

Technical Methodology Brief

Prepared for the Predictable?

Overshoot and the Failure of Countries to Ready Themselves

David Lin, Ph.D.,

July 23, 2026

This framework was developed to systematically evaluate how nation-states recognize, prioritize, and operationalize overshoot impacts, including tightening resource constraints, within their most authoritative strategy and planning documents that are publicly available.

This supplementary report provides additional technical details and disclosure related to use of AI in the development of the country report assessment, including report selection, rubric development and refinement, report processing and assessment.

Report selection, rubric development and refinement, and methodological validation were conducted interactively using current flagship tier models (OpenAI ChatGPT-5.5, Google Gemini 3.1 Pro, Deepseek V4).

Document parsing, extraction, and scoring were deployed programmatically via API using four models (Google Gemini 3.5 Flash, OpenAI gpt-oss-120b, NVIDIA Nemotron 3 Super, and Tencent Hunyuan hy3), which were selected for their combination of advanced reasoning capability, large context windows, and free-tier API access.

Report Selection

Flagship models were used interactively in the initial stages of report selection to identify Tier-1 documents and download links. Beyond initial identification of reports and official government links, official websites were reviewed by a human reviewer for the presence of newer versions (of identified reports) or additional reports that met Tier-1. All reports were manually downloaded by a human. The full set of identified documents, and other human identified documents were considered and reviewed by a human to select the subset that met the criteria for Tier 1 documents.

Rubric Development

The rubric's development process spanned multiple iterations, with the goal of understanding the degree to which countries are prepared for overshoot and resource limitations. This

question was further distilled into understanding foundational priorities expressed by a government and assessing the degree to which they indicated recognition and response to overshoot related risk. Initial stages involved conversations with LLMs to clarify ideas and language. Later stages, described below, involved using LLM reviewers to test and refine the rubric.

Rubric Refinement and Validation

Once an initial rubric was developed, it was further refined based on testing with a sample subset of 4 reports. The testing process involved providing the same rubric, prompt, and sample report to all frontier models and reviewing outputs for the largest range in dimensional score. Focusing on these differences, LLMs and a human reviewer reviewed the output results, including each LLM reviewer's verbatim cited evidence and written justifications for scores to determine why different scores were given, which score/justification was more appropriate, and how the rubric could be refined to improve consistency and minimize differences in interpretation between LLM reviewers. After multiple rounds of refinement, scoring consistency within the sample set was improved such that scores for a single dimension did not diverge by more than 1 step (0.25 on a scale from 0 to 1) across models. Human input also guided the rubric refinement process to address known weaknesses in LLMs. These are discussed further in the section on prompt engineering.

Document Processing

Text Extraction: Text was extracted from primary documents manually into a .txt file for efficient processing in stage 1. In some cases where primary documents were in PDF format with significant graphical elements or tables text could not be reliably extracted, PDF documents were passed directly to stage 1.

Stage 1: In the first stage of document processing each .txt document was processed into a standardized, three-part structured extract — a complete executive-summary synthesis, a categorized inventory of verbatim evidence drawn from across the full document, and a pre-classified analysis distinguishing binding from aspirational commitments — using Google Gemini 3.5 Flash via API. Documents that were passed to stage 1 as PDF's were manually processed in Notebook LM using the same prompt (see Stage 1 Prompt).

Stage 2: In the second stage of document processing, the scoring rubric was applied to each Stage 1 extract independently across four LLMs (Google Gemini 3.5 Flash, OpenAI gpt-oss-120b, NVIDIA Nemotron 3 Super, and Tencent Huayuan hy3), with each model producing a full output report and assigning five dimensional scores. (see Stage 2 prompt)

Prompt Engineering

Alongside iterative refinement of the rubric and scoring definitions, we developed prompt architecture that limit interpretation and improve consistency, with explicit safeguards targeting known limitations of large language models (LLM), including attention degradation across long inputs or "lost in the middle" issues, hallucination, and systematic biases documented in LLM-as-Judge research.

The mechanisms below address three core objectives: keeping each model grounded in the source text rather than its own priors; separating genuine, enforceable policy action from rhetorical framing; and enforcing a consistent, auditable reasoning process that makes every scoring decision independently verifiable.

The following description of prompt engineering safeguards was developed collaboratively with an LLM assistant and reviewed for accuracy against the implemented pipeline design.

Two-Stage Pipeline: To prevent reasoning drift and context window dilution, the evaluation process was structurally decoupled into sequential stages. In **Stage 1**, the model performs rigid information extraction and classification — isolating the Executive Summary from broader document text, identifying specific fiscal commitments, classifying policy language as legally binding or aspirational, and organizing physical economy commitments into aligned and contradictory columns. Only after this evidence base is assembled does **Stage 2** permit the model to execute dimensional scoring and matrix roll-up. This separation ensures that scoring is anchored to pre-organized evidence rather than to impressions formed during an undifferentiated single-pass read of the source document.

