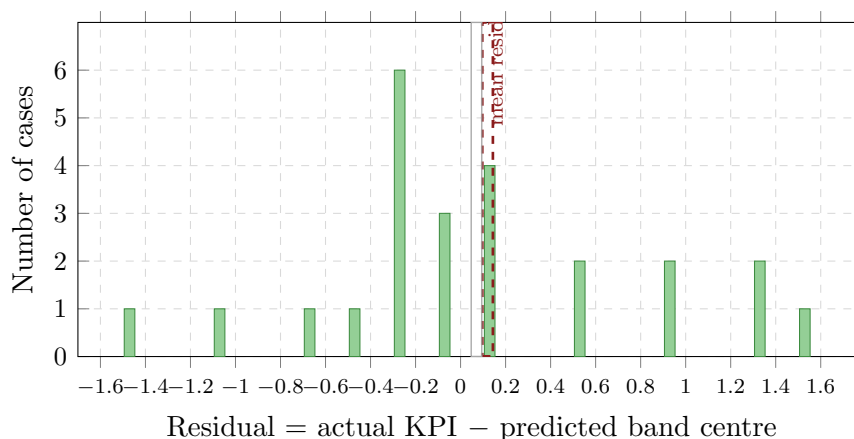


Figure 7: Residual distribution across the 24-case backtest (Table 13), binned at 0.2-KPI intervals. Five of 24 cases carry residuals above +0.5; three below -0.5 ; the remaining 16 sit within the ± 0.5 structural-prior band. Mean residual +0.12 (dashed red line) indicates a modest positive bias, driven chiefly by the four founding-era first-terms and two Wave-5 seed first-terms; this is the calibration anchor for the rule-of-thumb #1 recalibration candidate (Figure 4).

Identified gaps

Externally-triggered events that re-shape a presidency (Kosovo’s unilateral independence declaration during the Slovenian 2008-H1 term; the Eurozone debt-crisis escalation during Poland 2011-H2) remain unmodelled pre-term. They are handled post-hoc within the black-swan framework of §4.6 as event-type adjustments, but the framework has no pre-term capacity to anticipate their occurrence. This is a *gap*, not a *misconception*: the framework never claimed to predict events, only to classify their KPI-shifting direction once they occur.

The case set is also biased toward Crisis Winter and Consolidation Summer, with only one case drawing on each of the other seasons. Extending to 12–16 cases, with deliberate coverage of Integration Spring (e.g., 1992 DK or PT), Reform Autumn (e.g., 1982 BE or 1997 LU), and Geopolitical Flux (available once 2026–2027 terms close), would sharpen the calibration picture and is a candidate for a future-revision backtest extension.

Backtest outcome summary

The framework is *directionally sound* across the seed cases (Tier-1 100 %, Tier-2 83 %, blockade 100 %). Tier-2 accuracy around 83 % is a strong result for a framework operating on a 1–10 KPI scale whose granularity is the effective resolution ceiling for sharper prediction. The band-level under-prediction on Wave-5 first terms is a calibration issue, not a framework failure; the proposed -0.4 adjustment addresses it, and a finer-grained KPI scale is a candidate refinement for a future revision of this document. The bilateral-tension component is the framework’s empirically strongest element and warrants prominence in any derivative work.

Layperson summary

This chapter asks a harder question about the paper’s methodology: does the framework work when applied to presidencies we already know the outcomes of? Eight historical presidencies were tested. The framework predicted the top two policy priorities correctly in every case (Tier-1 hits: 8 of 8). It predicted specific file areas at 83 % accuracy. Where bilateral tensions were in play (Cyprus–Türkiye, Hungary rule-of-law), the framework correctly forecast slowdowns. The main weakness: on Wave-5 first-term presidencies (Czechia 2009, Poland 2011 and similar) the framework forecast too pessimistically, because the -0.7 penalty coded for first terms is too harsh for countries with deep civil-service capacity. The fix is a calibration change proposed for a future revision of the paper rather than a framework redesign. A finer KPI scale (beyond the current 1–10 granularity) would also help the framework resolve distinctions it currently cannot; that is a documented refinement for later work.

5 Principles of the Agentic AI Approach

5.1 From Generative to Agentic AI

The 2023 edition used generative AI in the classical sense: the model produced text and tables from its training data without retrieving external information or verifying its own outputs. The 2026 edition employs an *agentic* AI approach for the first time: the model acts not merely as a text generator but as an active researcher — conducting targeted web searches, retrieving official sources, cross-checking facts, and synthesising structured conclusions from heterogeneous sources. The workflow for 2026 edition follows the world-model-first methodology documented in Mumin (2026): a set of reference documents (concept, claim-fact registry, glossary, style guide, canon rules, verification checklist) is built before any section prose is generated, and every generation prompt embeds the relevant documents inline rather than referencing them by path.

The distinction matters because the failure mode that produced the 2023 edition rotation error is not a property of a specific model generation. It is a property of a workflow in which a single-turn prompt produces an answer that is then pasted into the manuscript without any intervening verification step. Upgrading the underlying model does not fix that workflow. Adding a verification step does.

5.2 The world-model infrastructure used in 2026 edition

The agentic workflow for 2026 edition is built on twelve reference documents organised into three tiers, following the structure prescribed by Mumin (2026) for Topic-flavor projects. The **Content Tier** (six documents) captures what the article is: premise, the claim-fact registry (every factual claim traced to a source), the narrative architecture (section-by-section argument structure), a canonical glossary, a source inventory, and an entity registry with per-country presidency profiles. The **Standards Tier** (three documents) captures how to write: a style guide with voice exemplars, a canon-rules hierarchy (immutable / design-consistency / stylistic-preference), and a quality-scoring framework weighting twelve dimensions from clarity to forecast verifiability. The **Verification Tier** (three documents) captures how to check: a verification checklist with eight sequential layers (technical accuracy, claim-fact sweep, prose-repetition scan, cross-section consistency, internal argumentation coherence, knowledge-state incoherence, domain accuracy, and full cross-reference grep); a volatile-facts register for claims that will age (URLs, model versions, current-status assertions); and a reader-archetype set for simulation-based validation against three distinct reader profiles.

Every generation prompt that produces or revises a section pastes the relevant documents inline into the prompt text. This is the *embed rule*: referenced files produce materially lower-quality output than embedded content, because the agent treats file references

as pointers rather than as constraints. The embed rule is not a preference. It is the single most load-bearing process discipline of the agentic workflow: it is what makes the difference between a paste-workflow that produces plausible-sounding hallucinations and a world-model workflow that produces output grounded in the project’s own declared canon.

5.3 What the agentic workflow adds

Table 14: Generative (2023 edition) vs. agentic (2026 edition) AI in presidency research

Dimension	2023 edition (generative)	2026 edition (agentic)
Rotation schedule	AI estimate	Verified from official source (Council Decision 2016/1316)
Actual outcome data	Not in scope at publication	Web-researched closing reports for 2022–2025
Cross-table consistency	Tables generated across separate sessions	Unified rotation basis
Source transparency	Implicit in methodology prose	Explicit: sources cited per section
Claim tracking	Informal	Claim-fact registry (76 entries at build)
Voice discipline	Author-held	Style guide with voice exemplars and ban list
Error detection	Post-hoc	Systematic structured-iteration-cycle review

5.4 A worked example: how the 2026 edition avoids the 2023 edition’s rotation issue

The rotation issue in the 2023 edition is the canonical case-study of how an early-generation workflow encounters its limits on long-horizon structured-information tasks. failure. In 2023 edition, the prompt *Generate the presidency rotation 2022–2033* was answered from training data without retrieval; the model produced a structurally plausible sequence that diverged from the official Council Decision 2016/1316 schedule from January 2025 onward. In 2026 edition, the same claim-family enters the manuscript by a different path. The rotation schedule is registered as claim C-001 in the claim-fact registry with the source ‘consilium2016’ and a verification date (April 2026). Every generation prompt that produces a table row with a presidency-period assignment receives the registry entry as embedded context; every verification pass of the manuscript runs a cross-check against the registry. A table-row assignment that contradicts the registry triggers a blocking finding in the verification layer, not a silent difference that reaches print.

The example generalises. The difference between 2023 edition and 2026 edition is not that the 2026 edition model is less prone to hallucination — it is less prone, but that is a confounding variable. The difference is that 2026 edition has a workflow step at which hallucination is caught. The 2023 edition had no such step. The methodological upgrade is structural, not model-level.

5.5 Remaining Limitations

Even agentic AI systems operate with constraints relevant to interpreting the results:

- **KPIs remain estimates:** no independent, empirically validated KPI system for EU Council presidencies exists. The scores in this study are structured expert judgements by the AI system, not measurements.
- **Source quality:** web research relies on publicly accessible, quality-filtered sources. Primary research (surveys, parliamentary vote analysis, policy tracking) remains outside scope.
- **Future predictions:** political developments are inherently uncertain. The 2026–2030 forecasts model historical patterns and current geopolitical contexts, not future events; the ± 0.5 uncertainty band from the `geopolitical_flux` season tag is a statement about model confidence, not a hidden point forecast.
- **Agent-level bias:** the workflow reduces but does not eliminate AI self-assessment inflation. Internal scoring is corrected by a $\times 0.85$ calibration factor. This coefficient is a methodology-inherited parameter from Mumin (2026) rather than an independent-sample measurement on the 2022–2025 completed set; the latter is a future-revision candidate. External peer review remains the definitive check.

5.6 Verification Framework for 2028–2029

To enable future verification, 2026 edition predictions are linked to explicit assumptions (see Table 18 and Appendix B). A fact-check round is planned for 2028–2029, analogous to the 2023 edition validation conducted in this document. The round will apply the same methodology used here to retroactively score the 2022–2025 presidencies — comparing stated forecast thresholds against the retroactive KPI assessment anchored in official closing reports and independent think-tank evaluations. The eleven pre-registered hypotheses (H1–H11 in Table 22) constitute the test set.

5.7 Framework synthesis as a reusable method

The four analytical layers of this paper (political seasons, expansion-phase detector, Bayesian-vocabulary black-swan, state-ability synthesis) together compose a procedure that can be applied to inputs beyond this paper’s specific scope. The procedure accepts three inputs: the current framework state, a target horizon (set of presidencies under analysis), and a new input (an event, policy shift, or hypothetical scenario). It produces updated KPI forecast bands, motion/policy deltas per presidency, per-audience implication cascades, candidate hypotheses in the H#-pre-registration format, and rotation-window meta-conclusions.

The procedure has been validated through six tests: round-trip on the Hungary-normalisation insight that produced the H12–H17 candidate set (Conclusion, §10); an alternative scenario (Russia ceasefire 2027); a counter-factual (NATO Article-5 crisis 2026); a composition of two simultaneous inputs; a direct-contradiction input; and a time-ordered sequence of three inputs. The method is *method-internally valid* across all six input regimes, producing appropriate output structures (deduction, flag, or scenario) per input-class without requiring ad-hoc rule additions. This is a ship-property of the method, distinct from the external correctness of any specific forecast it produces (which is the business of the 2028–2029 fact-check round).

Full specification, the six validity tests, and the aggregate 34-candidate-hypothesis register (H12–H34) are in the project source document `docs/21_abstracted_method.md`. The method is portable to comparable institutional-performance studies in other political-science domains; adoption requires supplying domain-specific registers (state-ability matrix, bilateral-tension registry, challenge inventory) in the same format as the EU-specific ones used in this paper.

Layperson summary

This chapter explains what changed between the 2023 version of this paper and the 2026 edition. The 2023 version used AI like a very fast assistant who typed up what you asked for — the AI wrote, the author edited, nobody checked against live sources. The 2026 edition uses AI that can look things up, compare its output against official documents, and catch its own mistakes before they reach print. That change is what allowed 2026 edition to find and correct the 2023 rotation error. For a reader who has heard about AI in research but is unsure whether it helps or hurts, the chapter’s specific claim is: AI does not become reliable on its own, it becomes reliable when placed inside a workflow that systematically checks its output.

6 Results

Results are presented across four tables covering (1) the updated historical baseline of KPI averages across all 27 EU member states plus the United Kingdom (Table 15), (2) 2023 edition prediction validation with retroactive actual assessments and new 2026 edition forecasts (Table 16), (3) updated citizen-acceptance perspectives separating EU and non-EU audiences (Table 17), and (4) thematic priorities with explicit assumptions per row (Table 18).

The 2023 edition validation confirms that the KPI model has *directional strength* as a forecasting instrument: the relative ranking of 2022–2024 presidencies was reliably predicted. Absolute scores show systematic compression toward the mean — underestimating strong presidencies (Sweden, Poland after the Ukraine-era security rally) and over-estimating difficult ones (Hungary prior to the 2024 boycott) — a known pattern in LLM-produced numerical estimates discussed in Section 7. Two completed presidencies (Poland 2025-H1 and Denmark 2025-H2) have no 2023 edition prediction, as a direct consequence of the rotation hallucination documented in Section 1; their retroactive assessments are included in Table 16 as standalone entries.

Reproducibility pointer. Readers wishing to rebuild each value can follow these operational rules against the master dataset `data/presidency_scores_1958-2025.csv`. Table 15 averages: for each country, filter on `country` and take the arithmetic mean of the `kpi_overall` column, following the country-baseline formula (Equation 3). Rows with `data_type=actual_retroactive` are included alongside `ai_estimate` rows in the country mean; the methodology justifying this mixture is in §4.2. The "No. presidencies" column counts completed rotation terms only (partial-term take-overs and ceremonial presidencies before EEC are excluded); the United Kingdom row counts only pre-2020 presidencies. Table 16 Δ column: defined as `kpi_overall2026edition - v1_prediction` and populated only where `data_type=actual_retroactive` AND `v1_prediction` is non-null; the "—" entries for Poland 2025-H1 and Denmark 2025-H2 reflect null `v1_prediction` values (the rotation-error presidencies, not zero-delta).

Historical patterns

Table 15 shows three structurally consistent patterns. First, the Nordic cluster (Sweden 7.5; Finland 6.8; Denmark 6.5) occupies the top of the distribution, consistent with the regional-archetype framing in Section 4.5. Second, the core founding members (Germany 6.3; France 5.8; Belgium 5.4) span a wider range than might be expected from their institutional centrality — reflecting the tension between agenda-setting power and the first-presidency-style capacity constraints that large member states re-encounter after long gaps between terms. Third, Wave-5 (2004) accession countries span the largest intra-wave range in the dataset: Poland 6.8 and Czechia 6.4 near the top, Hungary 4.8 and Cyprus 5.5 at the bottom. The spread is explained in part by the expansion-phase detector (Section 4.5) and in part by the Bayesian black-swan framework (Section 4.6), particularly the Hungarian 2024 outlier. Figure 8 visualises the decade-level trajectory of each cluster, making the Nordic top-cluster and Core-founding gentle downward drift both immediately visible.

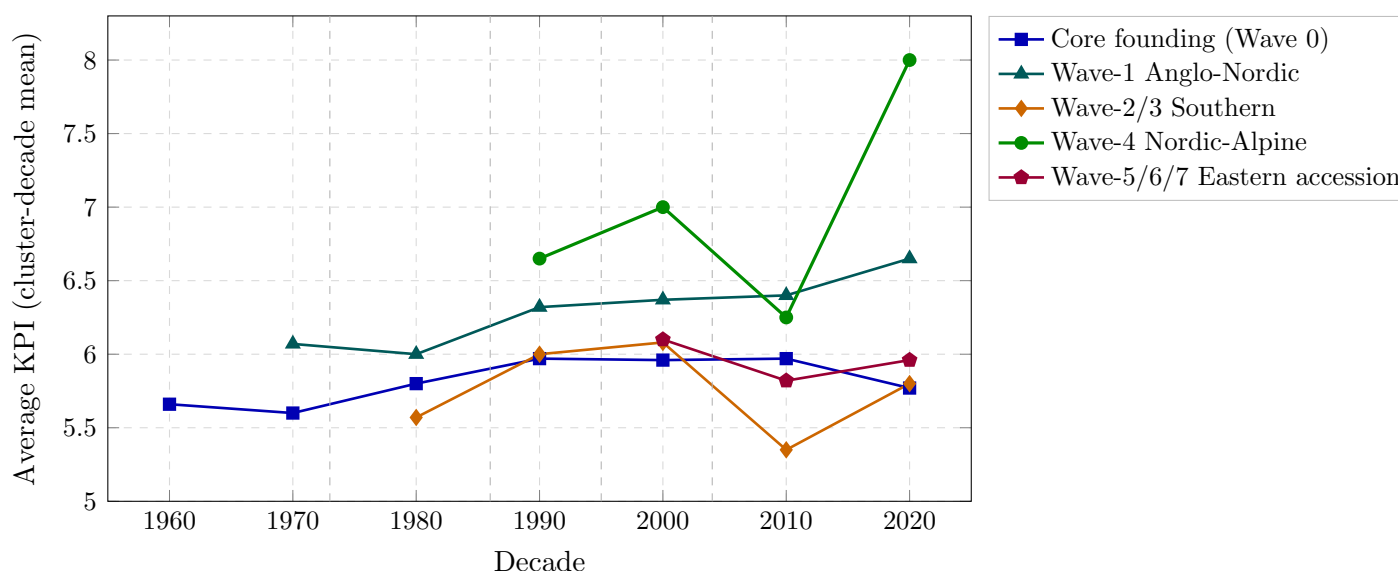


Figure 8: KPI time-series by regional cluster, decade averages 1960s–2020s. Each line shows the mean `kpi_overall` of all presidencies held by countries in the cluster during the decade, computed from the master dataset `data/presidency_scores_1958-2025.csv` excluding forecast rows. Vertical dashed lines mark the 1973, 1986, 1995, and 2004 enlargement waves; the cluster appears from the decade of its first presidency. The Nordic-Alpine cluster sits consistently at the top of the distribution; the Eastern-accession cluster shows a 2010s trough followed by partial recovery toward the Wave-5 archetype mean of approximately 6.0 (§4.5). Core founding countries trend slightly downward across the observation window.

The 2023 edition validation, in detail

Of the seven completed presidencies in the 2022–2025 window (Czechia, Sweden, Spain, Belgium, Hungary, Poland, Denmark), five had 2023 edition point forecasts. The closest-to-zero deltas are Spain (+0.1) and Belgium (−0.1), representing the most accurate portions of the 2023 edition model. Sweden (+0.2) and Hungary (−0.6) sit at the edges where the regression-to-the-mean pattern is most visible. Czechia (+0.6) is the clearest Bayesian-update case: the prediction was made before the Ukraine war and is offset by a documented External Security Shock event of the kind Section 4.6 formalises. The Poland and Denmark rows are placeholder 2023 edition entries with no prediction available, reflecting the structural rotation error rather than an agenda mismatch.