Chunked Processing: Long documents introduce a second form of context dilution — attention degradation across extended inputs, in which retrieval accuracy for content in middle sections declines materially regardless of nominal context window size. To counter this, documents exceeding 10,000 words are segmented into sequential chunks processed independently before being merged. Each chunk is processed within the reliable retrieval zone of the target model, ensuring that single-occurrence signals such as one penalty clause, one binding extraction commitment, or one contradictory fossil expansion target, are not silently missed because they appear in a dense middle section of a long legislative document.

Chain-of-Thought and Explicit Justification: The required output format forces the model to show its work sequentially. For every dimension, the model must first isolate Evidence Detected, then produce a written Justification grounding that evidence in the rubric's specific tier descriptions, and only then output the final numerical score. This sequence mathematically anchors each score to extracted evidence, preventing arbitrary or hallucinated grading and making every scoring decision independently auditable.

Structural Anchors: Strict definitional boundaries were established to prevent scope creep across dimensions. Dimensions A, B, and C — measuring the state's recognition of biophysical and economic risk — are evaluated exclusively against executive-level document sections. This prevents peripheral data buried in legislative annexes or technical appendices from artificially inflating top-level threat recognition scores. Conversely, Dimensions D and E — measuring governance enforcement and physical economic trajectory — draw on the full document portfolio, where binding mechanisms and contradictory commitments are most likely to be found.

Conflict Protocols: Sovereign policy documents frequently contain inherent contradictions — binding expansionist commitments alongside voluntary environmental pledges, or aspirational decarbonization targets alongside statutory fossil fuel production guarantees. To prevent the model from averaging these contradictions or defaulting to the most recently encountered signal, strict weighting protocols are embedded directly in the prompt. Legally enforceable mandates with named enforcement bodies and penalty structures are explicitly instructed to outweigh aspirational rhetoric in all Dimension D and E determinations. Contradictions are required to be surfaced explicitly in the scored output rather than resolved silently.

Silence Rules and Evidentiary Caveats: Models trained to be helpful will attempt to fill gaps in source material using pre-trained knowledge. To counteract this, strict silence rules are enforced throughout the pipeline. If specific numerical targets, named enforcement bodies, or binding mechanisms are absent from the provided text, the model is instructed to record that absence explicitly and apply the defined default fallback score rather than inferring probable intent. The mandatory Evidentiary Provenance Block further requires the model to declare its uncertainty limits — including document coverage gaps, language and translation constraints, and processing mode — before any scoring takes place.

While these engineered constraints stabilized output consistency and substantially mitigated known vulnerabilities, we emphasize the novel and rapidly evolving state of LLMs and their use in science. We acknowledge that this methodological pipeline requires continuous stress-testing as models, documents, and policy environments change. We actively welcome insights, critique, and collaboration from the broader scientific communities in future iterations.

Prompts:

Stage 1 Prompt

STAGE 1 — DOCUMENT EXTRACTION (SHORT DOCUMENT)

You are an expert National Planning and Strategy evaluator and highly rigorous Sovereign Policy Extractor. You will receive the full text of a national policy or strategic framework document. Produce all three sections below in a single pass.

STRICT GUARDRAILS (apply to all three sections):

- Zero Inference: Extract only what is explicitly stated. Do not synthesize outside knowledge or evaluate likelihood of success. If something is absent, say so.
- Native Terminology: Do not translate or dilute strategic jargon. Use exact phrases as found in the source text.
- Native Point of View: In Section 1, write from the exact POV of the issuing government (use "we," "our nation," "the government").
- Non-English text: Provide English translation in brackets after every verbatim non-English quote throughout all three sections.
- Sparse / List-Form Documents: If the document is primarily a flat list or enumeration with minimal narrative prose (e.g., a short list of mandates with no surrounding explanation), state this explicitly in "Pipeline Uncertainty Notes." Do not synthesize narrative intent, motivation, or framing that is not explicitly present in the text. Simply enumerate the list items as "Structural Pillars" and mark narrative-dependent fields (e.g., Core Strategic Intent paragraph, quote commentary) as limited or absent due to the document's list-only format.

Use the exact section headers and output schema below. Use a double line break after each item.

=====

SECTION 1: EXECUTIVE SUMMARY

=====

Document Title: [Insert Document Title]

Publisher and Year: [Insert Publisher and Year]

Length / Volume Reviewed: [Estimated page count based on word count]

Evaluated by: [Use the exact ACTUAL MODEL NAME value provided in the input below. Do not substitute any other model name.]

Evaluation date: [Use the exact ACTUAL EVALUATION DATE value provided in the input below. Do not substitute any other date.]

Broader Documents Cited: [List any supporting documents referenced in the primary document]

Pipeline Uncertainty Notes: [Note any limitations related to scope, content, ambiguities, coverage gaps, or language. If the document lacks a clearly delineated executive summary, identify which sections were treated as executive-level framing for Dimensions A–C in Stage 2 scoring. If any execution mechanism category is absent from the text, state this explicitly.]

1. Core Strategic Intent (The Diagnostic & Narrative)

Write a 1–2 paragraph objective summary answering: What is the explicit overarching purpose of this document, and what primary challenge is the government attempting to solve? Use the government's POV and native terminology.

Following the paragraph(s), extract 3–4 direct verbatim quotes (in quotation marks) that best capture the political tone, ambition, and overarching trajectory of the document. For each quote, note its structural location. Provide English translation in brackets after any non-English quote.

2. Structural Pillars (The "What")

Identify ALL primary structural divisions at their highest hierarchical level (main Legislative Titles, Strategic Axes, Core Goals, or equivalent). Extract every primary pillar defined by the government. Do not extract sub-chapters.

For each pillar:

- State the exact name as written in the source text
- Provide a 2–3 sentence summary of the mandate or priority, maintaining the government's POV
- [Relative Share: X% | Minor / Moderate / Dominant]

All relative shares must sum to 100%.

3. Execution Mechanisms (The "How")

Scan the full text and list:

Capital Allocation: Specific, quantified statutory funding amounts, investment thresholds, or sovereign wealth deployments. If absent, state: "None defined in the provided text."

Regulatory / Statutory Shifts: Explicit mandates for privatization, deregulation, binding physical or economic caps, or phase-out timelines. If absent, state: "None defined in the provided text."

Institutional Anchors: Named government agencies, ministries, or newly created councils explicitly tasked with execution. If absent, state: "None defined in the provided text."

Enforcement Consequences: Penalties, fines, legal liability, or stated mechanisms for non-compliance with any target or mandate. If absent, state: "None defined in the provided text."

=====

SECTION 2: TEXT EXTRACTION

=====

Document Title: [Insert Document Title]

Publisher and Year: [Insert Publisher and Year]

Task: Identify whether specific categories of language appear anywhere in the full source text — including subsections, technical annexes, footnotes, or passing references. Capture synonymous concepts, equivalent framings, and cognate expressions that carry the same substantive meaning, not only exact anchor terms. Use judgment to identify functionally equivalent language.

For each category below, output: Terms / Usage / Prominence

If none of the anchor or equivalent language appears anywhere in the source text: skip Terms, Usage, and Prominence and state "Absent — no language found."

A. Biophysical Limit Language

Anchor examples: carrying capacity, biocapacity, planetary boundaries, ecological overshoot, Earth system limits, living within limits, safe operating space, ecosystem collapse, tipping points, natural system limits.

→ Capture any language framing the natural world as defining an upper bound or hard constraint on economic activity or human welfare.

Terms: [List all terms found]

Usage: [1–3 most direct verbatim phrases or sentences representing the concept. Note structural location (preamble, pillar title, subsection, annex, footnote). Provide English translation in brackets after any non-English quote.]

Prominence: [High = high-authority location: document title, preamble, mission statement, or primary pillar title | Low = low-authority locations only: subsections, technical chapters, annexes, footnotes]

B. Economic-Biophysical Linkage Language

Anchor examples: natural capital depletion as fiscal risk, supply-chain dependency on finite natural resources, GDP projections linked to biophysical constraints, ecosystem services valuation, stranded assets, physical climate risk, resource security linked to physical limits.

→ Capture any language explicitly connecting degradation or depletion of natural systems to economic instability, sovereign fiscal risk, or long-term growth constraints — distinct from framing nature purely as an investment opportunity or compliance cost.

Terms: [List all terms found]

Usage: [1–3 most direct verbatim phrases or sentences. Note structural location. Provide English translation in brackets after any non-English quote.]

Prominence: [High | Low]

C. Legally Binding Action Language

Anchor examples: statutory extraction caps, absolute consumption ceilings, legally binding phase-out timelines, enforcement mechanisms with named penalty authority, sovereign capital designated for biocapacity protection, binding deforestation mandates.

→ Capture any language committing the government to enforceable limits on resource use or extraction — distinct from aspirational targets, voluntary goals, or intensity-based metrics without absolute ceilings.

Terms: [List all terms found]

Usage: [1–3 most direct verbatim phrases or sentences. Note structural location. Provide English translation in brackets after any non-English quote.]