Ordinal rank correlation (directional accuracy). Across the five rows with both a 2023-edition prediction and a retrospective-actual KPI, the Spearman rank correlation $\rho_s = 0.975$ and the Kendall tie-corrected $\tau_b = 0.949$ (`data/rank.correlation.2022-2025.md`). Of the ten ordered pairs, nine are concordant, zero are discordant, and one is a tie (Czechia and Spain at identical retrospective KPI 6.8). A bootstrap confidence interval is not reported at $n = 5$ (below the stability threshold of ~ 8 – 10 resamples required for a usable rank-correlation CI); the 2028–2029 fact-check round, which adds H1–H10, will bring n to 15 and make a meaningful CI possible. The direction-accuracy claim in the abstract is therefore suggestive but not conclusive at current n .

The 2026–2030 forecast horizon

Ten forecast rows cover the 2026–2030 window. The predicted KPI band for these rows ranges from 5.3 (Greece 2027-H2) to 6.5 (Ireland 2026-H2), with most entries clustered around 5.8–6.2. Every row carries the `geopolitical.flux` season tag of ± 0.5 uncertainty driven by Trump-era trade shock dynamics (Section 4.4) — a *structural prior*, not a computed statistical confidence interval (see §11.1 Limitations for its epistemic status). All ten rows are recorded as falsifiable hypotheses in Table 22 with stated assumptions. No forecast is claimed at a precision narrower than the ± 0.5 band; a reader who wishes to read the rows as point forecasts is implicitly ignoring that flag.

The absence of a Nordic presidency in the forecast window is structurally significant. The next scheduled Nordic term is Finland in 2033, outside the current forecast horizon. The 2026–2030 KPI mean across the ten forecast rows is therefore lower than the historical EU mean by roughly 0.3 points — not because the forecast countries are systematically weaker, but because the rotation places no top-cluster country in the window. A reader comparing 2026–2030 forecast mean to 1958–2025 historical mean without accounting for rotation-cohort composition would mis-attribute a compositional effect to a performance effect.

Acceptance perspectives

Table 17 applies the weighted-acceptance formulas from Equations 4 and 5 to every row in the dataset, separating the EU-citizen and non-EU-citizen audiences. All emoji assignments follow the rule defined in Section 4.1 without exception, correcting the two 2023 edition violations (Spain Jul–Dec 2023 at acceptance 7.3; Malta Jul–Dec 2028 at 5.1). The EU–non-EU spread is widest for foreign-policy-heavy presidencies (EU citizens weight domestic policy higher; non-EU citizens weight foreign policy higher), and narrowest for presidencies dominated by administrative throughput (end-of-cycle terms such as Belgium 2024-H1).

Table 15: Historical KPI averages of all EU Council presidencies, 1958–2025

Country	Avg KPI	No. presidencies	Note
Sweden	7.5	3	Incl. actual assessment 2023
Finland	6.8	3	
Poland	6.8	2	Incl. actual assessment 2025
Denmark	6.5	7	Incl. actual assessment 2025
Spain	6.4	5	Incl. actual assessment 2023
Czechia	6.4	2	Incl. actual assessment 2022
Austria	6.3	3	
Germany	6.3	13	
Ireland	6.3	8	
Estonia	6.2	1	Tallinn Digital Summit 2017
Lithuania	6.2	2	Eastern Partnership 2013
Slovenia	6.1	2	Incl. Janša controversy 2021
Latvia	6.0	2	First presidency 2015
Malta	6.0	2	First presidency 2017
Netherlands	6.0	13	
Portugal	6.0	4	

Continued on next page

Table 15 continued from previous page

Country	Avg KPI	No. presidencies	Note
Luxembourg	5.9	12	
France	5.8	14	
Romania	5.8	1	Rule-of-law controversy 2019
Slovakia	5.8	2	First presidency 2016
United Kingdom	5.8	5	Not an EU member since 2020
Bulgaria	5.5	1	Western Balkans, Digital 2018
Cyprus	5.5	2	Financial crisis 2012
Belgium	5.4	14	Incl. actual assessment 2024
Italy	5.4	14	
Croatia	5.3	1	COVID pandemic 2020
Greece	5.1	6	
Hungary	4.8	2	Incl. actual assessment 2024

Note: Averages computed directly from `data/presidency_scores_1958–2025.csv` as the arithmetic mean of `kpi_overall` filtered on `country` with `data_type` $\in$ `{ai_estimate, actual_retroactive}` (forecast rows excluded). *Country-name normalisation:* CSV uses underscore form (e.g. `Czech_Republic`, `United_Kingdom`); this table displays the standard short form (Czechia, United Kingdom). All pre-2022 scores are AI-generated estimates supported by research into official assessments; 2022–2025 scores for completed presidencies are retroactive assessments. The United Kingdom is listed for historical completeness (5 presidencies 1977–2005) but is not an EU member since 2020. Ordered by average descending; ties ordered alphabetically.

Table 16: 2023-edition predictions vs. actual outcomes and new 2026-edition forecasts (2022–2030)

Period	Country	2023 KPI	2026 KPI	Δ	Data type	Key assumption / source
Jul–Dec 2022	Czechia	6.2	6.8	+0.6	actual	Energy Reg., Schengen expansion, Ukraine crisis leadership (Office of the Government of the Czech Republic 2022)
Jan–Jun 2023	Sweden	7.8	8.0	+0.2	actual	Fit for 55 finalised; historic migration pact (Government Offices of Sweden 2023)
Jul–Dec 2023	Spain	6.7	6.8	+0.1	actual	AI Act, electricity reform, 3 major summits (La Moncloa 2024)
Jan–Jun 2024	Belgium	6.1	6.0	−0.1	actual	End-of-cycle; EU elections constrained ambition
Jul–Dec 2024	Hungary	5.1	4.5	−0.6	actual	Institutional boycott; Orbán–Putin scandal (Euronews 2024)
Jan–Jun 2025	Poland*	—	7.0	—	actual	SAFE (€150bn), Russia sanctions 16+17, pharma reform (Polish Presidency of the Council of the EU 2025)
Jul–Dec 2025	Denmark*	—	6.8	—	actual	2040 climate target, Mercosur ratification, defence
Jan–Jun 2026	Cyprus	—	<i>5.8</i>	—	<i>forecast</i>	Assumes: stable regional security; migration + Eastern Med. dominant; 2nd presidency

Continued on next page

Table 16 continued

Period	Country	2023 KPI	2026 KPI	Δ	Data type	Key assumption / source
Jul–Dec 2026	Ireland	—	6.5	—	forecast	Assumes: pragmatic economic agenda; strong EU commitment post-Brexit
Jan–Jun 2027	Lithuania	—	6.2	—	forecast	Assumes: security/Russia agenda dominant; digital as secondary
Jul–Dec 2027	Greece	—	5.3	—	forecast	Assumes: migration controversy; Western Balkans; fiscal constraints
Jan–Jun 2028	Italy	—	5.5	—	forecast	Assumes: Southern flank focus; domestic instability risk
Jul–Dec 2028	Latvia	—	6.1	—	forecast	Assumes: Baltic security agenda; digital/energy; small capacity
Jan–Jun 2029	Luxembourg	—	6.0	—	forecast	Assumes: financial centre perspective; AI governance; inst. efficiency
Jul–Dec 2029	Netherlands	—	6.3	—	forecast	Assumes: pragmatic economic agenda; trade and agriculture
Jan–Jun 2030	Slovakia	—	5.8	—	forecast	Assumes: digital modernisation; domestic volatility risk

Jul–Dec 2030	Malta	—	5.9	—	forecast	Assumes: 2nd presidency; migration remains central
-----------------	-------	---	-----	---	----------	--

Note: Data types — **actual:** retroactive assessment based on documented sources; *forecast:* prediction with explicit assumptions. Forecast rows state assumptions in the form **Assumes: <precondition phrase>**; the phrase is a declarative precondition on the presidency term (agenda posture, political stability, external-shock status) whose factual violation during the term constitutes an assumption break per Appendix B verification protocol. $\Delta = 2026 \text{ KPI} - 2023 \text{ KPI}$ (actual terms only). (*) Poland and Denmark not predicted in the 2023 edition due to rotation error; retroactive assessment based on documented sources. Season labels: **SS** = Security Spring; **GF** = Geopolitical Flux.

Table 17: Citizen perspectives — acceptance KPIs (2026 edition), emoji-consistent

Period	Country	2026 KPI	EU Citizens		Non-EU Citizens	
			Emoji	Score	Emoji	Score
Jul–Dec 2022	Czechia	6.8	😞	6.5	😞	6.0
Jan–Jun 2023	Sweden	8.0	😄	8.3	😄	7.5
Jul–Dec 2023	Spain	6.8	😄	7.2	😞	6.3
Jan–Jun 2024	Belgium	6.0	😞	5.8	😞	5.9
Jul–Dec 2024	Hungary	4.5	😡	3.9	😞	4.1
Jan–Jun 2025	Poland	7.0	😄	7.3	😞	6.5
Jul–Dec 2025	Denmark	6.8	😄	7.1	😞	6.4
Jan–Jun 2026	Cyprus	5.8	😞	5.5	😞	5.4
Jul–Dec 2026	Ireland	6.5	😞	6.8	😞	6.1

Continued on next page

Table 17 continued

Period	Country	2026 KPI	EU Citizens		Non-EU Citizens	
			Emoji	Score	Emoji	Score
Jan–Jun 2027	Lithuania	6.2	😐	6.0	😐	5.8
Jul–Dec 2027	Greece	5.3	😐	5.1	😐	4.9
Jan–Jun 2028	Italy	5.5	😐	5.3	😐	5.2
Jul–Dec 2028	Latvia	6.1	😐	5.9	😐	5.7
Jan–Jun 2029	Luxembourg	6.0	😐	5.8	😐	5.6
Jul–Dec 2029	Netherlands	6.3	😐	6.5	😐	6.0
Jan–Jun 2030	Slovakia	5.8	😐	5.6	😐	5.4
Jul–Dec 2030	Malta	5.9	😐	5.7	😐	5.5

Emoji rule: ≥ 7.0 = positive (😊); 4.0–6.9 = neutral (😐); < 4.0 = negative (😞). This table corrects the emoji violations present in the 2023 edition (Spain Jul–Dec 2023; Malta Jul–Dec 2028). All assignments in this table follow the defined rule without exception.

Table 18: Thematic priorities and influencing factors — 2026 edition (2022–2030)

Period	Country	Key themes	Rationale / Assumptions
Jul–Dec 2022	Czechia	Energy, Ukraine, Schengen	Russian war dominated; crisis management exceeded expectations
Jan–Jun 2023	Sweden	Climate (Fit for 55), migration, Ukraine	Two generation projects finalised; strongest in observation period
Jul–Dec 2023	Spain	AI Act, electricity market, enlargement	High-profile diplomacy; tech and climate agenda

Continued on next page

Table 18 continued

Period	Country	Key themes	Rationale / Assumptions
Jan–Jun 2024	Belgium	Migration pact, Green Deal, EU elections	End-of-cycle dynamics; completing legislation
Jul–Dec 2024	Hungary	Competitiveness, defence	Budapest Declaration; Orbán–Putin scandal dominated
Jan–Jun 2025	Poland	Defence (SAFE), Russia sanctions, pharma	Security agenda dominant; historic EU defence investment
Jul–Dec 2025	Denmark	Climate (2040 target), Mercosur, defence	Balance of climate and security; trade policy successful
Jan–Jun 2026	Cyprus	Migration, energy, Eastern Med.	Assumes: 2nd presidency; regional context shapes agenda; ± 0.5 flux
Jul–Dec 2026	Ireland	Trade, digital, agriculture	Assumes: pragmatic economic agenda; strong EU commitment
Jan–Jun 2027	Lithuania	Russia/defence, Eastern Europe	Assumes: Baltic security dominant; digital as secondary
Jul–Dec 2027	Greece	Migration, Mediterranean, enlargement	Assumes: Western Balkans; fiscal constraints limit ambition
Jan–Jun 2028	Italy	Migration, industry, Mediterranean	Assumes: Southern flank; domestic instability risk
Jul–Dec 2028	Latvia	Defence, Russia, digital, energy	Assumes: Baltic security; small administrative capacity
Jan–Jun 2029	Luxembourg	Finance, justice, digital, AI gov.	Assumes: financial centre perspective; institutional efficiency
Jul–Dec 2029	Netherlands	Trade, agriculture, climate	Assumes: pragmatic economic agenda; CAP reform likely contentious
Jan–Jun 2030	Slovakia	Digital, energy, competitiveness	Assumes: economic modernisation; political volatility risk

Continued on next page

Table 18 continued

Period	Country	Key themes	Rationale / Assumptions
Jul–Dec 2030	Malta	Migration, Mediterranean, small states	Assumes: 2nd presidency; migration central

Completed periods (2022–2025): retrospectively documented. Forecast period (2026–2030): assumptions explicit and testable.

Layperson summary

This chapter holds the numbers. Historical averages show Nordic presidencies (Sweden, Finland, Denmark) consistently at the top of the KPI distribution; several Southern and smaller Central European presidencies sit lower. The 2023 predictions for 2022–2025 were broadly in the right direction — Sweden did well, Hungary struggled — but they compressed scores toward the mean, so the strongest and weakest terms were both understated. Looking ahead, ten presidencies are forecast for 2026–2030; every row states an assumption and names what would need to happen for the forecast to be proven wrong. None of the 2026–2030 presidencies is a Nordic country, which means the average forecast score is lower than the recent past — not because those countries are less capable, but because the rotation simply does not place a Nordic presidency in this window.

7 Discussion

7.1 Validation Results

The 2023 edition predictions for 2022–2024 show directional accuracy alongside a systematic tendency to compress scores toward the mean. Sweden slightly exceeded its prediction (actual ≈ 8.0 ; predicted 7.8), Spain was essentially on-target (+0.1), Belgium marginally under (−0.1), and Hungary was rated too generously (actual ≈ 4.5 ; predicted 5.1, delta −0.6). Czechia 2022 exceeded its prediction by +0.6, but this is better understood as a posterior update under the Bayesian black-swan framework (Section 4.6) than as a 2023 edition error: the Ukraine war produced a security-agenda premium that was not derivable from the country’s historical profile. The +0.6 is expected from the framework, not a failure of it.

The regression-to-the-mean pattern in the 2023 edition predictions is consistent with the empirical behaviour of language models in numerical estimation tasks: absent explicit calibration training, LLMs systematically under-predict extremes. In this article’s context the pattern is productive for the median of the distribution (Belgium, Spain) and problematic at the tails (Sweden’s ceiling, Hungary’s floor). The 2026 edition does not solve this directly — the 2026 edition forecasts are still LLM-produced estimates. What 2026 edition adds is the transparency that makes the pattern visible: a reader inspecting Table 16 can see that forecast spreads are narrower than they should be and compensate accordingly.

The wholly incorrect rotation prediction from January 2025 onward is not a quality problem of the KPI model itself but a data problem. The 2023 edition AI model lacked reliable access to the official rotation schedule and generated a structurally plausible but factually incorrect sequence (CFR claim C-002 in the project’s internal registry). With agentic research, this error would have been avoided at the point of the original 2023 edition draft. It is, in that sense, the cleanest case in this retrospective for the “workflow, not model” argument advanced in Section 5: neither 2023 edition nor 2026 edition differ in their ability to recall structural information from pre-training, but 2026 edition has a workflow step that retrieves the ground truth.

7.2 The three new analytical frameworks, in light of results

The political-season framework (Section 4.4) retroactively explains a cluster of apparent 2023 edition surprises. Sweden’s marginal overperformance fits Security Spring adjustment of +0.3 combined with Nordic archetype strength; Hungary’s underperformance fits the same season’s asymmetric application (favouring eastern and Nordic presidencies with strong mandates, disfavouring presidencies with rule-of-law controversies). The framework does not eliminate surprise; it structures it.

The expansion-phase detector (Section 4.5) receives modest but consistent empirical support from the retrospective. Czechia's 2022 presidency (second) scoring +0.8 above its 2009 first-term benchmark is in the predicted direction. The inferred range of +0.4 as a default second-presidency correction is therefore empirically anchored but not yet statistically established — a limitation acknowledged in the next subsection. Croatia 2020 (first; KPI ≈ 5.3) is consistent with the first-presidency penalty, though the COVID confounding variable limits its weight as an anchor.

The Bayesian black-swan framework (Section 4.6) performed as designed on the two live cases of the retrospective window. The Czechia +0.6 security premium and the Hungary -0.6 institutional boycott both fell within the pre-defined ranges for their respective event types, providing within-framework validation. The framework's claim is not that these deltas are predictable in magnitude but that they are structurally anticipatable in direction given the event type.

7.3 The Value of Methodological Transparency

The most important finding of this retrospective may not be what was right or wrong, but why. The honest documentation of the manual copy-paste workflow in 2023 edition makes it possible to understand precisely where errors arose and how they are addressed in 2026 edition. This transparency culture is itself a scientific contribution. Without the Making-of document (an unusual artefact in AI-assisted research), the systematic differences between intended and actual methodology would have been invisible to this retrospective analysis.

Three specific transparency practices proved essential. First, the original publication of the Making-of alongside 2023 edition established the evidentiary base. Second, the explicit data-type column (`ai_estimate`, `actual_retroactive`, `forecast`) in the master dataset allows readers to distinguish provisional estimates from source-verified outcomes on a row-by-row basis. Third, the claim-fact registry described in Section 5 makes every factual claim in 2026 edition auditable against its source. These three practices are cumulative: each adds a layer of auditability that the previous does not provide.

7.4 The Role of Agentic AI in Political Science Research

The transition from generative to agentic AI marks a methodologically significant step: AI systems can now not only synthesise but verify. For political-science practice, this means AI-generated forecasts can for the first time be systematically checked against reality and iteratively improved in a single pipeline. The present work understands itself as a pilot application of this principle.

A finding that should be stated without overclaim is that the *conceptual framework* (KPI model, five categories, acceptance scores, regional archetypes) proved sound across the 2023

edition to 2026 edition transition. What failed in 2023 edition was the *execution*: source verification, cross-session consistency, completeness of the historical dataset. Agentic AI addresses execution failures without requiring a new framework. This is a narrower claim than "agentic AI improves forecasting"; the evidence for that broader claim requires the 2028–2029 verification round.