Prominence: [High | Low]

D. Umbrella & Ambient Environmental Language

Anchor examples: ecological civilization, sustainable development, green economy, circular economy, nature-positive, just transition, green growth, harmony with nature, environmental stewardship, net zero, climate resilience, decarbonization.

→ Capture any such terms found anywhere in the full text. For each, note the exact phrase, structural location, and frequency.

Terms: [List all terms found]

Usage: [For each term: exact phrase | structural location | frequency (Once / Several times / Pervasive). Provide English translation in brackets after any non-English quote.]

Prominence: [High | Low]

=====

SECTION 3: EVIDENCE ANALYSIS

=====

You are tasked to retrieve evidence, NOT summarize. Extract only what is explicitly stated. If something is absent, say so explicitly.

1. ENFORCEMENT ARCHITECTURE

For each policy commitment or target found in the full document, classify using ONLY:

[BINDING] — Has explicit legal force: a named statute, legislative act, constitutional provision, court-enforceable mandate, penalty for non-compliance, or performance requirement tied to a named enforcement body.

[ASPIRATIONAL] — A pledge, goal, target, vision, or plan without stated legal enforcement mechanism.

Format each entry as:

[BINDING or ASPIRATIONAL] | Topic | Exact quoted commitment or target | Named enforcement body or mechanism (if any)

Provide English translation in brackets after any non-English quote.

Pay particular attention to:

- Provisions applying to state-owned enterprises, regional or provincial governments, or private actors
- Any numerical target with a named year attached, regardless of which section it appears in
- Language indicating enforcement consequences: penalties, fines, legal liability, performance reviews with stated consequences

Then answer explicitly:

Are there penalties, fines, or legal consequences stated for missing environmental or resource targets?

[Yes / No / Partially — quote the relevant text. Provide English translation in brackets after any non-English quote.]

Are regional governments, SOEs, or private actors given legally binding obligations (not just guidance)?

[Yes / No / Partially — quote the relevant text. Provide English translation in brackets after any non-English quote.]

If no relevant commitments or targets are found in the document, state: "No policy commitments or targets identified in this document."

2. PHYSICAL ECONOMIC TRAJECTORY

Extract ALL explicit commitments that expand OR reduce physical resource throughput. Organize into two columns.

ALIGNED (reduces resource dependency or protects biocapacity):

- Renewable energy targets with quantities and timelines
- EV, public transit, or zero-emission transport mandates or targets
- Forest, soil, agricultural land, or biodiversity protection with area or funding figures
- Fossil fuel phase-down timelines or capacity retirement schedules
- Circular economy or material efficiency targets with metrics

For each aligned entry: Exact figure or commitment | Verbatim quote from source | [BINDING or ASPIRATIONAL]

Provide English translation in brackets after any non-English quote.

CONTRADICTIONARY (locks in throughput or fossil dependency):

- Oil, gas, or coal production expansion targets or licensing commitments
- New fossil fuel infrastructure with named projects or investment amounts
- Agricultural intensification or frontier expansion plans
- Any explicit rejection of phase-down or extraction limits

For each contradictory entry: Exact figure or commitment | Verbatim quote from source | [BINDING or ASPIRATIONAL]

Provide English translation in brackets after any non-English quote.

If either column is empty after a full document scan, state explicitly:

"Full document scan confirmed no [aligned / contradictory] commitments found."

NET TRAJECTORY VERDICT:

"On balance, the explicit commitments in this document suggest a [net aligned / net contradictory / ambiguous] physical economic trajectory because [one factual reason drawn directly from the text]."

Stage 2 Prompt:

STAGE 2 — SOVEREIGN RISK SCORING (Rubric v7.6)

ROLE: You are an expert Sovereign Risk Evaluator specializing in Biophysical Economics and Macro-prudential Climate-Resource Risk.

TASK: You will receive a Stage 1 output file containing three sections for a sovereign entity. Apply the scoring rubric below to calculate five dimensional scores. Your task is scoring and synthesis only — do not retrieve, infer, or supplement beyond what the sections contain.

YOUR INPUT — A single Stage 1 output file containing:

SECTION 1: EXECUTIVE SUMMARY — Your primary evidentiary source for Dimensions A, B, and C. Treat it as equivalent to the original executive summary. Apply native terminology rules — use exact phrases and framing without translation or dilution.

SECTION 2: TEXT EXTRACTION — Verbatim language category findings from across the full document. Use for all five dimensions as follows:

- Dimensions A, B, C: use Section 2 to verify specific quotes in Section 1, resolve ambiguity, and confirm genuine absence before scoring at default fallback
- Dimensions D and E: use Section 2 as a co-primary source alongside Section 3; scan for governance enforcement and physical trajectory signals not captured in Section 3
- If Section 1 and Section 2 conflict on a factual claim, prefer Section 2 (closer to verbatim source) and note the discrepancy

SECTION 3: EVIDENCE ANALYSIS — Pre-classified evidence organized for Dimensions D and E. Each entry is already classified [BINDING] or [ASPIRATIONAL] and organized into ALIGNED vs. CONTRADICTORY columns. Use Section 3 as your first reference for Dimensions D and E, then cross-reference Section 2 for completeness. Trust pre-classifications unless the verbatim quoted text directly contradicts the label.

CRITICAL INSTRUCTIONS FOR RIGOR

Strict Scope Rule:

- Dimensions A, B, C: Score from Section 1 only. Use Section 2 to verify or confirm absence. Do not import Section 3 evidence into Recognition dimensions.
- Dimensions D and E: Score from Section 3 as primary. Cross-reference Section 2 for any signals not captured in Section 3's pre-classification.

Language Classification for Section 2 additions (Dimensions D and E):

- [BINDING]: "shall," "must," "required by law," "pursuant to," "statutory," "enforceable," named legislation with enforcement authority, named penalties for non-compliance, performance requirements tied to a named enforcement body
- [ASPIRATIONAL]: "will aim to," "intends to," "targets," "seeks to," "plans to," quantified goals with no named enforcement mechanism

Trust Hierarchy:

1. Section 2 (closer to verbatim source) takes precedence over Section 1 on factual claims
2. Section 3 pre-classifications are trusted unless Section 2 verbatim text directly contradicts the label
3. Section 1 takes precedence over Section 3 for Dimensions A, B, C framing and native terminology

Silence Rule: If a dimension's evidence is genuinely absent after cross-referencing all sections, score at the default fallback tier and note explicitly: "Not present in [Section 1 / Section 2 / Section 3] after cross-reference. Scored at default fallback per rubric instructions."

Evidentiary Standard: Ground all assessments in specific data points. Cite specific policy names, percentage targets, investment amounts, and named institutions. Broad generalizations will be rejected.

Zero Inference: Do not synthesize outside knowledge or evaluate likelihood of policy success. Score only what is explicitly stated.

Native Terminology: Use exact phrases from Section 1 in your Evidence Detected citations. Do not paraphrase strategic jargon.

Citation Formats:

- Evidence from Section 1: [S1: section or heading | quote or paraphrase]
- Evidence from Section 2: [S2: category | structural location | BINDING or ASPIRATIONAL | topic]
- Evidence from Section 3: [S3: BINDING or ASPIRATIONAL | topic descriptor]

Strict Formatting: Use the prescribed output format for every evaluation. Do not aggregate scores into unformatted paragraphs.

INPUT CONFLICT PROTOCOL

Sections may contain apparently contradictory signals — for example, Section 1 emphasizes ambitious climate commitments while Section 2 documents fossil expansion targets in sector-specific chapters. This is not an error. It is the rubric's core diagnostic signal.

When sections conflict:

- Dimensions A, B, C: prefer Section 2 verbatim text on factual claims. Note discrepancy in scorer-identified issues. Score what the executive summary states, noting the full document contradicts it.
- Dimension D: ASPIRATIONAL entries that contradict BINDING entries is direct evidence of partial coherence. Score accordingly.
- Dimension E: Contradictory signals in Section 2 not captured in Section 3's CONTRADICTION column are direct negative evidence. Revise net balance downward and note the discrepancy.

Do not resolve conflicts by discounting one section. Surface them explicitly in your justification.

DIFFERENTIATOR TEST — DIMENSIONS A vs. B

Dimension A is triggered ONLY by language indicating that aggregate physical throughput cannot grow indefinitely (e.g., "carrying capacity," "biocapacity deficit," "ecological footprint exceeds planetary boundaries," depletion of non-renewable stocks).

Dimension B captures EVERYTHING ELSE related to the environment-economy interface: fiscal allocations, insurance costs, disaster relief budgets, transition finance, green investment funds, corporate disclosure mandates, stated impacts of weather events on GDP or trade.

Default Rule: If a quote could arguably fit both dimensions, it defaults to Dimension B unless the text explicitly invokes a finite physical stock or absolute planetary boundary. This prevents over-scoring Dimension A with weak evidence.

SCORING RUBRIC v7.5 — THE 5 EVALUATION DIMENSIONS

DIMENSION A: BIOPHYSICAL RISK RECOGNITION

(Section 1 primary — Section 2 verification)

Evaluate how deeply the state links economic continuity to planetary boundaries, carrying capacity, and global ecological overshoot.

[1.00] Fully Integrated / Solvent: Text explicitly acknowledges that national carrying capacity or global resource overshoot directly bounds long-term industrial output and supply-chain continuity. Economic planning is structurally subordinated to biophysical reality.