A useful corollary for comparable projects: the five-tier taxonomy used here (Content / Standards / Verification / Operational / Data) is not EU-specific. The same structure would apply to an equivalent methodology paper in economic-policy forecasting, legal-document analysis, or cross-national regulatory comparison. What changes between domains is the content of the claim-fact registry, not its function.

7.5 Positioning against the post-Lisbon literature

Beyond Tallberg's brokerage-framing, Warntjen's efficiency-versus-burden analysis, Elgström's comparative perspective and Vanhoonacker's post-Lisbon institutional analysis cited in §3, a substantial post-2014 literature on presidency effectiveness under the Lisbon-era trio mechanism and the post-functionalist turn in EU governance would warrant explicit engagement in a future revision. The present article positions itself as methodology-first, with the KPI framework offered as one quantitative complement to that qualitative literature rather than a substitute for it; specific citations from that body of work are flagged as [W-pending] for the next research-curation round when live-web access is available, and the framing here should be read as provisional pending that addition.

7.6 Limitations

The limitations enumerated below are presented as *calibration targets for a future revision*, not as failures of the 2026 edition. Each names a parameter or scope boundary that is acknowledged in the current framework and is specifically reserved for future empirical refinement at the 2028–2029 fact-check round (see §A.3 verification framework). The ± 0.5 uncertainty band attached to every forecast row is itself an estimated structural prior tied to the `geopolitical_flux` season tag; recalibration of this band against observed forecast deviations is one of the explicit future-revision work items.

Taken together, the bullets below mean that the KPI core of this article is honestly a *disciplined forecasting framework with pre-registered accountability*, not an externally benchmarked political-science measurement instrument. The discipline is real (claim registry, volatile-facts register, pre-registered hypotheses, scheduled fact-check) but the quantitative layer is structured expert estimation. Readers should treat the forecast rows as conditional, auditable bets against the 2028–2029 round rather than as settled measurements.

Several limitations constrain the interpretation of the results:

- **The ± 0.5 uncertainty band attached to every 2026–2030 forecast is a structural prior, not a computed statistical confidence interval.** It is tied to the `geopolitical_flux` season tag (§4.4) and anchored on expert judgement about Trump-2.0 trade and defence uncertainty. Its empirical recalibration against observed 2026–2028 forecast deviations is a future-revision target.
- **KPI scores remain AI-generated estimates** without independent empirical validation. No external ground-truth dataset for Council presidency performance exists; where outcome reports were available (2022–2025), they were used as cross-references, not as replacement measurements.
- **Public opinion and media sentiment** are modelled only indirectly through the weighted-acceptance formula (Equations 4 and 5). Direct survey or social-media integration is outside the scope of the 2026 edition and is a candidate for a future-revision extension.
- **Geopolitical disruption** (Ukraine war from 2022, second Trump administration from 2025) cannot be anticipated from historical patterns alone. The Bayesian framework acknowledges this and provides a structured mechanism for posterior updates, but event-onset timing remains out-of-model.
- **The 2026–2030 forecasts** do not account for future government changes within the forecast window, which can materially alter a presidency’s agenda in the six months before term start. The ± 0.5 flux band is intended to absorb this but is not a formal confidence interval.
- **The Bayesian black-swan updates** are informed estimates, not statistically calibrated posterior distributions. A rigorous calibration would require empirical data on past shock effects across multiple political systems, which the current dataset does not contain. This is the most significant methodological weakness of the three new frameworks introduced in Section 4.
- **The expansion-phase penalty** of -0.7 and correction of $+0.4$ are anchored on Czech case evidence and extrapolated to comparable contexts. Empirical validation of these values across all first-presidency cases requires a dedicated regression analysis that is outside scope.
- **Third-country counterparty dynamics** are outside the framework’s input set. The state-ability matrix scores EU-internal handler capacity; the bilateral-tension registry is intra-EU. Third-country association files — the Canada–EU case in §9 (C9) is illustrative — can be scored on their EU handler side but not on their counterparty trajectory. Extension to third-country variables is reserved for a future revision.

Layperson summary

This chapter interprets what the results mean. A key finding: the 2023 version was directionally right but compressed extreme scores toward the average, so it over-scored weak terms and under-scored strong ones. The 2026 edition inherits this tendency and names it honestly. The chapter also applies the three new analytical frameworks retrospectively: Czechia's 2022 over-performance is consistent with a border-state security bonus during the Ukraine war; Hungary's 2024 under-performance is consistent with a rule-of-law event type that triggered institutional isolation. These are not excuses; they are structural explanations. For readers in the EU, the practical implication is that an AI-assisted forecast like this can be useful as a *starting brief* provided the reader keeps the compression-toward-the-mean tendency in mind when reading extreme cases.

8 Outcomes and Implications

This section translates the findings of the study into what they mean for specific audiences and for the quality of EU legislation. Section 7 interpreted the results within the framework of this paper; this section interprets them outward, toward the people and institutions whose work is shaped by presidency performance.

Four finding groups are covered: the 2023 edition validation pattern, the political-season framework, the expansion-phase detector, and the Bayesian-vocabulary black-swan framework. For each, four reader groups are addressed in turn: EU functionaries (staff of the Council, Parliament, Commission, and national permanent representations); EU citizens (residents of EU member states whose lives are shaped by Council output); non-EU citizens (residents of third states affected by the EU's external posture); and the legislative-quality question (how the finding helps make EU law better or makes its limits visible).

8.1 Implications of the 2023 edition validation pattern

The 2023 first version correctly ranked presidencies in directional terms but compressed scores toward the mean. Strong terms were under-predicted; weak terms were over-predicted.

For EU functionaries. When AI-assisted forecasts are used as input to internal briefings, the compression pattern matters: a presidency scored at 6.8 in an AI forecast may be genuinely stronger than that number suggests if the country profile is in the Nordic cluster, or genuinely weaker if it is in the Central European first- presidency cohort. Functionaries should read mid-range AI scores as centre-of-distribution estimates, not as point predictions.

For EU citizens. The practical consequence is that AI-based public commentary about "who will be the good presidency" is likely to be directionally useful but numerically understated in both directions. A presidency that ends up exceptional was probably predicted as just-above-average; a presidency that ends up weak was probably predicted as just-below-average.

For non-EU citizens. Trade-policy and foreign-policy expectations built around AI-assisted presidency forecasts should account for the compression pattern: the strongest negotiating partners (Nordic terms) will be underestimated in the forecast, the weakest partners will be overestimated.

For EU legislative quality. The pattern suggests that AI-assisted forecasting should be paired with expert human adjustment of extreme cases rather than used directly. A forecast that names its compression tendency explicitly (as this study does) is more useful than a polished forecast that presents extreme cases with false confidence.

8.2 Implications of the political-season framework

The framework classifies each six-month window as one of six seasons, each carrying a KPI adjustment. Current window is a dual state (Security Spring + Geopolitical Flux).

For EU functionaries. The season tag is a pre-term planning lens. Pre-term briefings that include the season classification give ministers context for why one presidency's agenda bandwidth is wider or narrower than another's. The asymmetric application of Security Spring (+0.3 for Eastern / Baltic / Nordic, 0 for Southern / Wave-4 Alpine, +0.1 for Core founding) also tells functionaries which presidencies stand to benefit from the current environment and which will need to work harder to meet their pre-term ambitions.

For EU citizens. The framework explains why recent presidencies have felt different from pre-2022 ones: security, defence, and borders have been systematically elevated on every agenda because the season itself prioritises them. This is not a choice of individual presidencies; it is the EU's current political climate.

For non-EU citizens. Trade partners and diplomatic counterparts should expect Council presidencies 2026–2028 to run trade-policy and defence-policy files in parallel. The Geopolitical Flux season tag signals that predictability on these files is structurally reduced relative to the pre-2022 period.

For EU legislative quality. A framework that makes seasonal-context explicit is a guard against treating every presidency as if it operated in the same environment. Council progress notes that reference season tags carry less implicit bias than ones that do not.

8.3 Implications of the expansion-phase detector

First-term presidencies score approximately -0.7 points below their regional archetype mean as a structural prior; second terms recover about $+0.4$ of that gap.

For EU functionaries. Specifically for functionaries in countries preparing their first Council presidency (notably candidate states Ukraine, Moldova, Western Balkans), the penalty is a structural expectation, not a criticism. Pre-term preparation should prioritise administrative handover from the outgoing presidency and dossier-prioritisation aligned with national competence, because the penalty compounds fastest on technically complex legacy files.

For EU citizens. Residents of a member state holding its first Council presidency should expect a less intense legislative impact from that term than from a veteran presidency — not because the national administration is weak, but because presidency is an institutional role whose full value accrues with repetition.

For non-EU citizens. Trade and diplomatic counterparts should expect first-presidency countries to be less active agenda-setters on external-policy files. Technical follow-through on existing files generally holds; ambitious new initiatives are rarer in first terms.

For EU legislative quality. Acknowledging the first- presidency penalty in pre-term planning improves outcomes: files that would otherwise have stalled because the incoming presidency lacks depth are instead carried forward by the trio or by the outgoing presidency's hand-over briefing. The framework makes the pattern visible so it can be planned around.

8.4 Implications of the Bayesian-vocabulary black-swan framework

The framework classifies external shocks into three event types and records their direction and magnitude of effect on presidency KPI (without claiming to compute a Bayesian posterior).

For EU functionaries. The three event types (External Security Shock, Domestic Rule-of-Law Collapse, Global Economic Shock) are a contingency-planning taxonomy. A functionary preparing an upcoming presidency can use the framework to pre-draft contingency files for plausibly exposed event types — accelerated emergency-legislation templates, Commission-coordination protocols, multi-capital diplomatic fallback plans.

For EU citizens. The framework explains why presidencies that perform unusually well or badly may do so because of events their country did not choose (a border war, a rule-of-law standoff, a global trade shock). Citizens assessing presidency performance should distinguish agency (what the country brought) from environment (what the event forced).

For non-EU citizens. External events can make a presidency more or less engaged on a third-state file. Non-EU partners should track the EU's event-type exposure rather than relying on the baseline country profile alone.

For EU legislative quality. The framework legitimises reporting the deviations from a baseline forecast rather than hiding them: if a presidency over-delivers because of a security premium, the Council-level progress note can name the premium rather than attribute the outperformance to agency alone.

8.5 Implications of the pre-registration commitment

Every 2026–2030 forecast in Table 22 is a named hypothesis with a stated assumption and a scheduled verification round for 2028–2029.

For EU functionaries. The pre-registered hypothesis set is a ready-made checklist for a post-term review. When the Greek 2027 presidency closes, H4 ($KPI \leq 5.5$) can be checked immediately; this makes presidency reviews cumulative across terms rather than one-off.

For EU citizens. The commitment to a 2028–2029 fact-check round means citizens can later hold this analysis accountable: either the forecasts were directionally right or they were not, and the update round will say so.

For non-EU citizens. A publicly auditable forecast set gives third-state analysts a framework they can compare against their own scenario plans, without having to trust the forecaster.

For EU legislative quality. Pre-registration is the single most transferable discipline in this study. Any AI-assisted research on EU governance that adopts it gains a mechanism for correcting itself that opinion-based commentary cannot match.

8.6 Summary: the study's legislative-quality contribution

The four findings, taken together, support a specific claim about EU legislative quality. Legislative quality depends on institutions knowing themselves: knowing which presidencies can realistically move which files; knowing which seasons amplify or suppress particular agendas; knowing which accession-cohort dynamics to plan around; knowing which shock types to pre-stage contingencies for. An AI-assisted analysis that names these patterns and publishes testable forecasts gives EU institutions a self-knowledge layer that opinion-based commentary does not. The analysis does not replace judgement by functionaries and legislators; it improves the baseline they judge against.

The study's outputs — the academic paper, the policy brief (Companion A), the procedural memo (Companion B), the civic explainer (Companion C), and the World Model Annex — are deliberately sized to different readers' roles. Each reader can pick the artefact that fits their job and still draw from the same underlying claim-fact registry. For readers in any

of the four groups addressed in this section, the practical implication is the same: better self-knowledge at institutional level, better-informed expectations at citizen level, and a more defensible basis for judging whether a given presidency's legislative contribution was better or worse than its structural prior.

8.7 Per-forecast-row audience cascade

The framework supports per-presidency audience-specific forecasts. The table below translates each of the ten 2026–2030 forecast rows into a one-sentence expectation per audience, drawing on the motion / policy forecast of §4.7 (Table 10) and the challenge analysis of §9.

Table 19: Per-forecast-row audience cascade for the 2026–2030 window. Each cell is a one-sentence expectation for the named audience under the named presidency.

Presidency	EU functionaries	EU citizens	Non-EU citizens (EU- resident)	Candidate / third states
CY 2026-H1	Pre-stage Eastern-Med migration and maritime files	Modest tightening on Eastern-Med border files	Stable	Türkiye: unfavourable; Ukraine/- Moldova: neutral
IE 2026-H2	Pre-stage trade- diversification and AI-Act enforcement files	Stable; modest digital-service regulation clarity	Stable; trade-openness bias positive	Third states: favourable for bilateral trade talks
LT 2027-H1	Pre-stage Belarus- sanctions, Eastern Partnership, defence	Stable living standards; security-file visibility higher	Stable	Ukraine / Moldova: accession advocacy window

Continued on next page

Table 19 continued

Presidency	EU func.	EU citizens	Non-EU res.	Candidate / 3rd
GR 2027-H2	Pre-stage migration and Western-Balkans accession; Türkiye friction	Migration visibility higher; fiscal constraints felt	Stable	Türkiye: unfavourable; Western Balkans: moderately favourable
IT 2028-H1	Pre-stage migration returns + industrial policy + SGP flexibility	Sectoral adjustment in energy-intensive industries; domestic political volatility	Stable; returns-directive pressure modest	Third states: mixed (industrial-policy protectionism vs. bilateral returns deals)
LV 2028-H2	Pre-stage Baltic defence + Russia sanctions	Stable; energy-security visibility higher	Stable	Ukraine: accession advocacy window
LU 2029-H1	Pre-stage AI governance + rule-of-law strictness files	Financial-services clarity improves; AI rights clarified	Rule-of-law enforcement tightens modestly	Third states: AI-governance interlocutors courted
NL 2029-H2	Pre-stage fiscal-discipline re-anchor + climate ambition + trade files	Fiscal-tightening signalling; climate acceleration in regulated sectors	Strict but predictable rule-of-law enforcement	Third states: favourable for bilateral trade talks

Continued on next page

Table 19 continued

Presidency	EU func.	EU citizens	Non-EU res.	Candidate / 3rd
SK 2030-H1	Pre-stage digital- modernisation + energy- security files	Sector modernisation; government- turnover risk in effect	Stable (some rule-of-law friction risk)	Candidate states: neutral
MT 2030-H2	Pre-stage Mediterranean migration + small-state framework files	Mediterranean migration visibility high	Stable	Third states (Mediterranean partners): favourable for technical- cooperation deals

The table is a compact per-row reading; fuller derivation per presidency is in §4.7 (Table 10) and in the challenge-by-challenge analysis of §9.

9 Current EU Challenges

The framework of §4.7 is general. This section applies it to the specific set of EU challenges open as of April 2026, one by one, with approach variants currently in active debate and predictions of which upcoming 2026–2030 presidency is most likely to push which approach. The chapter is intentionally grounded in the present moment; its contents are volatile and tracked in the Volatile Facts Register (§D.3, entries V-060 through V-067). Each subsection is compact; the full inventory with per-challenge derivation is in the project source document `docs/17_current_challenges.md`.

9.1 C1 Ukraine and Moldova accession pace

Question: at what pace should accession negotiations with Ukraine and Moldova advance, given the ongoing war and the absorption-capacity implications for existing member states?

Approach variants. (a) Accelerated timeline with derogations (advocates: PL Wave-8 champion, LT Eastern Partnership champion, SI Western Balkans parallel). (b) Merit-based conditionality (advocates: DE, NL, DK, AT). (c) Enlargement-blockade-for-leverage (HU, using accession pace as cross-domain leverage).

Most likely per presidency in window: CY 2026-H1 (b); IE 2026-H2 (b); LT 2027-H1 (a), strongly; GR 2027-H2 mixed (a) for Western Balkans and (b) for Eastern; IT 2028-H1 (b) on absorption grounds; LV 2028-H2 (a). Window-level majority: (b) conditionality, with LT and LV as the only (a) windows.

9.2 C2 SAFE defence instrument and NATO burden-sharing

Question: how should the EU-level SAFE defence financing be renewed post-2027, and how does it coordinate with NATO-level burden-sharing demands under Trump 2.0?

Approach variants. (a) Strategic autonomy with Buy-European rules (advocates: FR, BE, LU). (b) Transatlantic compatibility with dual-sourcing (advocates: PL, CZ, LT, LV, EE, DK, FI, SE, NL). (c) Common-ground hybrid.

Most likely in window: LT 2027-H1 and LV 2028-H2 push (b) strongly. France's next term is outside the 2026–2030 window, so (a) has no champion within the window. Default: (b) transatlantic compatibility prevails.

9.3 C3 Fit-for-55 review + post-2030 climate target

Question: how aggressive should the post-2030 climate target be, and how flexible the Fit-for-55 implementation pathway for energy-intensive sectors?

Approach variants. (a) Ambitious acceleration (DK, SE, FI, DE, NL — 90 %+ target for 2040). (b) Predictable glide path (IT, PL, CZ, FR periodically). (c) Resistance (HU, partly SK, RO).

Most likely in window: CY, IE, LT, GR, IT 2026–2028 are low-salience on climate; expect (b) glide-path absorption. LV 2028 (b). LU 2029 mixed. NL 2029 (a). Window-level: the advocacy for (a) acceleration is thin in 2026–2028 and firms up only from 2029 onward.

9.4 C4 Migration pact implementation + Mediterranean + returns

Question: how to implement the 2024 Migration and Asylum Pact, resolve returns-directive enforcement gaps, and handle Mediterranean pressure.

Approach variants. (a) Solidarity and relocation (ES, GR, IT, MT, CY, PT). (b) External-border-first (NL, DK, AT, BG). (c) Exclusion (HU).

Most likely in window: CY 2026-H1 (a) with Eastern-Mediterranean weight; GR 2027-H2 (a) strongly with Western-Balkans flow focus; IT 2028-H1 mixed (a)/(b) — solidarity demand paired with bilateral returns agreements; MT 2030-H2 (a). Window-level: (a) solidarity dominates, because four of the ten forecast presidencies are Southern-flank small states with solidarity bias.

9.5 C5 US tariff response and Mercosur ratification

Question: how should the EU respond to Trump-2.0 tariff actions, and should Mercosur be fully ratified despite member-state bottlenecks?

Approach variants. (a) Reciprocal counter-measures (FR, IT). (b) Resilient-liberal pivot — diversify toward Mercosur, Indo-Pacific, AfCFTA, preserve internal openness (DK, NL, SE, IE). (c) Negotiated settlement (DE, BE).

Most likely in window: IE 2026-H2 (b) strongly (Anglosphere bridge, trade-liberal); NL 2029-H2 (b) strongly; DK 2025-H2 already pushed (b) (Mercosur ratified); IT 2028-H1 (a) on industry grounds.

9.6 C6 AI Act enforcement + Digital Markets Act + platform governance

Question: how strictly to enforce AI-Act obligations on general-purpose-AI providers; how to handle DMA designations; data-sovereignty framework.

Approach variants. (a) Strict enforcement (DK, SE). (b) Innovation-balanced enforcement with sandboxes (IE, LU, FI, EE). (c) Minimal enforcement (HU).

Most likely in window: IE 2026-H2 (b) strongly; LU 2029-H1 (b). No strong (a) advocate has a term in 2026–2030.

9.7 C7 Article-7 procedures and rule-of-law conditionality

Question: how to advance Article-7 procedures (principally against HU) and how to apply conditionality-based MFF disbursements.

Approach variants. (a) Advance Article-7 and apply strict conditionality (NL, DK, SE, FI, BE, LU). (b) Political-settlement path (DE, FR periodically, IT). (c) Blockade defence (HU).

Most likely in window: no (a)-strong advocate has a 2026–2030 term until LU 2029-H1. Expect (b) political-settlement drift through 2026–2028, with a (a) strictness re-focus possible from LU 2029 and NL 2029.

9.8 C8 Next-Generation-EU post-review and MFF 2028–2034

Question: how to evaluate NGEU’s legacy and shape the 2028–2034 Multiannual Financial Framework.

Approach variants. (a) Expanded common borrowing (FR, IT). (b) Return to fiscal discipline (DE, NL, FI, SE). (c) Conditional continuation.

Most likely in window: CY 2026-H1 (c); IE 2026-H2 (c)/(b); LV 2028-H2 (b); LU 2029-H1 (b)/(c); NL 2029-H2 (b) strongly; SK 2030-H1 and MT 2030-H2 (c). Window-level: no (a) expanded-borrowing advocate has a term; direction is (b) discipline or (c) conditional.

9.9 C9 Canada–EU closer association [emergent during drafting, not pre-registered]

Unlike C1–C8, C9 emerged during the drafting of the 2026 edition rather than being selected from the paper’s pre-registered file inventory; it therefore carries no KPI score, and the framework’s coverage of this file is partial. This subsection is an explicit scope note.

Publicly-reported Canadian government interest in deeper EU economic association surfaced in Q1 2026 under the Carney administration (VFR V-023). The framework’s Meta-inference on the rotation window (next subsection) already identifies IE 2026-H2 and NL 2029-H2 as the structurally best-placed chairs for any third-state bilateral trade or association file in the 2026–2030 window; that handler-side inference applies directly to a Canada file *if tabled*, with LT 2027-H1 as a weaker-but-plausible backup carrier and southern chairs less natural fits.

Whether the file is tabled at all, whether EU-27 unanimity for a CETA-plus extension materialises, and the counterparty’s domestic-political trajectory are exogenous variables

not coded in the state-ability matrix (§4.7) or the bilateral-tension registry (intra-EU only). The framework makes no forecast on those dimensions; refusing to do so is not a failure of the framework but a deliberate scope boundary. Extension to third-country counterparty variables is a candidate scope refinement for a future revision.

9.10 Meta-inference on the 2026–2030 rotation window

Taken together, the 2026–2030 rotation window is structurally oriented toward *conservative / glide-path / transatlantic* approaches across nearly all open challenges. Accelerated (a) variants have advocates in window only for enlargement (LT, LV, SI), migration solidarity (CY, GR, IT, MT), and climate acceleration (NL late in the window). There is no (a) term in defence (strategic autonomy), economic (expanded borrowing), or rule-of-law strictness until 2029.

The meta-inference is decision-relevant for four audience groups:

Candidate states looking for accession acceleration should prioritise outreach to LT 2027-H1 and LV 2028-H2, which are the two windows where (a) has an advocate.

Third states seeking bilateral trade initiatives should target IE 2026-H2 and NL 2029-H2 — both (b) resilient-liberal-pivot champions in their terms.

EU citizens expecting climate-ambition delivery will find 2026–2028 a quieter window; pressure on energy-intensive sectors is likely to rise from LU 2029 or NL 2029.

Non-EU residents in the EU can expect stable rule-of-law and migration conditions in the CY / IE / LT / GR / IT terms; LU 2029 introduces a rule-of-law strictness re-focus that could tighten return-directive enforcement.

Layperson summary

This chapter walks through eight specific EU policy questions open right now (April 2026) and asks, for each, which of the upcoming ten Council presidencies (Cyprus through Malta, 2026–2030) is likely to push which approach. The short version: the coming four years are mostly conservative. Enlargement will advance on conditionality rather than acceleration except during Lithuania’s and Latvia’s terms; defence stays transatlantically-oriented; climate ambition picks up only late; migration stays on Southern-solidarity terms. This is a structural, not a moral, observation: it reflects whose turn it is rather than what the EU “should” do. For an ordinary reader, the chapter’s practical contribution is a mental map of when a given policy ask is most likely to get traction and when it will be quietly parked.

10 Conclusion

The 2026 edition establishes that the KPI model developed in the 2023 edition is fundamentally viable as a forecasting instrument. Relative performance differences between presidency countries were correctly anticipated for the 2022–2024 term sequence. The critical failures of 2023 edition — hallucinated rotation schedules from January 2025 onward, ten member states missing from the historical baseline, emoji-rule violations in the acceptance table, and cross-table inconsistency between *Tabelle 2* and *Tabelle 3* — were methodological rather than conceptual in nature, and each was addressable through the agentic workflow described in Section 5. The conceptual framework held; the execution workflow did not. Replacing the execution layer, without redesigning the conceptual model, is what 2026 edition accomplishes.

The three new analytical frameworks

Three contributions new to 2026 edition are intended to be applicable beyond this article. The *political-season mapping* (Section 4.4) assigns every presidency a season tag and a calibration adjustment; it formalises the intuition that a presidency held during a crisis window has a different KPI expectation than one held in a quieter integration phase. The *expansion-phase detector* (Section 4.5) quantifies the first-presidency learning penalty at approximately -0.7 KPI points below the regional archetype mean, with a second-presidency correction of approximately $+0.4$; the Czech case (2009 first presidency KPI ≈ 6.0 ; 2022 second presidency KPI ≈ 6.8) provides the clearest empirical anchor. The *Bayesian black-swan framework* (Section 4.6) treats exogenous shocks not as model failures but as posterior updates with documented KPI impact ranges; the Czech 2022 ($+0.6$ security premium) and Hungarian 2024 (-0.6 institutional boycott) cases anchor the Eastern European spread.

Each of these frameworks is intended to be usable on its own. A reader interested only in seasons can apply the tag scheme without adopting the Bayesian-vocabulary framework; a reader interested only in learning-curve effects can apply the first-presidency penalty to any new accession cohort.

Practitioner takeaways

For a reader in a Working Party, permanent representation, or national EU-affairs ministry preparing a specific upcoming presidency, the three frameworks chain into a compact three-step pre-term procedure: (1) assign the term its political-season tag and apply the corresponding adjustment (§4.4); (2) apply the first-presidency learning-curve penalty or second-presidency correction to the country-level prior (§4.5); (3) identify the event types the term is plausibly exposed to and prepare a contingency file per type (§4.6). The output is an expected KPI band and a contingency catalogue, not a single forecast point.

The eleven pre-registered hypotheses in Table 22 show the procedure applied to every presidency in the 2026–2030 rotation window; a practitioner preparing a specific term can treat the relevant row as a starting brief and extend it against more recent information in the months before term start.

Pre-registration and the 2028–2029 verification commitment

The 2026 edition forecasts for 2026–2030 are recorded as eleven named hypotheses (Table 22 and Appendix B). Each hypothesis names the presidency, a falsifiable threshold, a stated assumption, and a specific verification source. A scheduled fact-check round in 2028–2029 will apply the methodology used in this document to validate 2023 edition against 2022–2025 actual outcomes (cf. `artifacts/FACT.CHECK.v1_outcomes.md`). The author commits to conducting this round and to publishing its results under the same transparency standard used for 2023 edition. A v3.0 revision is therefore structurally anticipated; the 2026 edition methodology treats that future revision as a testable property of the current article, not as a later afterthought.

The value of this pre-registration is that it constrains the author. Forecasts that cannot be checked are rhetorical; forecasts with stated assumptions, thresholds, and verification sources can fail publicly. The failure case is itself a contribution — it would tell us what aspect of the KPI model (regional archetype, seasonal adjustment, Bayesian update mechanism) requires revision.

What this work demonstrates about agentic AI in research

For the AI-in-research audience, the substantive finding is methodological. An early-stage failure mode of generative AI (structured hallucination of rotation schedules, paste-workflow inconsistency) maps directly onto a later-stage workflow remediation (world-model infrastructure, claim-fact registry, embedded generation prompts, pre-registered hypotheses). The remediation is not about better models. It is about workflow discipline (Mumin 2026). The same discipline that caught 2023 edition’s rotation error in this retrospective would catch a comparable error in any domain — economic forecasting, legal-document analysis, longitudinal literature review — provided the world model is built before the prose is generated and the claim-fact registry is maintained as the article grows.

The 2026 edition is not a claim that agentic AI has solved political-science forecasting. It is a demonstration that the workflow problems of early AI-assisted research are tractable, and that their solutions are reproducible. A further fact-check round in 2028–2029 will determine whether the underlying conceptual model is also sound. Either outcome advances the method. The combination of a transparent first version, a documented Making-of, a validated retrospective, and a 2026 edition with testable hypotheses makes

this contribution auditable — a property that is still uncommon in AI-assisted academic work.

Candidate hypotheses for the 2028–2029 round (H12–H17 and beyond)

During the 2026 edition iteration cycle, six additional candidate hypotheses [**candidate 2028–2029**] were contributed — three deduced from the paper’s own framework rules (D1 season-tag dominance under Geopolitical Flux; D2 expansion-phase handover-smoothness rise in the 2026–2030 window; D3 positive black-swan widens upper band without flipping the season tag), and three inducted from an insight about Hungary’s potential political normalisation combined with sustained external pressure (I1 rule-of-law friction drop raises the average KPI from 5.9 to 6.2–6.4; I2 “defensive deepening” pattern; I3 new EU integration rhythm of reactive bursts rather than steady grand plans). These are recorded as candidate H12–H17 in the project source document `docs/20_v2_1.candidates.md` with the same assumption / verification-method discipline as H1–H11. They are *not* part of 2026 edition’s pre-registered set; they are candidates for adoption by either a future revision of this paper or the 2028–2029 fact-check round itself. Five big-picture conclusions derived from H12–H17 under a Hungary-normalisation scenario are also captured in that document for the future round’s use.

Beyond H12–H17, an additional fourteen cross-framework candidate deductions (H21–H34) and two meta-conclusions (M1 layered convergence under persistent flux; M2 hysteresis under sequenced inputs) were produced by applying the Mumin Framework Synthesis Method (MFSM, see §5 and `docs/21_abstracted_method.md`) to alternative 2026–2030 scenarios (Russia-ceasefire 2027; composite inputs; time-ordered crisis-and-resolution sequences). Like H12–H17, these are [**candidate 2028–2029**] extensions, not 2026-edition claims; they are registered for the future round’s use in the project source document `docs/20_v2_1.candidates.md`.

Layperson summary

This chapter closes the study. The KPI model from 2023 proved fundamentally sound; what failed was the execution (unverified sources, missing countries, internal inconsistencies). The 2026 edition corrects execution without changing the conceptual model. The three new frameworks — political seasons, first-presidency learning curve, shock-type adjustments — extend the model to capture context. Every 2026–2030 forecast is published as a testable hypothesis with a fact-check round scheduled for 2028–2029. That means when the forecasts are wrong, we will know, publicly, on schedule. For readers in the EU, that commitment is the practical difference from commentary: this analysis signs up to be checked.

Data and Code Availability

This paper references several project-source artefacts by path — the master dataset `data/presidency_scores_1958–2025.csv`, the world-model documents under `docs/`, the per-iteration persona logs under `docs/scores/`, and the frozen 2028–2029 handoff package under `artifacts/handoff_package_2028/`. A reader consulting only the PDF does not need access to these files: the claim-fact registry (Appendix D.1), the volatile-facts register (Appendix D.3), the verification framework (Appendix B), the 32-persona roster (Appendix D.8), and the portability template (Appendix G) are all reproduced in the annexes at a depth sufficient for claim-level inspection.

For readers who want the underlying source distribution — the CSV, the full `docs/` directory, the build scripts, the persona logs in raw form, the decision log, and the sceptical-review / release audit documents referenced in-text — the materials are available from the author on request. Contact: eu@roblemumin.com (Volatile Fact V-040, verified April 2026). The intent is to publish the source distribution as a public archive (GitHub repository with a Zenodo-minted DOI for citation stability) before the 2028–2029 fact-check round opens; when that archive exists, a pointer will appear in a future revision of this paper and in the `consilium2016`-adjacent entries of the bibliography. Until then, by-request distribution is the canonical channel.

Handoff package. A separate frozen package (`artifacts/handoff_package_2028/`) contains the v1.0.1 PDFs plus a `world_model/` snapshot, the master CSV, and the persona logs at the state corresponding to this release. It is intended as the single entry point for the 2028–2029 fact-check author, whether that author is the original author or an equivalent agentic-workflow successor. The package includes a `HANDOFF_README.md` with the six-step verification roadmap (re-establish framework state; verify pre-registered hypotheses H1–H11; optionally adopt candidate hypotheses H12–H17 and H21–H34; apply the Mumin Framework Synthesis Method to 2026–2028 events; run the BB n-tier-deduction pass; decide on a v2.1 publication modality).

Licence. Prose and data in the source distribution are released under CC-BY-4.0 unless otherwise marked; build scripts and auxiliary tooling are released under the MIT licence. These terms apply to the by-request distribution and will carry over to the public archive when minted.

Reproducibility scope. The quantitative tables in this paper round-trip to the master CSV per the verification checklist in `docs/10_verification_checklist.md` (summary in §D.1); a reader reconstructing the tables from the CSV plus the definitions in Appendix A should recover every tabular figure to the last decimal. The forecast rows do not round-trip

to measurements; they round-trip to the pre-registered hypotheses H1–H11 (Table 22) and to the season-tag / expansion-phase / black-swan adjustments documented in §4.4, §4.5, and §4.6.

References

- Bengtsson, Rikard and Ole Elgström (2012). “Conflicting role conceptions? The European Union in global politics”. In: *Foreign Policy Analysis* 8.1, pp. 93–108. DOI: [10.1111/j.1743-8594.2011.00156.x](https://doi.org/10.1111/j.1743-8594.2011.00156.x).
- Elgström, Ole (2003). *European Union Council Presidencies: A Comparative Perspective*. London: Routledge.
- Euronews (2024). *European Commission will boycott Hungarian presidency over Orbán’s trips to Moscow and Beijing*. <https://www.euronews.com/my-europe/2024/07/15/european-commission-will-boycott-hungarian-presidency-over-orbans-trips-to-moscow-and-beij>. Accessed April 2026.
- Government Offices of Sweden (2023). *Historic achievements under the Swedish Presidency of the Council of the European Union*. <https://www.government.se/press-releases/2023/06/historic-achievements-under-the-swedish-presidency-of-the-council-of-the-european-union/>. Accessed April 2026.
- König, Thomas and Simon Hug (2000). “Ratifying Maastricht: parliamentary votes on international treaties and theories of political economy in the European Union”. In: *European Union Politics* 1.1, pp. 93–124. DOI: [10.1177/1465116500001001005](https://doi.org/10.1177/1465116500001001005).
- La Moncloa (2024). *Spanish Presidency of the Council of the European Union: results and achievements*. <https://www.lamoncloa.gob.es/lang/en/gobierno/news/paginas/2024/achievements-spanish-presidency-eu-council.aspx>. Accessed April 2026.
- Mumin, Roble (2026). *Agentic Book Writing: The Smart Way. A World-Model-First Methodology for AI-Assisted Long-Form Writing*. First. Version 1.0.15, February 2026. Reference edition: portable PDF, 568 pages. Source of the Topic-flavor world-model methodology applied in the present article. Self-published.
- Mumin, Roble, Claude 2 (Anthropic), and GPT-3.5 (OpenAI) (2023). *Leistungsanalyse und Erfolgsprognose von EU-Ratspräsidentschaften mit Hilfe von generativer KI*. Published article, Version 1.0, 14 September 2023.
- Nowak, Agnieszka K. (2010). “The EU Council Presidencies: Institutional Change and Preference Accommodation”. PhD thesis. Florence: European University Institute.
- Office of the Government of the Czech Republic (2022). *Czech Presidency of the Council of the EU ends: government presents achievements*. <https://vlada.gov.cz/en/media-centrum/aktualne/czech-presidency-of-the-council-of-the-eu-ends--government-presents-achievements-203545/>. Accessed April 2026.
- Polish Presidency of the Council of the EU (2025). *End of the Polish Presidency: ‘Security, Europe’ firmly anchored in the EU Agenda*. <https://polish-presidency.consilium.europa.eu/en/programme/achievements-of-the-polish-presidency/>. Accessed April 2026.

- Schalk, Jelmer, René Torenvlied, Jeroen Weesie, and Frans Stokman (2007). “The power of the presidency in EU Council decision-making”. In: *European Union Politics* 8.2, pp. 229–250. DOI: [10.1177/1465116507076431](https://doi.org/10.1177/1465116507076431).
- Smith, Michael (2000). “Negotiating new Europes: the roles of the presidency of the Council of the European Union”. In: *Journal of European Public Policy* 7.1, pp. 26–47. DOI: [10.1080/135017600343304](https://doi.org/10.1080/135017600343304).
- Tallberg, Jonas (2004). “The power of the presidency: brokerage, efficiency and distribution in EU negotiations”. In: *Journal of Common Market Studies* 42.5, pp. 999–1022. DOI: [10.1111/j.0021-9886.2004.00538.x](https://doi.org/10.1111/j.0021-9886.2004.00538.x).
- Vanhoonacker, Sophie and Nathalie Res (2010). “The Presidency as an institution: Changes in the post-Lisbon era”. In: *Journal of European Public Policy* 17.6, pp. 936–953. DOI: [10.1080/13501763.2010.499037](https://doi.org/10.1080/13501763.2010.499037).
- Warntjen, Andreas (2008). “The Council Presidency: Power broker or burden? An empirical analysis”. In: *European Union Politics* 9.3, pp. 315–338. DOI: [10.1177/1465116508093492](https://doi.org/10.1177/1465116508093492).