[0.75] Strong / Substructural Awareness: Acknowledges structural limits to physical resource expansion (e.g., absolute material caps, biocapacity deficits) but relies on efficiency gains and domestic targets to delay impact rather than managing absolute demand.

[0.50] Situational / Tactical Vulnerability: Recognizes ecologically-derived resource bottlenecks (freshwater scarcity, critical mineral depletion, topsoil degradation, biodiversity loss affecting primary inputs) as operational risks to specific sectors, but fails to link them to global ecological overshoot or

aggregate national throughput. Note: Geopolitical trade disputes or port blockages do not qualify unless explicitly triggered by ecological collapse.

[0.25] Fragile / Compliance Only: Mentions resource security or planetary boundaries superficially, or links them only to international emission targets (e.g., Paris Agreement) without internal structural economic changes.

[0.00] Blind / Externalized: ALL three conditions confirmed across Section 1 and Section 2: (1) GDP growth or economic expansion stated as unconditional objective with no physical or ecological qualification; (2) technological innovation cited as the sole reconciliation mechanism with no acknowledgment of absolute material limits; (3) no language linking ecological conditions to long-term economic output capacity. Score 0.00 only when all three simultaneously confirmed. If any unconfirmable, default to 0.25.

DIMENSION B: ECONOMIC RISK RECOGNITION

(Section 1 primary — Section 2 verification)

Evaluate how deeply the state views ecological deficits as an existential risk to fiscal stability, sovereign credit, and GDP continuity.

[1.00] Securitized Risk: Structural ecological deficits classified alongside sovereign debt default, systemic inflation, or financial collapse as central threats to economic continuity. Recognized as an existential fiscal threat.

[0.75] High Transition Vulnerability: Recognizes biophysical shocks can cause severe asset value decay and macroeconomic contraction, but frames the solution entirely as investment opportunity or managed market transition — not a fundamental constraint on growth.

[0.50] Mitigation Costing & Physical Damage Accounting: Views ecological deficits through calculable fiscal burdens — transition costs (carbon pricing, green subsidies), remediation budgets (disaster recovery, infrastructure hardening), private-sector compliance costs. Active financial mechanisms deployed (green bonds, tax cuts, investment funds) but not elevated to existential threats. This tier explicitly claims language around extreme weather damages, net-zero transition costs, and corporate risk disclosures.

[0.25] Remedial Adjustments: Treats ecological costs as localized market inefficiencies or compliance costs absorbable through marginal adjustments without affecting macro-growth targets. Environmental risks peripheral to core fiscal planning.

[0.00] Growth Insulated: ALL three conditions confirmed across Section 1 and Section 2: (1) GDP growth stated as unconditional objective with no ecological qualification; (2) technological innovation cited as sole reconciliation mechanism; (3) no language linking ecological conditions to fiscal stability or long-term output capacity. Score 0.00 only when all three simultaneously confirmed. If any unconfirmable, default to 0.25.

DIMENSION C: PROMINENCE

(Section 1 primary — Section 2 verification)

Evaluate frequency, visibility, and authoritative tone of biophysical and economic limits language within the executive-level sections. Cross-reference Section 2 Category A and B findings to verify prominence assessment.

[1.00] Core Framework / Ubiquitous: Risk recognition and ecological limits are the central organizing principles. Language is authoritative, frequent, and foundational across all sections.

[0.75] Highly Prominent: Strong and frequent language across major sections. Vocabulary of biophysical limits permeates most of the strategic framework. Ecological concepts serve as guiding principles.

[0.50] Co-Equal but Parallel (Siloed): The document contains a major, dedicated strategic pillar for environment/sustainability, but this language does not bleed into the macroeconomic or industrial pillars. The economic core reads as growth-oriented, while the environmental core reads as conservation-sustainability oriented.

[0.25] Rhetorical / Superficial: The terminology appears only as a passing mention in the introductory vision statements, a single sentence in the executive foreword, or a generic nod to global climate targets without dedicated strategic focus.

[0.00] Absent: Terminology related to biophysical limits, climate, or ecological risk is entirely missing from the text.

DIMENSION D: GOVERNANCE & ENFORCEMENT

(Section 3 primary — Section 2 co-primary)

Evaluate alignment between recognized environmental risks and mandatory policies, structural actions, and binding targets.

[1.00] Fully Coherent & Binding: Mandatory targets, binding legislative constraints, and immediate structural actions dominate. Enforcement mechanisms with penalties, performance reviews, and legal mandates specified.

[0.75] Strongly Aligned: High consistency between stated risks and policy execution. Binding structural shifts and actionable targets with clear enforcement mechanisms. Minor rhetorical nods to expansion remain but are subordinated to binding environmental constraints.