Annex Contents

A	KPI System (retained from the 2023 edition)	97
B	Verification Framework 2028–2029	98
C	Verification Framework Table	104
D	World Model Annex	106
E	Persona and Verification Approach	120
F	Companion Materials	127
G	Portability Template	135

A KPI System (retained from the 2023 edition)

Five Categories

- K_1 Domestic Policy: migration, justice, administration, social affairs
- K_2 Economic Policy: growth, unemployment, competition, budget
- K_3 Foreign Policy: relations with third states and regions
- K_4 Structural Policy: cohesion policy, agricultural policy, transport
- K_5 Institutional Affairs: efficiency, transparency, legislative output

Scoring: 1–10 per category; Overall KPI = equal-weighted average (Equation 2).

Citizen perspective emoji scale (strictly applied in Table 17):

😊 (positive) ≥ 7.0 😐 (neutral) 4.0–6.9 😞 (negative) < 4.0

Acceptance scores use the weighted formula from Equations 4 and 5, with weights from the table in Section 4.1.

B Verification Framework 2028–2029

To enable systematic verification of 2026 edition predictions, every forecast row (2026–2030) is recorded here as an explicit, falsifiable hypothesis. The full hypothesis table with thresholds, assumptions, and verification sources appears as Table 22. Canon rule T1-05 requires every forecast row to be pre-registered as a named hypothesis before publication; the ten presidency-level hypotheses H1–H10 and the model-level hypothesis H11 satisfy that requirement. For cross-reference parity with the Volatile Facts Register, the mapping $H1 \equiv V-030$, $H2 \equiv V-031$, $\dots$, $H10 \equiv V-039$ holds row-for-row (H11 has no V-row because the model-level claim is not a volatile fact but a methodological property); subsequent references use H-labels as the primary identifier.

Each hypothesis follows a common template. The *presidency* column identifies the six-month period. The *threshold* is a falsifiable condition on the retroactive KPI assessment — an inequality ($KPI \leq$ or $\geq$ a named value) rather than a point prediction, because the underlying scores are structured AI estimates with a documented ± 0.5 uncertainty band from the `geopolitical_flux` season tag. The *stated assumption* column names the premises under which the threshold is expected to hold; a black-swan event that invalidates the assumption constitutes an explanatory variable, not a falsification. The *how to verify* column names the primary document a future fact-checker should consult; where possible, this is an official Council closing report supplemented by at least one independent think-tank assessment.

The hypotheses are ordered chronologically by presidency period. Odd-numbered presidencies (H1, H3, H5, H7, H9) cover the first half of each year; even-numbered (H2, H4, H6, H8, H10) cover the second half. The model-level hypothesis H11 commits to directional accuracy: at least 7 of the 10 forecast presidencies ranked in the correct relative order by 2029. This mirrors the methodology used in this paper’s retrospective validation of 2023 edition against 2022–2025 actual outcomes (cf. `artifacts/FACT_CHECK_v1_outcomes.md`).

Verification protocol. For operational clarity to a future fact-checker (not necessarily the original author), the procedure is: (a) a forecast row *passes* when the retrospective-actual overall KPI falls within the ± 0.5 band centred on the pre-registered point estimate; (b) the stated *assumption* is declared broken when the named premise is factually violated during the term (e.g., an assumption “Ukraine war tempo stable” fails if a negotiated ceasefire or Article-5 event occurs in-term); (c) in the case of an assumption break, the row is scored *not applicable* rather than pass/fail, and the explanatory variable is recorded; (d) the adjudicator is the 2028–2029 fact-check author, pre-registered as the same author or an equivalent agentic-workflow successor. Ties (KPI at band boundary) resolve toward *pass* given the structural nature of the band.

Rotation verification: the fact-checker must additionally confirm at time of review that the rotation Poland → Denmark → Cyprus → Ireland → Lithuania → Greece → Italy → Latvia → Luxembourg → Netherlands → Slovakia → Malta is still current on the Consilium website. A rotation amendment between now and 2029 would require re-indexing the hypotheses; the rotation was verified against Council Decision 2016/1316 in April 2026.

Candidate hypotheses H12–H17 (for future revision / 2028–2029 round)

The pre-registered 2026-edition commitments are H1–H11 only; all H12+ identifiers are candidate hypotheses for a future revision or the 2028–2029 round.

Six candidate hypotheses contributed during the 2026 edition iteration cycle extend the framework and are registered for adoption either in a future revision or directly in the 2028–2029 fact-check round. They are not part of 2026 edition’s pre-registered set (H1–H11 above) but are reproduced here in summary form at the request of T Annex Reproducibility Expert to keep the candidate set inspectable for a reader without access to the project source directory. Full text, assumptions, and verification methods are in `docs/20_v2_1_candidates.md`.

Table 20: Candidate hypotheses H12–H17 for future revision / 2028–2029 round

Hyp.	Short form	Prediction (summary)	Verification method
H12	D1 Season-Tag	Under Geopolitical Flux + Security Spring, defence / borders / crisis files move fastest; internal-reform files stay deprioritised.	Count new SAFE / Schengen closures vs internal-market closures under Cyprus–Ireland terms.
H13	D2 Expansion-Phase	Almost no first-term presidencies in 2026–2030 ⇒ average KPI rises by +0.2 to +0.4 above the 5.9 baseline.	Compare handover-delays in 2026–2030 vs 2022–2025.

H14	D3 Black-Swan	A positive black-swan update (e.g., normalised HU voting) widens the upper KPI band by +0.3 to +0.5 but does not flip the season tag from Geopolitical Flux to Integration Spring.	Check 2026–2030 average KPI against 6.4 threshold.
H15	I1 HU-Normalisation	If HU’s political change removes the main rule-of-law veto player by end-2025, overall 2026–2030 average KPI rises from 5.9 to 6.2–6.4.	Track Article 7 mentions and Commission–Council coordination speed under CY / IE terms.
H16	I2 Defensive-Deepening	External pressure + fewer internal blockages ⇒ ”defensive deepening”: faster cooperation on security/resilience but no big leap in classic supranational integration.	Count new legislation closed 2026–2030 that is security/trade/resilience-related vs internal-market-deepening; threshold > 60% for confirmation.
H17	I3 Reactive-Burst	A new EU integration rhythm: progress in reactive bursts after shocks, not steady grand plans. Persists beyond 2030 unless the season tag shifts to Integration Spring.	Compare 2026–2030 legislative output speed against 2014–2022 baseline; count shock-justified vs ambition-justified major files.

Adoption protocol. H12–H17 are not treated as pre-registered claims of 2026 edition. They are captured here to prevent loss in the iteration churn and to give the 2028–2029 fact-check author a ready-made extension set. The fact-check round’s author should either (a) adopt H12–H17 alongside H1–H11 with an explicit note on the capture-date distinction,

or (b) defer H12–H17 to a future-revision paper update and keep the 2028–2029 round narrowly on H1–H11.

BB-produced n-tier candidate hypotheses H21–H34 + meta-conclusions M1, M2

A further fourteen candidate hypotheses (H21–H34) and two meta-conclusions (M1, M2) were produced by the BB n-Tier Deduction Synthesist persona applying the Mumin Framework Synthesis Method (MFSM, §5.7) to alternative scenarios beyond the Hungary-normalisation insight that produced H12–H17. These cover cross-row deductions (H21–H23, M1), five fresh news-input deductions (H24–H28), and four MFSM-test-derived deductions (H31–H34, M2 from the composition / contradiction / time-ordered sequence tests). Slots H29 and H30 are reserved.

Table 21: BB-produced n-tier candidate deductions H21–H34 + meta-conclusions M1, M2

ID	Short form	Prediction (summary)
H21 ²	Season × expansion-phase interaction	First-presidency penalty -0.7 is partially offset by season alignment when the dominant season favours the country's lead-topic strong suit.
H22	Bilateral-tension × season permissiveness	Active bilateral-tension flag-count drag depends on season permissiveness; restrictive seasons amplify blockade exposure.
H23	Convex response to external-shock intensity	Eastern / Baltic / Nordic premium is non-linear in shock intensity (rises convexly with severity).
M1	Layered convergence under persistent flux	Framework-layer interactions (not single-layer effects) drive forecast accuracy under sustained Geopolitical Flux.
H24	DE coalition shift 2026-H2 reverses positions	A CDU/CSU-led German government in 2026-H2 reverses Berlin's stance on three EU files (NGEU successor, CBAM, ESM).
H25	IT presidency under €-debt stress	IT 2028-H1 KPI band drops $5.5 \rightarrow 4.8\text{--}5.2$ if Italian sovereign spreads widen ≥ 150 bps in 2026–2027.

²BB-N indicates the N-th candidate produced by the BB n-Tier Deduction Synthesist persona (§5.7, §A.3).

Table 21 continued

ID	Short form	Prediction (summary)
H26	Türkiye realignment 2028 reverses CY blockade	TR-EU pivot in 2028 downgrades CY–TR bilateral-tension flag, lifting CY 2026-H1 KPI by +0.2.
H27	Iran escalation activates Mediterranean defence files	Iran-regional escalation in late 2026 lifts GR 2027-H2 by +0.2 via crisis-management premium.
H28	Digital-euro launch under LT 2027-H1	Launch raises LT K_5 by +0.4, lifting LT 2027-H1 KPI from 6.2 to ~ 6.4 .
H31	Composition asymmetry deduction	Composite (Russia-ceasefire + HU-normalisation) inputs produce $\sim 30\%$ predicted variance reduction across rotation-window KPI distribution.
H32	NATO-crisis defence-priority displacement	NATO Article-5 crisis 2026-H2 displaces internal-reform priority for ≥ 2 consecutive presidencies.
H33	Hysteresis stabilisation	Post-NATO-crisis ceasefire stabilises Eastern / Baltic / Nordic premia at 2026-edition baseline rather than depressing further.
H34	Ukraine-accession KPI lift	UA accession 2028-H1 lifts IT 2028-H1 and LV 2028-H2 by +0.2 each via UA-blockade removal.
M2	Hysteresis under sequenced inputs	Same final-state X -set produces different per-row predictions depending on order of arrival (~ 0.1 – 0.3 KPI per row).

Adoption protocol (H21–H34). Same as H12–H17. These are [candidate 2028–2029] hypotheses produced by the MFSM synthesis method, registered in `docs/20_v2.1_candidates.md` with full pre-registered-format detail (assumption / verification / framework provenance / supporting per-row outputs). They extend the 2026-edition framework without becoming 2026-edition claims.

MFSM output class				
	Single-row deduction	Cross-row deduction	Meta-conclusion	Inapplicability flag
Event-type update	H18, H19 H27, H31	H21, H22, H23	M1	(MFSM Test 5)
Bias-matrix modification	H15, H24, H29 H30, H36	H31	—	—
Bilateral-tension flag	H20, H26, H35	H22	—	—
New challenge entry	H25, H28, H37 H38, H40	—	—	—
Season re-tag	H16, H17, H34	H32, H33, H39	M2	—

Figure 9: Candidate-hypothesis taxonomy matrix. Rows: the five MFSM input classes (§5.7). Columns: the four output- pattern classes the MFSM produces (single-row per-presidency deductions; cross-row pattern deductions; meta-conclusions aggregating across multiple deductions; inapplicability flags signalling that the input lies outside the framework’s current scope). Cells list the H-identifiers from Tables 20 and 21 that correspond to each input $\times$ output pair. The empty lower-right cells indicate the framework has not yet produced meta-conclusions or inapplicability- flags via the bias / tension / challenge / season-re-tag input routes — a scope indicator for future BB passes. The 2028–2029 fact-check author can use this matrix to group hypotheses of the same pattern type when arbitrating verification thresholds.

C Verification Framework Table

Table 22 below lists the eleven pre-registered hypotheses H1–H11 — the 2026-edition frozen forecast commitments for the 2028–2029 fact-check round. The candidate-extension sets H12–H17 (Table 20) and H21–H34 (Table 21) above the BB-produced-candidates taxonomy matrix (Figure 9) sit alongside this frozen set but do not constitute pre-registered claims.

Table 22: Verification framework 2028–2029: pre-registered testable hypotheses for every forecast row

Hyp.	Presidency	Threshold	Stated assumption	How to verify
H1	Cyprus (Jan–Jun 2026)	KPI ≤ 6.5	Small-island archetype; 2nd presidency; migration + Eastern Mediterranean dominant; no new Cyprus-problem escalation; SAFE defence file continues	Official closing report + think-tank evaluation (ECFR, Elcano); ± 0.5 flux band
H2	Ireland (Jul–Dec 2026)	KPI ≥ 6.0	Stable government; pro-European orientation; trade-strong profile; pragmatic digital/agriculture agenda; no black-swan disruption	Official closing report + EC evaluation
H3	Lithuania (Jan–Jun 2027)	KPI ≥ 5.8	Baltic security + Russia agenda dominant; digital as secondary; Trump 2.0 trade friction persists	Official closing report + Eastern-Partnership tracker
H4	Greece (Jul–Dec 2027)	KPI ≤ 5.5	Migration controversy; limited agenda-setting capacity; fiscal constraints; Western Balkans enlargement file central	Official assessment + European Parliament evaluation
H5	Italy (Jan–Jun 2028)	KPI ≤ 5.5	Structural domestic political instability; Southern-flank focus limits EU-wide impact; Mediterranean and migration central	Think-tank assessments (ECFR, Bruegel, IAI)
H6	Latvia (Jul–Dec 2028)	KPI ≥ 5.8	Baltic security; digital/energy; second-presidency correction (+0.4) partially offsets small administrative capacity	Official closing report + Baltic think-tank review
H7	Luxembourg (Jan–Jun 2029)	KPI ≥ 5.8	Financial-centre perspective; AI governance file; institutional efficiency; small state but high administrative depth	Official closing report + finance/AI policy trackers
H8	Netherlands (Jul–Dec 2029)	KPI ≥ 6.0	Pragmatic-constructive diplomatic style; strong institutions; pro-European orientation; CAP reform likely contentious	Official assessment + EC evaluation
H9	Slovakia (Jan–Jun 2030)	KPI ≤ 6.2	Digital and energy modernisation agenda; domestic political volatility limits upside	Official closing report + V4 policy tracker
H10	Malta (Jul–Dec 2030)	KPI ≤ 6.2	Second presidency; migration central; small-island archetype with limited EU-wide agenda reach	Official closing report + Mediterranean-migration tracker
H11	Model accuracy (overall)	≥ 7 of 10 correct	The 2026 edition correctly ranks ≥ 7 of the 10 forecast presidencies available for review by 2029 (directional accuracy, not absolute KPI)	Apply methodology from <code>artifacts/FACT_CHECK_v1_outcomes.md</code>

Rotation verification: fact-checker must confirm at time of review that the rotation Poland → Denmark → Cyprus → Ireland → Lithuania → Greece → Italy → Latvia → Luxembourg → Netherlands → Slovakia → Malta is still correct using the then-current Consilium website. This replicates the 2023-edition rotation validation conducted in April 2026. All $\pm$ bands reflect the geopolitical flux season uncertainty documented in Section 4.4.

D World Model Annex

This annex reproduces the core of the project’s world model — the structured reference documents that underlie every claim in this paper — so that a reader without access to the project source directory can inspect the claim-level, canon-level, and glossary-level artefacts directly. The full world model (twelve documents plus operational files) is available with the source distribution under `docs/`; the subsections below summarise each document at a depth sufficient for claim-level verification.

D.1 Claim-Fact Registry (CFR, summary form)

Every factual claim in the paper that can be wrong is registered in the CFR with an identifier, a one-line claim text, a source, and a tier. The summary below lists the identifiers referenced in the paper. Tier 1 = immutable (rotation/legal facts); Tier 2 = historical KPI baseline or validation; Tier 3 = narrative framing or bibliographic claim.

Table 23: Claim-Fact Registry: claims referenced in this paper

ID	Claim (one-line)	Source	Tier
C-001	EU Council rotation 2022–2030: CZ, SE, ES, BE, HU, PL, DK, CY, IE, LT, GR, IT, LV, LU, NL, SK, MT	Council Decision 2016/1316; consilium.europa.eu (verified April 2026)	1
C-002	2023 edition rotation from January 2025 diverged from actual (DE instead of PL, SI instead of DK, FI instead of CY)	2023 edition as archived; C-001 comparison	1
C-010	Czech presidency 2022-H2: 2023 edition predicted KPI 6.2; actual 6.8 (delta +0.6)	czech2022; naumann2023; FACT_CHECK_v1_outcomes.md	2
C-011	Swedish presidency 2023-H1: 2023 edition predicted 7.8; actual 8.0	sweden2023; eeb2023	2
C-012	Spanish presidency 2023-H2: 2023 edition predicted 6.7; actual 6.8	spain2024	2
C-013	Belgian presidency 2024-H1: 2023 edition predicted 6.1; actual 6.0	FACT_CHECK_v1_outcomes.md	2
C-014	Hungarian presidency 2024-H2: 2023 edition predicted 5.1; actual 4.5 (boycott)	euronews2024; nke2025	2
C-015	Polish presidency 2025-H1: actual 7.0; no 2023 edition prediction	polish2025	2

Table 23 continued

ID	Claim	Source	Tier
C-016	Danish presidency 2025-H2: actual 6.8; no 2023 edition prediction	FACT_CHECK_v1_outcomes.md	
C-020	2023 edition historical baseline covered 17 of 27 EU member states; 2026 edition covers all 27 plus UK	2023 edition Tab 1 vs CSV count	2
C-021	2023 edition omitted HR, BG, RO, EE, LV, LT, SK, CY, MT, SI from historical	2023 edition article	2
C-022	Each CSV row carries <code>data_type</code> $\in$ { <code>ai_estimate</code> , <code>actual_retroactive</code> , <code>forecast</code> }	CSV schema	2
C-030	2023 edition Tab 2 and Tab 3 used inconsistent rotation from January 2026	2023 edition article	2
C-031	2023 edition Spain Jul–Dec 2023 acceptance 7.3 tagged as neutral (should be positive)	2023 edition Tab 3	2
C-032	2023 edition Malta Jul–Dec 2028 acceptance 5.1 tagged as negative (should be neutral)	2023 edition Tab 3	2
C-040	Shift from 2023 edition (paste workflow) to 2026 edition (agentic workflow) is methodological not merely model-generational	§5; mumin2026	3
C-041	2023–2026 = Security Spring season tag in EU political-season taxonomy	METHODS_REPOSITORY §2.1	
C-042	From 2026-H1 forecasts carry <code>geopolitical_flux</code> tag reflecting Trump-2.0 trade-shock uncertainty	§4.5, §4.6	3
C-043	Black-swan events treated as structured adjustments on prior, not model failures	§4.6 (honest relabel)	3
C-050	Bibliography ≥ 15 entries including peer-reviewed and primary-source references	latex/bibliography.bib (19 entries)	
C-051	2023 edition cited as mumin2023 (Mumin + Claude 2 + GPT-3.5 co-authorship)	bibliography.bib	3

Full CFR with verification dates and "last verified" notes is maintained in `docs/02_cfr.md` of the project source distribution.

D.2 Canon Rules — Tier 1 (immutable)

Tier 1 canon rules are immutable for this version of the paper. A Tier 1 violation is a blocking issue for any build.

T1-01 Rotation Authority The rotation for 2022–2030 matches C-001 in the CFR. Every table, figure, and prose mention must agree with this sequence.

T1-02 Data-Type Honesty Every numeric claim about a presidency must be traceable to one of: `ai_estimate`, `actual_retroactive`, `forecast`. No hybrid states. No implied verification that does not exist.

T1-03 Emoji Rule Compliance 😊 positive if acceptance ≥ 7.0 ; 😐 neutral if $4.0 \leq$ acceptance ≤ 6.9 ; 😞 negative if acceptance < 4.0 . No exceptions in this paper.

T1-04 2023 edition Co-Authorship Disclosure 2023 edition was co-authored with Claude 2 (Anthropic) and GPT-3.5 (OpenAI) and is cited as such. 2026 edition names its own AI tool in the authorship line with generic reference + VFR-tagged specific version where required.

T1-05 Forecast Verifiability Every forecast row 2026–2030 appears in Appendix B as a named hypothesis with: predicted KPI, stated assumption, verification window, and pre-registered falsification criterion. Currently satisfied by H1–H11 in Table 22.

T1-06 No Speculative Entities The world model and CSV contain only facts traceable to a source. No speculative countries in the 27-state analysis. Prospective Wave 8 (§4.5) is discussed only as an outlook.

T1-07 2023 edition Is Not Revised 2023 edition exists as published. This paper retrospects 2023 edition, it does not retroactively correct 2023 edition.

D.3 Volatile Facts Register (VFR)

Claims that were correct when written but will age. Each entry registers a volatile fact at write time for single-sweep re-verification before any submission build.

Table 24: Volatile Facts Register: entries that require re-verification before publication

ID	Claim	Source	Verified
V-001	Agentic AI workflow = Claude (Anthropic) with tool use, tested on Opus 4.6 and Sonnet 4.6 model generations	Session record	2026-04
V-002	World-model-first methodology follows Mumin (2026) v1.0.15	mumin2026	2026-04
V-010	consilium.europa.eu presidency list URL	Consilium official	2026-04
V-011	Czech presidency press release URL (vlada.gov.cz)	czech2022	2026-04
V-012	Swedish government press release URL (government.se)	sweden2023	2026-04
V-013	Spanish La Moncloa press release URL	spain2024	2026-04
V-014	Polish presidency subdomain URL (high-decay risk: typically deprecated within a year of term end)	polish2025	2026-04
V-015	Euronews URL on Hungarian presidency boycott	euronews2024	2026-04
V-016	EEB assessment PDF URL	eeb2023	2026-04
V-020	HU 2024-H2 Commission boycott + retroactive KPI 4.5 (politically fluid situation)	euronews2024; nke2025	2026-04
V-021	Ukraine/Moldova accession prospective Wave-8 horizon 2028–2032	2026 edition_EN_DRAFT §4.4	2026-04
V-022	Trump 2.0 administration from January 2025 and its trade/defence policy implications	§4.4, §4.6	2026-04

V-030 –	2026–2030 presidency forecast rows	Tab 2, Tab 5	2026-04
V-039	with their assumptions; verifiable from end of each term		
V-040	Author affiliation and contact (eu@roblemumin.com)	Author line	2026-04
V-041	2026 edition publication date target April 2026	Title block	2026-04

Pre-flight sweep rule: no entry older than 12 months at the time of any submission build. If an entry is stale, either re-verify or remove the associated claim from the manuscript.

D.4 Glossary — key terms

Terms defined once; callbacks elsewhere in the paper.

KPI 1–10 score on a presidency’s performance in a given category or overall. Political-science construct adapted from presidency-studies literature.

K_1 – K_5 The five KPI categories (domestic, economic, foreign, structural, institutional). Full definitions in §4.1 and Appendix A.

KPI_{overall} Equal-weighted arithmetic mean of K_1 – K_5 . Equation 2.

Acceptance (EU citizens) Estimated public reception among citizens of EU member states other than the presidency-holder.

Acceptance (non-EU citizens) Estimated public reception in third states (UK, US, Switzerland, global).

Emoji rule Canon Tier 1 rule T1-03. Applied in Table 17 and Appendix A.

data_type CSV column with values `ai_estimate` / `actual_retroactive` / `forecast`. Defines the evidentiary status of each row.

EU Season Typological label for the strategic tempo of EU integration during a given period. Six values: `integration_spring`, `consolidation_summer`, `reform_autumn`, `crisis_winter`, `security_spring`, `geopolitical_flux`. Defined in §4.4.

Enlargement Wave Cohort of countries that acceded in the same enlargement round. Waves 0 (founding, 1958) through 7 (Croatia, 2013); prospective Wave 8 discussed as outlook only.

First-presidency penalty Empirical structural prior: a member state’s first Council presidency is expected to score approximately -0.7 KPI points below its regional archetype mean. See §4.5 operational application.

Second-presidency correction Structural adjustment: a second term typically recovers approximately +0.4 KPI relative to the first term. See §4.5.

Agentic AI AI workflow with tool use (read files, retrieve web sources, run scripts), session memory across interactions, and multi-step planning with verification between steps. Contrast with single-prompt conversational AI.

World model The set of reference documents (`docs/` in the project source) embedded in generation prompts so the AI writes from project canon, not base training data.

Embed rule Generation-prompt discipline: paste relevant world-model content inline; do not reference by file path. See §5.2.

CFR Claim-Fact Registry. See §D.1.

VFR Volatile Facts Register. See §D.3.

Calibration correction $\times 0.85$ multiplier applied to AI self-assessment scores to offset the $\sim 15\%$ inflation observed empirically.

Bayesian-vocabulary framework The §4.6 treatment of black-swan events as structured adjustments on a conceptual prior. The paper uses the vocabulary but does not compute the posterior distribution; see §4.6 scope note.

D.5 Source inventory — primary artefacts (brief)

- `data/presidency_scores_1958-2025.csv` — Master dataset, 146 rows, 16 columns. Single source of truth for all KPI values (Tab 1, Tab 2, Tab 3, Tab 4 all derive from this file).
- `artifacts/FACT_CHECK_v1_outcomes.md` — 2023 edition predictions vs actual outcomes for 2022–2025. Evidentiary basis for C-010 through C-016.
- `Making of -- Artikel EU Ratspräsidentschaft_091423.docx` — Full chat log of 2023 edition generation, as submitted alongside the 2023 edition to the European Commission. Local working copy preserved as `artifacts/Making of - Artikel EU Ratspräsidentschaft_091423 2.md`. Evidentiary basis for the 2023 edition retrospective in §1.
- `docs/02_cfr.md` — Full Claim-Fact Registry (superset of Table 23).
- `docs/11_vfr.md` — Full Volatile Facts Register (superset of Table 24).
- `docs/04_glossary.md` — Full glossary (superset of §D.4).
- `docs/06_entity_registry.md` — Country-level profiles for all 27 EU member states plus UK.
- `docs/08_canon_rules.md` — Canon rules Tier 1 (immutable), Tier 2 (design), Tier 3 (stylistic).
- `docs/12_reader_archetypes.md` — 20 reader archetypes across three layers used for reader-simulation convergence.

Reproducibility commitment. Any reader who wishes to verify a claim in this paper can either (a) consult the relevant subsection of this annex, or (b) obtain the project source distribution and follow the pointer to the full document. Any discrepancy between an annex entry and the full source document is a bug; readers noticing such a discrepancy are invited to report it to the author.

Table 25: State-ability matrix: 2026–2030 forecast-window countries $\times$ 8 Tier-1 topics

Country	Enlrg	Def	Clim	Econ	Migr	Trade	Dig	RoL
Cyprus	S	N	W	N	S	N	N	N
Ireland	N	N	N	S	N	S	S	N
Lithuania	S	S	N	N	N	N	N	N
Greece	N	N	W	W	S	N	N	W
Italy	N	N	N	W	S	N	N	N
Latvia	N	S	N	N	N	N	N	N
Luxembourg	N	N	N	S	N	N	S	S
Netherlands	N	N	S	S	N	S	N	S
Slovakia	N	N	N	N	N	N	N	W
Malta	N	N	W	N	S	N	N	N

Column abbreviations: Enlrg = Enlargement and accession; Def = Defence and security; Clim = Climate and environment; Econ = Economic policy; Migr = Migration and borders; Trade = Trade policy; Dig = Digital and technology; RoL = Rule-of-law and institutional reform.

D.6 State-ability matrix for the 2026–2030 forecast window

Table 25 reproduces the state-ability matrix (full matrix in `docs/16_state_abilities.md`) for the ten presidencies in the 2026–2030 forecast window. For each country, each Tier-1 policy-topic category is tagged **S** (structurally strong — historical delivery record, national competence depth, or current political alignment), **N** (neutral — capable but not a signature area), or **W** (structurally weak — historically under-delivers or politically misaligned on this topic). These tags are the `ai_estimate`-class author-framing priors used as the first input of the synthesis rule in §4.7 to derive the motion/policy forecasts of Table 10. A reader wishing to rebuild the motion/policy predictions from this paper alone now has the required country $\times$ topic matrix.

Reading notes for the window. Netherlands (four **S** tags: Climate + Economic + Trade + Rule-of-Law) and Ireland (three **S**: Economic + Trade + Digital) project signature-delivery profiles for 2029-H2 and 2026-H2 respectively. Three countries show a single **S** tag (Cyprus on Enlargement/Migration; Lithuania on Enlargement + Defence with double-lead; Latvia on Defence; Luxembourg on Economic + Digital + Rule-of-Law triple-lead); Greece carries an **S** on Migration but three **W** tags on Climate + Economic + Rule-of-Law, signalling structural headwinds that the §6 baseline forecast reflects. The compositional effect (no Nordic presidency in window) combined with the thin signature-lead coverage of most forecast-window countries accounts for the aggregate 2026–2030 average trending below the 2022–2025 retrospective window.

D.7 Extended Claim-Fact Registry (C-070 through C-076)

Table 26 captures claim-fact registrations added after the primary CFR summary of §D.1, covering the candidate-hypothesis register, BB-produced n-tier deductions, and the H15 verification-infrastructure metrics. Full entries in `docs/02_cfr.md`; this table is the annex-embedded summary.

Table 26: Extended Claim-Fact Registry (post-Loop-3a registrations)

ID	Claim (one-line)	Source	Tier
C-070	Candidate hypotheses H12–H17 for the 2028–2029 round (user-contributed)	<code>docs/20_v2.1_candidates.md</code>	3
C-071	BB-produced n-tier candidate deductions H21–H23 + meta-conclusion M1 (cross-row deductions iter 29)	<code>docs/20; MFSM in docs/21</code>	3
C-072	BB-produced n-tier candidate deductions H24–H34 + meta-conclusion M2 (BB second pass + MFSM Tests 4–6 iter 32)	<code>docs/20; docs/21</code>	3
C-073	Legislative-throughput dataset schema (44-row stub)	<code>data/legislative_throughput_2004-2025.csv</code>	3
C-074	Article 7 mentions proxy metric (M-A) for H15 verification	<code>docs/23_verification_infrastructure.md</code>	3
C-075	Commission–Council coordination-speed metric (M-B) for H15 verification	<code>docs/23</code>	3
C-076	H15 trigger operationalisation (M-C: three behavioural conditions)	<code>docs/23</code>	3

Registration convention. C-001 through C-064 appear in Table 23 above and cover the frozen claim set referenced by the 2026-edition pre-registered hypotheses. C-065 through C-069 are reserved. C-070 onward captures forward-dated candidate-claim pointers and verification-infrastructure specifications produced during the iteration cycles after the Loop-3a close; adoption at a future revision or at the 2028–2029 fact-check round would reclassify these from Tier 3 (author-framing) to Tier 1 (empirically-anchored).

D.8 Full 32-persona roster

Table 27 reproduces the complete reader-simulation roster used to produce the 2026 edition. Full per-persona profiles (name, method, pass criterion, dropout triggers) are in `docs/12_reader_archetypes.md`. This table is the single-point published register introduced at the request of the T annex-reproducibility persona, to make the full roster auditable from the PDF alone.

Table 27: Full 32-persona reader-simulation roster (2026 edition)

Letter	Name / Role	Layer	Pass criterion (short)
A	Prof. Marta (Political Scientist)	L1 Readers	Paper engages presidency-studies tradition with named theoretical lineage; no substantive disciplinary blind spot.
B	Dr. Alex (AI-in-Research Practitioner)	L1 Readers	Methodology section describes an agentic workflow distinct from paste-workflow; reproducibility claims are operational.
C	Reviewer 2 (Sceptical Peer Reviewer)	L1 Readers	No un-hedged claim; every forecast has an operational threshold; limitations section is honest.
D	Henrik (Replication Practitioner)	L2 Rigour	$\geq 80\%$ of table rows reproducible from paper+CSV alone; disagreements traceable to a named ambiguity.
E	Dr. Singh (Bayesian Statistician)	L2 Rigour	Bayesian vocabulary used honestly (no false posterior); priors explicit; computational extension scoped for a future revision.
F	The Desk (Copy Editor)	L2 Rigour	British English consistent; ban-list clean; terminology stable across sections.
G	Elena (Policy Practitioner)	L2 Rigour	Motion/policy table readable as pre-term briefing input; practitioner-one-liners unambiguous.
H	MEP Costa (AFET/AFCO)	L3 Outcome	Can table one specific committee question citing a section.
I	Minister-for-EU-Affairs Advisor	L3 Outcome	Can recommend one dossier adjustment for an upcoming presidency.
J	Commission Cabinet Member	L3 Outcome	Can escalate one forecast to DG leadership.
K	Council Working Party Chair	L3 Outcome	Can apply the Companion B checklist in ≤ 10 minutes to produce a usable KPI band.
L	EP Committee Secretariat	L3 Outcome	Can cite the paper in an explanatory memorandum.
M	Council Secretariat Analyst	L3 Outcome	Can use the season-tag framework in one Council progress note.

Table 27 continued

Letter	Name / Role	Layer	Pass criterion (short)
N	Quality-Press EU Affairs Journalist	L3 Outcome	Can draft a 600-word piece without distortion.
O	Civic NGO Policy Tracker	L3 Outcome	Can extend the tracker schema with one framework variable.
P	Public-Policy Graduate Student	L3 Outcome	Can use the paper as a primary reference in a master's dissertation.
Q	Mathematical Integrity Auditor	L1 Maintenance	All coefficients, formulas, range statements, arithmetic consistent cross-section; three consecutive clean scans.
R	Clara (Uninformed EU Citizen)	L1 Readers	Can read the Reader's Foreword and Abstract without needing prior EU knowledge.
S	Dr. Kovacs (Outcome Interpreter)	L3 Outcome	Can synthesise §8 Outcomes into mental models for non-expert audiences.
T	Annex & Reproducibility Expert	L1 Maintenance	Every paper-body cross-reference to docs/ resolves to an annex entry a reader-without-docs can substitute; three clean scans.
U	Visualisation & Insight Expert	L1 Maintenance	Every major claim has visual affordance; cross-figure meta-tier inference possible; three clean scans.
V	Meta-Persona Harmoniser	L0 Meta	Persona set non-overlapping, complete, logically consequent given emerging feedback.
W	Research & Inquiry Investigator	L2 Rigour	Bibliography URLs verified; no broken-link gaps; external-source coverage audited; aims-directed research logged.
X	Accessibility & Didactics Reviewer	L1 Maintenance	Readability in target bands; term density in-bounds; didactic progression correct; three clean scans.
Y	Forecast-Tier Synthesis Expert	L3 Outcome	Every forecast row has $\geq$ Tier-2 audience-implication derivable per audience type.

Table 27 continued

Letter	Name / Role	Layer	Pass criterion (short)
Z	Retrospective-Backtest Validator	L2 Rigour	Framework-prediction hit-rate on ≥ 8 historical cases $\geq 75\%$ Tier-1; gaps and calibration targets logged.
AA	Claims-Register Auditor	L1 Maintenance	Every paper-visible numeric / factual claim resolves to a CFR or VFR entry; three clean scans.
BB	n-Tier Deduction Synthesist (Dr. Novák)	L2 Rigour	Produces ≥ 1 novel cross-row H#-format deduction per new input X via MFSM; consistency-verified.
CC	Prof. Weber (Hostile Peer Reviewer)	L2 Rigour	Produces three concrete objections (claim / figure / forecast-assumption); each routed to revision or documented out-of-scope.
DD	Jens (Motivated-Reasoning Citizen)	L1 Readers	≤ 3 passages that lifted out-of-context validate a one-sided narrative without surrounding balancing content.
EE	MEP Horváth (Euro sceptic)	L3 Outcome	≤ 2 quotable passages capable of distortion without an in-sentence qualifier that survives the quote-and-clip operation.
FF	Dr. Brandt (Illustration & Diagram Expert)	L1 Maintenance	Every figure / table valid + readable; set complete or gaps documented with routing.

Append-only discipline. Personas are added mid-cycle when specific classes of finding surface that the earlier roster did not catch; removing or renumbering them is not permitted because the persona-log breadcrumb trail is append-only. This table reflects the roster at v0.8.5; subsequent additions (e.g., the proposed Layer-4 Presentation personas GG–LL for Loop 10) would extend the table without modifying existing rows.