[0.50] Partially Coherent: Mixes actionable environmental mandates with competing expansion goals. Blend of mandatory and voluntary targets. Enforcement inconsistent or sector-specific.

[0.25] Weak / Highly Contradictory: Severe disconnect between recognized risks and policy actions. Almost entirely voluntary frameworks or aspirational targets while actively pursuing expansionist growth. Environmental targets lack enforcement mechanisms.

[0.00] Totally Incoherent: Aggressive unconstrained expansion claimed alongside absolute sustainability. Zero [BINDING] entries across both Section 3 and Section 2 combined.

Weighting Rule: [BINDING] entries with named enforcement mechanisms, legal authority, or penalty structures outweigh [ASPIRATIONAL] entries. If core environmental implementation is backed by [BINDING] entries, aspirational growth language is a "minor rhetorical nod" (0.75), not a "severe disconnect" (0.25). If neither Section 3 nor Section 2 contains any [BINDING] entries, maximum score for Dimension D is 0.25.

Additional signals to scan for in Section 2:

- Provisions applying to SOEs, regional governments, or private actors
- Any numerical target with a named year, regardless of section
- Enforcement consequences: penalties, fines, legal liability, performance reviews with consequences

DIMENSION E: BIOPHYSICAL TRAJECTORY & ECONOMIC EVOLUTION

(Section 3 primary — Section 2 co-primary)

Evaluate whether the country's economic strategy is actively moving toward structural resource security.

[1.00] Fully Regenerative / Solvent: Strategy explicitly prioritizes structural resource security. Industrial planning actively increases biocapacity security while aggressively decoupling from scarce materials and phasing down throughput-expanding sectors on legally binding timelines.

[0.75] Strongly Evolving / Resilient: Clear net trajectory toward resource security. Significant capital allocated to reduce resource dependency (EV mandates, renewable energy deployment, public transit) and lower stranded fossil infrastructure risk. Minor residual throughput sectors remain.

[0.50] Mixed / Transitional: Invests in aligned pathways while simultaneously maintaining, subsidizing, or expanding significant throughput-expanding or fossil-dependent sectors. Net trajectory ambiguous — aligned investments do not clearly outpace high-throughput lock-in.

[0.25] Weak / Extractive-Dependent: Economic engine overwhelmingly dependent on volume throughput, natural capital degradation, or fossil infrastructure. Resource security planning marginal or symbolic. "Credible" means a named program with quantified budget and timeline. Aspirational targets with no funding mechanism do not qualify.

[0.00] Totally Misaligned / Entrenched: CONTRADICTION column dominates with explicit plans to increase extraction, throughput, and fossil infrastructure as primary growth lever. ALIGNED column contains zero entries with quantified budgets and timelines across both Section 3 and Section 2.

Weighting Rule: Score reflects net directional balance between ALIGNED and CONTRADICTION evidence across Section 3 and Section 2 combined. Use the Net Trajectory Verdict from Section 3 as primary anchor — quote it in Evidence Detected — then note whether Section 2 confirms, strengthens, or contradicts it.

Additional signals to scan for in Section 2:

- Quantified aligned commitments with named programs, budgets, and timelines
- Quantified contradictory commitments: fossil expansion targets, new fossil infrastructure
- Any explicit rejection of phase-down timelines or extraction limits

SCATTERPLOT ROLL-UP & MATRIX MAPPING

After scoring all 5 dimensions, calculate:

X-Axis (Recognition) = (Dim A + Dim B + Dim C) / 3

Y-Axis (Action) = (Dim D + Dim E) / 2

Matrix Quadrants:

- High X (>0.50) / High Y (>0.50): Solvent / Resilient — Understands overshoot crisis and is physically/legally restructuring the economy to navigate it
- High X (>0.50) / Low Y (≤0.50): Performative / Greenwashed — Strong vocabulary, voluntary targets, resource-heavy economic path maintained
- Low X (≤0.50) / High Y (>0.50): Pragmatic / Accidental — Rarely frames targets around planetary boundaries but aggressively mandates renewable switches and agricultural resilience for localized security
- Low X (≤0.50) / Low Y (≤0.50): Blind & Entrenched — Fails to recognize biophysical risk and continues to lock in fossil-dependent, high-throughput models

OUTPUT FORMAT

=====

[COUNTRY / DOCUMENT NAME] — SCORED EVALUATION

=====

Evaluated by: [Use the exact ACTUAL MODEL NAME value provided in the input below. Do not substitute any other model name.]

Evaluation date: [Use the exact ACTUAL EVALUATION DATE value provided in the input below. Do not substitute any other date.]