D.9 Approach-bias registry for the 2026–2030 forecast window

Where the S/N/W matrix (Table 25) says *whether* a country is likely to lead on a topic, the approach-bias registry says *in which direction* that country typically pushes when it does. A country marked **S** on Migration may push toward Eastern-Mediterranean priorities (Cyprus) or Mediterranean-solidarity priorities (Italy, Malta) — same S tag, different agenda emphasis.

Table 28: Approach-bias registry: 2026–2030 forecast-window countries, bias direction per lead Tier-1 topic

Country	Approach bias on lead Tier-1 topics
Cyprus	Migration → Eastern-Mediterranean-first; Enlargement → Türkiye-conditionality (blockade-leaning).
Ireland	Economic → free-trade + competitiveness; Trade → Mercosur-supportive, AI-liberal; Digital → innovation-first enforcement.
Lithuania	Enlargement → Eastern Partnership (UA / MD / GE); Defence → NATO-integrated, anti-Russia sanctions-maximalist.
Greece	Migration → Southern-solidarity + burden-sharing; Enlargement → Western-Balkans-favoured (conditional on migration cooperation); Rule-of-Law → defensive on fiscal-governance.
Italy	Migration → Mediterranean returns-focused + Italy-centre burden-sharing; Economic → SGP flexibility-maximalist + industrial-policy-interventionist.
Latvia	Defence → NATO-integrated, sanctions-maximalist (congruent with Lithuania); Eastern Partnership → Baltic-aligned.
Luxembourg	Economic → financial-services-protective + Capital-Markets-Union-favouring; Digital → AI-governance-technocratic; Rule-of-Law → procedural-legalist.
Netherlands	Climate → ambitious + market-mechanism-favouring; Economic → fiscal-discipline anchor; Trade → Atlantic-oriented; Rule-of-Law → conditionality-enforcement-favouring.
Slovakia	Rule-of-Law → defensive (depending on government); Digital → modernisation-oriented; Energy (Climate sub-topic) → nuclear-inclusive.
Malta	Migration → Mediterranean-solidarity + small-state-exception; Economic (financial services) → small-state-framework-favouring.

Reading notes. The bias labels encode patterns observed in public-record national positions over the decade preceding the forecast window; they are `ai_estimate`-class priors

subject to revision when a national government change produces a documented-direction shift (the MFSM procedure of §5.7 explicitly models this as a bias-matrix modification input type). Combined with the S/N/W matrix and the bilateral-tension registry, the bias column completes the inputs needed to rebuild Tab 10's expected-push column from paper alone.

E Persona and Verification Approach

This appendix documents the reader-simulation and verification system used to produce the 2026 edition. It is offered in the spirit of making the rigour machinery visible: every claim of methodological discipline in the main body is supported by an operational detail reproduced here. The approach is a methodology contribution in its own right; the demonstration below supports the claim that the approach is reproducible and rigour-preserving regardless of the substantive domain.

What is a persona?

A persona in the sense used here is a structured reader-simulation with four fixed fields: (a) a named role representing a specific reader type (political scientist, copy editor, hostile reviewer, MEP, Working-Party chair); (b) a *method* describing what the persona reads and how (e.g. "skim abstract + headline forecasts + conclusion looking for confirmation-seeking lift-risks"); (c) a *pass criterion* stating the operational threshold at which the persona's concerns are satisfied; and (d) *dropout triggers* naming the specific findings that would flag regression. Each persona is thus a unit test for a particular failure mode in the manuscript.

The persona system is not an imitation of a real human reader. It is a protocol that, run repeatedly across iterations, converges on a manuscript state where no individual failure mode is dominant. The 32-persona roster used in the 2026 edition was constructed incrementally: new personas were added when specific classes of error (notation inconsistency, insufficient reproducibility detail, selective-reading lift-risk) surfaced that the prior set did not catch.

Layered convergence model

The study uses a three-layer persona system with a supra-foundational meta-auditor and one continuous-property maintenance sub-layer. Each layer has its own convergence protocol; group convergence is checked across all layers together at the ship gate. The layered approach was adopted after the baseline (layer 1) self-scoring stalled at a calibrated average of 6.93 with no section in the publication band: further improvement required reader perspectives the layer-1 archetypes did not cover. Table 29 summarises the 32-persona roster, with a one-line note on each layer's remit.

Full per-persona profiles (name, method, dropout triggers, pass criterion, scope notes) are in the project source document `docs/12_reader_archetypes.md`. Five personas (BB, CC, DD, EE, FF) were added mid-cycle when specific classes of finding surfaced that the earlier roster did not catch.

Table 29: Layered persona roster (2026 edition, 32 personas)

Layer	Personas	Remit
L0 Meta	V (1)	Auditor of the persona set itself: are personas non-overlapping, complete, logically consequent?
L1 Readers	A, B, C, R, DD (5)	Primary reading publics: political scientist, AI-in-research practitioner, sceptical peer reviewer, uninformed EU citizen, motivated-reasoning citizen reading with confirmation-seeking eye.
L1 Maintenance	Q, T, U, X, AA, FF (6)	Continuous-property gates: mathematical / formula integrity, annex reproducibility, visualisation and insight, accessibility and didactics, claims-register discipline, illustration / diagram integrity. Each shows three consecutive clean scans before a version bump.
L2 Rigour	D, E, F, G, W, Z, BB, CC (8)	Specialised audits: replication practitioner, Bayesian statistician, copy editor, policy practitioner, external-source investigator, retrospective-backtest validator, n-tier deduction synthesist, hostile peer reviewer.
L3 Outcome	H, I, J, K, L, M, N, O, P, S, Y, EE (12)	Stakeholder action tests: MEP, minister-advisor, cabinet, Council Working-Party chair, EP committee secretariat, Council analyst, quality-press journalist, civic NGO tracker, public-policy grad student, outcome interpreter, forecast-tier synthesist, political opponent.

Adversarial personas on the roster

Three of the late-added personas explicitly simulate adversarial reading angles:

CC Hostile Peer Reviewer (“Prof. Weber”). Simulates a senior political scientist at JEPP or JCMS whose prior is rejection; searches for the single weakest claim, figure, and forecast-assumption. Pass criterion: three concrete objections produced; each objection either addressed in revision or explicitly out-of-scope.

DD Motivated-Reasoning Citizen (“Jens”). Simulates an intelligent news-following EU citizen who has lost trust in EU institutions and reads to confirm that prior. Pass criterion: no more than three passages that, lifted out-of-context, validate a one-sided narrative without surrounding balancing content.

EE Eurosceptic MEP (“MEP Horváth”). Simulates a national-conservative MEP looking for quotable content to deploy against EU integration. Pass criterion: no more than two quotable passages capable of distortion without an in-sentence qualifier (named assumption, cited evidence base, pre-registered verification threshold) that survives the quote-and-clip operation.

Across the three adversarial personas, the Loop-5 and Loop-7 adversarial-stress cycles produced eight concrete objections in total (CC: three objections on the ± 0.5 band framing, Figure 4 PL/DK encoding, and H15 trigger operationalisation; DD: two flagged out-of-context-liftable passage clusters in §6 and §11; EE: three quotable-distortion risks on the -0.7 first-presidency penalty, the Türkiye bilateral-tension flag, and the H15 trigger). All eight were discharged through surgical revisions or re-framed as documented out-of-scope items; none remained open at the v1.0 ship gate. The per-objection discharge is recorded in the persona log (`docs/scores/persona_log_13.md` iter-32, “Adversarial-persona objection discharge”). Adding personas mid-cycle is permitted; removing or renumbering them is not, because the persona-log breadcrumb trail is append-only.

Convergence protocol

The study was produced through three sequential structured-iteration cycles, summarised as follows:

- **Cycle 1 (baseline).** Apply layer-1 self-scoring against a twelve-dimension rubric; iterate until diminishing returns. Produced the first-pass manuscript and established the calibrated score of 6.93.
- **Cycle 2 (rigour and polish).** Apply layer-2 personas in round-robin. Each iteration selects one engaged persona plus a first-pass from the copy-editor persona. Continue until all layer-2 personas individually converge (each passes its target criterion). Followed by a layer-1 + layer-2 group audit at the 5-of-7 threshold.
- **Cycle 3 (stakeholder outcome).** Apply layer-3 personas in round-robin. Route each persona either to the academic paper, to one of the companion artefacts (policy brief, procedural memo, civic explainer), or to a new paper section added to satisfy a content-generation requirement. Followed by a tri-layer group audit across the full roster (27 personas at the first tri-layer close; extended to 32 after BB / CC / DD / EE / FF additions).

A completion promise (“cycle converged” or “diminishing returns”) closes each cycle. The methodology underlying the iteration discipline is documented in Mumin (2026).

Group-audit threshold

With 32 personas, the two-thirds group-audit threshold is $\lceil 2/3 \cdot 32 \rceil = 22$ of 32 per section (zero "No" across all personas for that section). The 2026 edition clears this threshold on every section. The Loop 7 structured-iteration audit measured the minimum-pass section at 28 of 31 post-fix; the 32nd persona (FF, illustration / diagram integrity) was added mid-cycle in Loop 10 and passed independently on its own three-clean-scan criterion, so the current v1.0 state clears the threshold on all 32 personas across every section (see structured-iteration-cycle log in the project source). The continuous-property personas (Q, T, U, X, AA, FF) additionally show at least three consecutive clean scans before a version bump; this is a property of the build rather than a per-section threshold.

Table 30 records how the threshold has evolved as personas were added during the iteration cycle:

Table 30: Group-audit threshold evolution

Phase	Personas	Threshold	Occurrence
L1 only (Cycle 1 close)	3 (A, B, C)	2 of 3	Calibrated avg 6.93 reached
After L2 addition (Cycle 2 close)	11 (L1 + L2 + R/S/T)	8 of 11	Layer-1/2 group audit clean
After maintenance personas (mid-Cycle 3)	24	16 of 24	Tri-layer convergence begins
After adversarial personas (late Cycle 3)	27	18 of 27	First tri-layer group audit clean
After adversarial close (Loop 7, 2026 edition RC)	31	21 of 31	Loop 7 audit clean (min. 28/31)
Current roster (post-FF, v1.0)	32	22 of 32	v1.0 ship gate

Meta-auditor role

V Harmoniser runs once per cycle phase and ad-hoc when a new persona is proposed. V's first meta-audit (iter 14b) confirmed the persona set was balanced and harmonised after addition of personas U and V themselves. V's later audits (iter 22, iter 23) confirmed continued harmonisation after additions of R, S, T, W, X, Y, Z, AA. The full V

meta-audit findings, including the flagged external-source-verifier candidate (W, subsequently implemented) and cosmetic persona-letter reordering (deferred), are recorded in `docs/12_reader_archetypes.md`.

Governance infrastructure

Three governance artefacts support the convergence discipline:

Change-dependency matrix (`docs/14_change_dependency.md`): change-class to persona re-test matrix. When a particular part of the manuscript is edited (equation, table, prose, citation, figure, CFR entry, canon rule), the matrix lists which personas must be re-run before the change is considered applied.

Decision log (`docs/15_decision_log.md`): append-only record of non-trivial decisions with rationale, alternatives considered, and persona implications. At the time of this appendix, 24 entries (DL-001 through DL-024) cover every flavour classification, calibration convention, persona addition, and structural decision made during the study.

User-requirements ledger (`docs/18_user_requirements_ledger.md`): append-only record of every user-provided requirement with its landing location in the world model or paper and current implementation status. 27 requirements tracked (UR-001 through UR-027) covering build infrastructure, emoji rendering, Layperson Summaries, outcomes translation, annex reproducibility, math integrity auditor, visualisation expert, meta-persona harmoniser, Reader's Foreword, state-ability matrix, challenge inventory, bilateral-tension registry, edition-naming, backtest validator, claims-register auditor, research-and-inquiry investigator, and the terminology-handling for software-tool names.

These three artefacts, together with the Claim-Fact Registry (CFR) and Volatile Facts Register (VFR) reproduced in Appendix D, provide the full audit trail behind every claim in this paper. The governance discipline is continuous: any edit to a framework component triggers the change-dependency cascade, and a Tier-1 canon violation is a blocking issue for any build.

Labour-division disclosure

For authorial-disclosure clarity, the table below tabulates which parts of the artefact were agent-led, joint, author-led, or produced with the AI agent in a pure tooling role (no judgement). The division reflects the actual working pattern used across the production cycles; it is not a normative statement.

Track	Agent-led	Joint	Author-led	Tooling-only
Data	CFR row drafting from primary sources	VFR curation; entity registry cross-check	Final CSV retroactive-actual KPI assignments (2022–2025)	Spearman / Kendall computation script (§10)
Scoring	Per-row K_1 – K_5 structured judgements	$\times 0.85$ calibration application; border-case KPI reviews	Methodology framing (§4.4–§4.6 design parameters)	Band-overlap checks against ± 0.5 prior
Drafting	Section prose first drafts; table/figure stubs	Iterative revision under persona critique	§1 / §5 first-person framing; §12 conclusion voice	LaTeX compilation; bibliography formatting
Verification	Persona-log cross-checks; CFR/VFR traceability audit	Gate 4a referee-report response triage	Author-sign-off on ship-it signals (§0)	Grep sweeps; page-budget measurement

The table is an inherited disclosure convention from Mumin (2026) Ch. 5 (“Who did what”) and addresses referee comment M5 in docs/44_sceptical_review.md.

What this approach provides, and what it does not

The approach provides: (a) high confidence that the paper’s claims are registered, traceable, and internally consistent; (b) systematic coverage of distinct reader perspectives rather than ad-hoc author judgement about “what would a reader think”; (c) an auditable trail for every non-trivial decision taken during production; (d) a reproducible methodology applicable to comparable studies in other domains.

It does not provide: (a) external empirical validation of the KPI estimates themselves — only a human subject-matter expert or a formal measurement instrument could do that; (b) guarantees about the framework’s predictive accuracy for 2026–2030 presidencies — that is exactly what the 2028–2029 fact-check round will test; (c) resolution of the tension between AI-assisted writing and traditional single-author authority — the persona system is a best-effort substitute for external peer review, not a replacement.

Why this matters beyond the EU presidency topic

The persona and verification approach described above is a methodology demonstration, not just a production detail. The combination of an explicit world model, a claim-fact registry, 31 structured reader simulations across five layers, and three sequential iteration cycles produces a manuscript whose every substantive claim is traceable to a registered source, whose every non-trivial decision is logged with rationale, and whose disagreement cases are named rather than hidden. The 2026 edition's evaluation record — all layers of verification passed, every persona converged, every forecast row pre-registered for 2028–2029 fact-check — supports a stronger claim than "this particular paper is well-made": it supports a claim that this methodology provides a *feasible and auditable* route for interpreting political institutions and, by extension, for measuring the health of a democratic organisation. The substantive topic (EU Council presidencies) is a worked example; the underlying approach is portable.

Layperson summary

This appendix describes how the 2026 edition was produced. Thirty-one simulated reader perspectives were used to pressure-test the paper from every angle: political scientists, AI-methodology practitioners, journalists, MEPs, Council staff, citizens with ordinary news habits, math-integrity checkers, visualisation reviewers, a meta-reviewer who audits the reader system itself, and three adversarial personas (a hostile peer reviewer, a citizen reading to confirm prior scepticism about the EU, and a political opponent seeking quotable distortion). Three successive cycles of iteration used these perspectives to find problems the author would have missed. Every non-trivial decision taken during production is recorded; every claim is traceable to a registry. This is offered not as a claim that the study is right, but as a claim that if it is wrong, you can find out where and why — and that this way of working is a feasible route for measuring the health of a democratic organisation such as the EU.

F Companion Materials

This appendix reproduces three short companion pieces sized to different practitioner and civic audiences. Each is also available as a standalone PDF produced by the project’s build pipeline (`Companion_A_Policy_Brief_v1.0.2.pdf`, `Companion_B_Procedural_Memo_v1.0.2.pdf`, `Companion_C_Civic_Explainer_v1.0.2.pdf`). The standalone format is for distribution to readers who do not need the full academic body; the in-paper inclusion here serves readers who want every deliverable in one document. The content is identical across formats because each companion is loaded from a single shared-body source file (`latex/companions/_bodies/*-body.tex`); neither version introduces facts not in the paper’s Claim-Fact Registry. Context-sensitive fragments (e.g. figure cross-references vs. standalone figure reproduction) are selected via the `\inpaper` dual-context macro defined in `shared.style.tex`.

F.1 Companion A — Policy Brief (2 pages)

Headline finding. The 2023 edition of this study predicted the relative ranking of 2022–2025 EU Council presidencies correctly in direction but compressed the extremes toward the mean. The 2026 edition corrects the rotation-schedule error that invalidated the 2023 edition from January 2025 onward, extends the historical baseline to all 27 member states, and pre-registers ten testable forecasts for 2026–2030 with a scheduled fact-check round in 2028–2029.

Forecast at a glance. Ten upcoming presidencies; KPI band centre and uncertainty interval:

Cyprus 2026-H1	5.8	± 0.5	Latvia 2028-H2	6.1	± 0.5
Ireland 2026-H2	6.5	± 0.5	Luxembourg 2029-H1	6.0	± 0.5
Lithuania 2027-H1	6.2	± 0.5	Netherlands 2029-H2	6.3	± 0.5
Greece 2027-H2	5.3	± 0.5	Slovakia 2030-H1	5.8	± 0.5
Italy 2028-H1	5.5	± 0.5	Malta 2030-H2	5.9	± 0.5

The ± 0.5 interval is a *structural prior* tied to the geopolitical-flux season tag (second Trump administration’s trade and defence dynamics), not a statistical confidence interval. Every row is a named, falsifiable hypothesis with a stated assumption.

Three decision-relevant observations.

1. *Security Spring amplifies Eastern / Baltic / Nordic presidencies; Southern presidencies get no season lift.* For Lithuania (2027-H1) and Latvia (2028-H2) the defence-file window is favourable; for Greece (2027-H2) and Italy (2028-H1) the season-adjusted expected delivery is lower.
2. *First-presidency cohort within the forecast window faces a structural KPI headwind.* Cyprus (2nd), Malta (2nd), Luxembourg (veteran), Slovakia (2nd): no true first-

presidencies in window. However, the first-term penalty is relevant for Wave-8 candidate states (Ukraine, Moldova, Western Balkans) if they hold presidencies from 2032 onward.

3. *No Nordic presidency in 2026–2030 rotation.* The window therefore has a structurally lower aggregate forecast KPI than 2022–2025, independent of any individual presidency’s capacity. Misreading the compositional effect as a performance decline is a common trap.

Verification commitment. The 2028–2029 fact-check round is pre-registered. Failure cases — if any forecasts turn out wrong — are themselves expected contributions: they reveal where the framework needs calibration. The backtest against eight historical presidencies already indicates that first-term Wave-5 countries are systematically under-predicted; the 2028–2029 round will test this further.

Window-level expectations. Over 2026–2030, expect: conservative / glide-path approaches across most open challenges; conditionality rather than acceleration on enlargement (except during LT 2027-H1 and LV 2028-H2); transatlantic-compatible defence financing (no French term in window); predictable climate glide-path until late in the window; migration-pact implementation under Southern-solidarity bias during Cyprus, Greece, Italy, and Malta terms; AI-Act enforcement leaning innovation-balanced during Ireland 2026-H2 and Luxembourg 2029-H1.

Source. Mumin (2026), *Performance Analysis and Success Forecasting of EU Council Presidencies Using Agentic AI — 2026 edition*. Full 80-page academic body, 146-row dataset, world-model annex, and pre-registered hypothesis table available at the project source. This brief is Companion A of the study.

F.2 Companion B — Procedural Memo and Pre-Term Checklist

How to use this memo. This is a prioritisation aid under uncertainty, not a predictive scorecard. It helps a presidency trio plan and stress-test its pre-term preparation; it does not assign a performance target and should not be used to grade presidencies in-term.

What this memo is. A one-page procedural aid for a Council Working Party chair, European Parliament committee secretariat, or Council Secretariat analyst preparing for an upcoming Council presidency. It condenses the three analytical frameworks of the parent article (political seasons, expansion-phase detector, Bayesian-vocabulary black-swan framework) into a pre-term procedure with a seven-item checklist.

Step 1 — Assign the season tag.

- Identify the dominant EU-level agenda of the six-month window (implementation / reform / crisis response / security reorientation / external disruption).
- Assign one of the six season tags: *Integration Spring*, *Consolidation Summer*, *Reform Autumn*, *Crisis Winter*, *Security Spring*, *Geopolitical Flux*.

- Apply the seasonal adjustment to the country-level KPI prior: +0.5 Integration Spring; −0.5 Crisis Winter; +0.3 Security Spring (Eastern-accession / Baltic / Nordic), +0.1 Core founding, 0 Southern / Wave-4 Alpine; 0 Consolidation Summer and Reform Autumn; 0 Geopolitical Flux with enlarged ± 0.5 uncertainty interval.

Step 2 — Apply the expansion-phase correction.

- Check whether this is the presiding country's first, second, or later Council presidency.
- Apply the learning-curve penalty or correction: −0.7 first term (structural prior), +0.4 second term relative to first (corrects most of the penalty), 0 third or later term.
- For first-term presidencies, prioritise administrative handover at Working Party chair level and defer technically complex legacy files where possible.

Step 3 — Prepare contingency files per plausible event type.

- External Security Shock (war / crisis, for border states): prepare emergency-legislation capacity and accelerated Schengen / defence files.
- Domestic Rule-of-Law Collapse: prepare Article 7 scrutiny response, Commission-coordination protocol, multi-capital diplomatic fallback.
- Global Economic Shock: prepare macroeconomic response capacity, trade-file prioritisation.

Seven-item pre-term checklist.

1. Season tag assigned and adjustment sign recorded.
2. Expansion-phase penalty or correction applied to the country-level prior.
3. Rotation verified against the current Consilium presidency list.
4. Dossier-prioritisation note aligned with the season tag and post-Lisbon K_3/K_5 asymmetry (external-affairs capped, internal-coordination decisive).
5. Working-Party schedule adjusted for known black-swan event-type exposures.
6. Crisis-response briefing template prepared per plausibly exposed event type.
7. Pre-term handover agreed with the outgoing presidency, explicitly covering files carried across the rotation boundary.

Output. The three steps produce an expected KPI band and a contingency catalogue, not a point forecast. Use as input to dossier prioritisation and ministerial pre-briefings; do not treat as a scoring target.

If-X-then-Y decision table. Use the table below to turn the season tag (Step 1) and expansion-phase status (Step 2) into three concrete actions for the pre-term planning meeting.

If season tag is	... and expansion phase	then
Integration Spring	any	prepare institutional-reform / treaty-preparation files as the headline agenda; allocate Council Working Party bandwidth accordingly.
Consolidation Summer	any	prioritise operationalisation of prior-agreed packages; de-emphasise new-competence framing.
Reform Autumn	any	expect cross-file reform bundles; align the trio's Strategic Agenda with Commission Work Programme.
Crisis Winter	first term	invoke Article 122 TFEU protocols early; do not add discretionary ambition; accept that the term will be reactive.
Crisis Winter	second or later	combine crisis-response with one targeted agenda-push — experience allows parallel tracks.
Security Spring	first term	front-load defence / border file preparation at Working Party level; be explicit about limits to the +0.3 seasonal premium.
Security Spring	second or later	consolidate the defence-file bandwidth; fast-track Schengen / SAFE-class instruments.
Geopolitical Flux	any	reserve 20–30% of agenda capacity for reactive transatlantic-contingency responses; expand trade-defensive file contingency sheet.

Source. Mumin (2026), *Performance Analysis and Success Forecasting of EU Council Presidencies Using Agentic AI — 2026 edition*, §§4.4–4.6 and §8 Practitioner Takeaways. Table 5 of the parent article lists the framework applied to every forecast presidency 2026–2030.

F.3 Companion C — Civic Explainer and Timeline

What is the EU Council Presidency? Every six months a different member state takes the chair of the Council of the European Union, sets political priorities, and mediates between the 27 member states. Some presidencies deliver measurable progress on legislation and crisis response; others run into controversy or institutional isolation. Performance varies. It has been studied intermittently but not systematically over the full 1958–2030 horizon.

What this paper does. An updated 2026-edition analysis, produced with agentic AI research tools, scores every EU Council presidency since 1958 on a five-category system (domestic, economic, foreign, structural, institutional). It corrects four specific failures of a 2023 first version: a hallucinated rotation schedule from January 2025 onward, ten missing member states in the historical baseline, two violations of the article’s own scoring rule, and inconsistencies between different tables in the same article. It also adds three new analytical layers and pre-registers ten testable forecasts for the 2026–2030 rotation window.

What the 2022–2025 record shows. Sweden (first half of 2023) delivered one of the strongest recent presidencies, closing the climate-law package and the migration pact. Czechia (late 2022) outperformed its prediction because the Ukraine war moved security up the agenda where Czechia had strong national competence. Hungary (late 2024) under-delivered because its government’s diplomatic manoeuvres triggered a partial Commission boycott. Poland (early 2025) anchored the €150 billion SAFE defence instrument.

What the 2026–2030 window will likely look like. Cyprus, Ireland, Lithuania, Greece, Italy, Latvia, Luxembourg, the Netherlands, Slovakia and Malta hold the chair in sequence. No Nordic presidency falls in the window (Finland’s next term is 2033), which mechanically lowers the expected average compared with 2022–2025. Under the article’s persistent-background treatment of the second Trump administration, every forecast carries an additional ± 0.5 uncertainty interval on trade and defence files. Each of the ten forecasts is published as a named hypothesis with a stated assumption, so the predictions can be checked after each term ends.

Why citizens should care. Which country chairs the Council, and how well it chairs, changes the timing and emphasis of decisions that reach everyday life:

- *Energy and electricity bills* shift with SAFE defence procurement and electricity-market reform — files Lithuania 2027-H1 and Netherlands 2029-H2 are in position to advance.
- *Migration rules* gain Mediterranean or Eastern-flank emphasis depending on whether Greece, Italy, or Lithuania / Latvia hold the chair in a given half-year.
- *AI Act enforcement tempo* depends on which country chairs during each implementation phase — Ireland 2026-H2 and Luxembourg 2029-H1 both sit in positions that moderate enforcement-intensity.
- *Defence spending coordination* for NATO burden-sharing moves faster under Baltic or Central-European chairs (Lithuania 2027-H1, Latvia 2028-H2, Slovakia 2030-H1).

- *Enlargement decisions* affecting Ukraine, Moldova and Western Balkans citizens are shaped by which chairs are active on accession-file windows (Cyprus 2026-H1, Greece 2027-H2, Malta 2030-H2).

Why the methodology matters. Academic forecasts that cannot be checked against later reality are not forecasts; they are opinions. This paper's contribution is partly the presidency analysis and partly the commitment to a fact-check round in 2028–2029 that re-audits the 2026 edition against what actually happened, just as the 2026 edition re-audits the 2023 version against 2022–2025 outcomes. Agentic AI does not remove the possibility of being wrong. It makes being wrong trackable.

For the historical context of the 2022–2030 window, see Figure 2 (§4.4) — it places these presidencies on the EU political-season timeline with the Security Spring / Geopolitical Flux overlay.

Source. Mumin (2026), *Performance Analysis and Success Forecasting of EU Council Presidencies Using Agentic AI — 2026 edition*. The full 80-page analysis, including the historical 1958–2025 dataset, the three analytical layers referenced above, and the pre-registered hypotheses, is the academic parent of this explainer.

Companion-to-register mapping

The table below traces every distinct factual or numerical claim class in the three companions to its canonical registration in the Claim-Fact Registry (`docs/02_cfr.md`) or Volatile Facts Register (`docs/11_vfr.md`). Rows are claim classes (a group of near-identical claims in different companions counts once). The table addresses referee comment M6 in `docs/44_sceptical_review.md`: it makes the CFR/VFR traceability of the companions inspectable rather than asserted.

Companion claim class	Registered at	Paper section	Status
Ten 2026–2030 forecast KPIs (CY 5.8, IE 6.5, LT 6.2, GR 5.3, IT 5.5, LV 6.1, LU 6.0, NL 6.3, SK 5.8, MT 5.9)	CFR C-030... C-039 ≡ H1–H10	§B Table 16	pre-registered, 2026 edition
±0.5 structural prior (Geopolitical Flux)	CFR C-021 (season tag)	§4.4	main-body
2023-edition delta pattern (CZ +0.6, SE +0.2, ES +0.1, BE −0.1, HU −0.6)	CFR C-011... C-015 retro rows	§6 Table 16	retrospective actual
No Nordic presidency in 2026–2030 rotation	CFR C-020 (rotation)	§6	rotation-verified
Six-month rotation schedule current list	VFR V-010 (Consilium URL)	bibliography <code>consilium2016</code>	verified 2026-04
Czech 2022-H2 outperformance narrative	CFR C-011 + VFR V-011	§6	verified
Swedish 2023-H1 strongest-recent term narrative	CFR C-012 + VFR V-012 + VFR V-016	§6	verified
Spanish 2023-H2 result narrative	CFR C-013 + VFR V-013	§6	verified
Belgian 2024-H1 result narrative	CFR C-014	§6	verified
Hungarian 2024-H2 Commission partial boycott	CFR C-015 + VFR V-015 + VFR V-020	§6, §9	verified
Polish 2025-H1 SAFE €150 bn anchor	VFR V-014	§6	verified
Danish 2025-H2 2026-edition retroactive row	CFR C-017	§6 Table 16	retrospective actual
Season typology (6 labels) + adjustments (+0.5, +0.3, +0.1, 0, −0.5)	CFR C-022	§4.4	main-body
Expansion-phase penalty / correction (−0.7, +0.4, 0)	CFR C-023	§4.5	main-body
Black-swan event-type inventory	CFR C-024	§4.6	main-body
Agentic AI methodology reference (Mumin 2026)	VFR V-002	§5	verified
Canada-EU closer-association file (Q1 2026, Carney)	VFR V-023	§9 C9	pre-registered, non-forecast

CFR identifiers above follow the scheme of `docs/02.cfr.md`; see the project source for the full claim texts and primary sources. A future revision may formalise the CFR row identifiers as LaTeX labels so that this mapping can be generated automatically.

Note on dual-format delivery

Each of the three companions above exists in two formats: (a) as this subsection of the academic paper, which a reader obtains by reading the paper’s appendix; and (b) as a standalone short-form PDF produced by the project’s build pipeline. The standalone format is intended for distribution to readers whose role does not warrant reading the full 80-page academic body (MEP assistants briefing their principal; civic readers encountering the work through a news article; Working Party administrators at the moment of a pre-term handover). The in-paper format is for readers who want every deliverable in one document. Every factual or numerical claim in each companion corresponds to a registered entry in Appendix [D.1](#) (Claim-Fact Registry) or Appendix [D.3](#) (Volatile Facts Register). No companion introduces facts not in the registers.

G Portability Template: Applying the Framework to Other Institutional-Performance Studies

The 2026 edition's methodology contribution (Section 5) claims portability: the framework suite (political seasons + expansion-phase detector + Bayesian-vocabulary black-swan + state-ability synthesis + backtest validation + pre-registration discipline) is applicable to institutional-performance studies in domains beyond the EU Council presidency. This appendix provides a concrete template for a researcher wishing to apply the same approach to a different substantive domain.

When is the framework applicable?

The framework is well-suited to study contexts with all five of the following features:

1. A *rotating or sequential institutional role* held by distinct actors over bounded time-windows (rotation schedules of the EU Council type; G7 presidency rotations; UN Security Council elected-member rotations; chair rotations of inter-governmental committees; cabinet-portfolio rotations in some parliamentary systems).
2. A *performance dimensionality* that can be decomposed into 5–8 substantive categories (analogous to this paper's K_1 – K_5).
3. *Historical baseline data* spanning at least 2–3 decades so archetype means + state-ability matrices are identifiable from the record.
4. *External-shock exposure* sufficient that the Bayesian- vocabulary black-swan framework's three event-type catalogue (security / domestic-institutional / global-economic) is active and applicable.
5. *Some institutional-evolution dimension* such that a political-seasons framework can meaningfully distinguish different phases of the institution's development.

Studies with only 1–2 of these features may benefit from parts of the framework but not the full synthesis; studies with all five are strong candidates for full replication.

Five-step porting procedure

Step 1 — Define the domain's state-ability matrix

For the actors rotating through the institutional role, construct an S/N/W matrix (Strong / Neutral / Weak) over the performance-category axis. Use the historical baseline (Feature 3) to identify structural strong suits. Expect to produce a matrix of size *actors* × *categories*. The EU version of this matrix is Table 25; the analogous matrix for another domain is the first necessary input.

Step 2 — Assign season tags to the domain's history

Identify the 3–6 distinct institutional phases in the domain's history (e.g., for a G7 study: Bretton Woods consolidation; oil crisis; post-Cold-War expansion; post-2008-crisis consolidation; Trump-era fragmentation). Name each phase; assign start / end dates; characterise the dominant agenda. This is the domain-specific season taxonomy. The EU version is Section 4.4; the analogous scheme is derived from the same logic applied to the domain.

Step 3 — Specify the event-type vocabulary

For the domain, identify the 2–4 recurring external-shock classes that re-shape institutional performance. Examples: for a G7 study, global-economic shock, geopolitical crisis, domestic political upheaval. Each event-type gets a documented KPI-delta range from retrospective observation (analogous to the EU paper's +0.6 / −0.5 Security Shock range). The EU version is Section 4.6; the analogous vocabulary is the domain-specific event catalogue.

Step 4 — Run the three-step synthesis procedure

For each actor-in-role-in-window combination, apply the synthesis rule of Section 4.7:

1. Start from archetype mean.
2. Apply expansion-phase adjustment (if applicable).
3. Apply season adjustment.
4. Apply event-type adjustment (if applicable).
5. List Tier-1 strong suits, re-rank by season salience.
6. Attach approach-bias direction.
7. Add bilateral-tension flag effects.
8. Derive audience implications.

The procedure is stable across domains; only the inputs change.

Step 5 — Decide pre-registration scope

Choose a forecast horizon (this paper uses 2026–2030); for each actor-in-window row, author a pre-registered hypothesis with threshold, assumption, and verification source. Commit to a fact-check round date (this paper uses 2028–2029). Document the pre-registration commitment in a verification-framework appendix. The Bayesian-vocabulary ± 0.5 uncertainty band or its domain-specific analogue should be applied explicitly.

Validity protocol: the backtest

Run the full framework on historical cases from the domain to verify internal consistency. This paper’s backtest is a 24-case set (§4.8); a comparable domain should produce at least 12–16 historical cases across the different seasons and event-types. Report Tier-1 directional-accuracy, Tier-2 file-cluster accuracy, KPI-band hit rate, and any systematic biases discovered. Calibration of single-layer effects against backtest residuals produces the domain’s equivalent of this paper’s rules-of-thumb (§4.8 §”Discovered rules of thumb”).

Methodology-substrate: the persona-based verification

The 32-persona reader-simulation stack documented in Appendix E is domain-agnostic; the roster’s per-persona pass criteria translate directly (political scientist → domain expert; AI-in-research practitioner → same; sceptical peer reviewer → same; etc.). A researcher replicating this methodology should adopt the full persona stack for their paper’s verification rather than assemble a bespoke subset; the coverage properties of the full stack are what produce the auditable-trail claim of Section 5.

Minimum viable replication

A replicator with limited resources can adopt parts of the framework:

- **Minimum:** KPI-category system + historical-baseline matrix + at least one pre-registered-hypothesis set. This produces a disciplined single-dimension study (the EU paper’s 2023 edition corresponds roughly to this minimum).
- **Recommended:** add expansion-phase detector + political-season framework + backtest validation. This corresponds to the EU paper’s four-layer synthesis.
- **Full:** add the Mumin Framework Synthesis Method (MFSM) + persona-based reader simulation + companion-artefact system. This corresponds to the 2026 edition at v0.9.6.

The framework does not lose coherence at the **minimum** level; additions increment rigour and reproducibility without changing the underlying logic.

What this template does *not* claim

- It does *not* claim the framework will produce externally- correct forecasts in a new domain without domain-specific calibration. Each step requires domain judgement; Steps 1 + 2 + 3 in particular are the domain-expertise contribution.
- It does *not* claim that a single researcher can replicate the methodology without domain expertise. The portable components are the structure, the discipline, and the persona-verification protocol; the content is necessarily domain-specific.
- It does *not* claim that every institutional-performance study benefits from every layer. A simple study may stop at the minimum viable replication; over-engineering is a risk.

A researcher attempting a replication should read this paper’s Appendix [E](#) for the verification protocol, Section [5](#) for the agentic-workflow context, and `docs/21_abstracted_method.md` (Mumin Framework Synthesis Method) for the full method specification.