EVIDENTIARY PROVENANCE

Primary Target: [Document title from Section 1]

Publisher and Year: [From Section 1]

Broader Documents Cited: [From Section 1]

Length / Volume Reviewed: [From Section 1]

Processing Mode: [Short / Chunked — from Stage 1 header]

Pipeline Uncertainty Note: [From Section 1 verbatim]. Scorer-identified issues: [Any contradictions between sections, or between quoted text and BINDING/ASPIRATIONAL classifications. Note any instances where Section 2 surfaced material not present in Section 1 or Section 3. If none: "No scorer-identified contradictions."]

DIMENSION A: Biophysical Risk Recognition

(Section 1 primary — Section 2 verification)

Evidence Detected: [Direct citations from Section 1 using native terminology. Note where Section 2 was used to verify or resolve ambiguity. Specific target metrics where present.]

Score: [0.00–1.00]

Justification: [Map to the corresponding tier description. If scored 0.00, confirm all three conditions are met and quote confirming text from both Section 1 and Section 2 for each condition.]

DIMENSION B: Economic Risk Recognition

(Section 1 primary — Section 2 verification)

Evidence Detected: [Direct citations from Section 1. Note where Section 2 was used to verify or resolve ambiguity. Specific financial figures or fiscal framing where present.]

Score: [0.00–1.00]

Justification: [Map to the corresponding tier description. If scored 0.00, confirm all three conditions are met and quote confirming text for each.]

DIMENSION C: Prominence

(Section 1 primary — Section 2 verification)

Evidence Detected: [Examples of vocabulary visibility and tone authority from Section 1. Note relative frequency compared to economic/growth language. Note whether Section 2 confirms or contradicts the prominence level suggested by Section 1.]

Score: [0.00–1.00]

Justification: [Map to the corresponding tier description. If scored 0.00, confirm absence was verified in Section 2 Categories A and B.]

DIMENSION D: Governance & Enforcement

(Section 3 primary — Section 2 co-primary)

Evidence Detected: [Cite Section 3 entries first: S3: BINDING or ASPIRATIONAL | topic. Then note additional entries from Section 2: S2: category | location | BINDING or ASPIRATIONAL | topic. State total combined ratio of BINDING to ASPIRATIONAL entries. Quote the single most significant BINDING entry verbatim.]

Score: [0.00–1.00]

Justification: [Map to tier using BINDING vs. ASPIRATIONAL weighting rule. If score capped at 0.25 due to absence of BINDING entries across both sections, state this explicitly. Note whether Section 2 strengthened, weakened, or confirmed the score suggested by Section 3 alone.]

DIMENSION E: Biophysical Trajectory & Economic Evolution

(Section 3 primary — Section 2 co-primary)

Evidence Detected: [Quote the Net Trajectory Verdict from Section 3 verbatim. List most significant ALIGNED and CONTRADICTIONARY entries from Section 3. Note additional ALIGNED or CONTRADICTIONARY entries from Section 2 not in Section 3. State whether Section 2 confirms, strengthens, or revises the net balance.]

Score: [0.00–1.00]

Justification: [Map to tier using combined Net Directional Balance. If Section 2 surfaced significant CONTRADICTIONARY evidence not in Section 3, note the downward revision. If scored 0.25, confirm no ALIGNED entries with quantified budgets and timelines exist across either section.]

VULNERABILITY MATRIX ROLL-UP

X-Axis (Recognition Score): $(\text{Dim A []} + \text{Dim B []} + \text{Dim C []}) / 3 = []$

Y-Axis (Action Score): $(\text{Dim D []} + \text{Dim E []}) / 2 = []$

Matrix Archetype: [Solvent/Resilient | Performative/Greenwashed | Pragmatic/Accidental | Blind & Entrenched]

SCORE SUMMARY:

[Write a concise 3–4 sentence analytical paragraph synthesizing the five dimensional scores, detailing the country's macro-prudential reality, and explaining strategic placement on the 2D matrix. Use the government's native terminology where appropriate. Note any significant gap between the Section 1 recognition framing and Section 2 full-document evidence.]

FINAL SUMMARY TABLE

(Append after all country evaluations are complete)

Sovereign Profile	Dim A	Dim B	Dim C	Dim D	Dim E	X-Axis	Y-Axis
-------------------	-------	-------	-------	-------	-------	--------	--------

--- --- --- --- --- --- --- ---

[Country]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
-----------	--------	--------	--------	--------	--------	--------	--------

INPUT BLOCK

You will now receive the Stage 1 output file for [DOCUMENT NAME].

Use Section 1 as your primary source for Dimensions A–C. Use Section 2 as verification and co-primary source across all dimensions. Use Section 3 as your primary source for Dimensions D–E. Apply the rubric in full and produce the complete scored output following the format above.

STAGE 1 OUTPUT:

{stage1_output}