Political OS Test Suite — Evaluation Cases

Test cases for validating the four Political OS constraint specifications

---

Purpose

This document provides a set of test cases to run through each Political OS (Classical Liberal, Marxist, Critical Justice, Theocratic). Each case should be evaluated mechanically using the OS's evaluation algorithm. The test suite is designed to:

1. **Differentiate** — expose cases where the OS produce divergent classifications
2. **Stress-test** — probe edge cases where an OS's axioms are under tension
3. **Cross ideological lines** — include policies from across the political spectrum
4. **Validate consistency** — confirm each OS produces internally coherent analysis

How to Run the Tests

Prerequisites

You need: - Access to an LLM (Claude, GPT-4, etc.) - The following documents from this repository:

Document	Purpose	When to Load
`classical_liberal_political_os.md`	Abstract OS specification	Methods 1-5: one OS per session
`marxist_political_os.md`	OS under test	Methods 1-4: one OS per session
`critical_justice_political_os.md`	OS under test	Methods 1-4: one OS per session; Method 6: as diagnostic program
`theocratic_political_os.md`	OS under test	Methods 1-4: one OS per session; Method 6: as host OS

Document	Purpose	When to Load
`us_democratic_political_os.md`	Implementation specification (US system)	Method 5: paired with Classical Liberal OS
`political_os_test_suite.md` (this document)	Test case descriptions	Reference only — copy/paste individual test cases into sessions
`comparative_political_os_analysis.md` (The Governance Stack)	Layered governance model and cross-OS structural analysis	Method 3 only — load instead of an OS document

Critical rule: Each LLM session gets **exactly one OS document**. Never load two OS documents in the same session — the LLM will blend their constraints and produce invalid results. The test cases from this document are pasted in as prompts, not loaded as context.

Do NOT load the parent ontology (`constraint_emergence_ontology.md`) or the Logical Encapsulation template (`ontology_templates.md`). The OS documents are self-contained constraint specifications — that is the point of Logical Encapsulation. Loading the parent ontology would invite the LLM to reason about the meta-framework (constraint manifolds, Markov objects, emergent properties) rather than mechanically applying the OS's evaluation algorithm. The ontology explains *why* the method works; the OS documents are *what* gets tested.

Method 1: Single OS, Single Test (Basic)

The simplest approach. Use this to familiarize yourself with the method.

Context window contains: One OS document (e.g., `classical_liberal_political_os.md`) — nothing else.

1. **Start a fresh LLM session** (no prior context — this is critical)
2. **Paste the entire contents** of one Political OS document (e.g., `classical_liberal_political_os.md`)
3. **Paste a test case description** from the Test Cases section below (just the Description paragraph — not the "Why this case" or "Expected divergence")
4. **Prompt**:

> Evaluate this using the framework's evaluation algorithm exactly as defined in the document. Follow every step of the algorithm in order — including any pre-evaluation steps. For each step, show your reasoning. If the algorithm produces a system state classification, report it using the framework's state taxonomy. If the algorithm terminates before reaching classification, report where and why it terminated.
>
> Use ONLY the framework provided. Do not introduce external assumptions.

5. **Record the result** using the Scoring Template at the end of this document

Method 2: Single OS, Full Suite (Systematic)

Run all 15 test cases through one OS in a single session.

Context window contains: One OS document — nothing else. Test cases are pasted as follow-up messages.

1. Start a fresh LLM session
2. Paste the entire Political OS document
3. Prompt:

> I will now give you a series of test cases. For each one, evaluate it using the framework's evaluation algorithm exactly as defined in the document. Follow every step in order — including any pre-evaluation steps. If the algorithm produces a system state classification, report it. If the algorithm terminates before reaching classification, report where and why it terminated. Be mechanical — follow the algorithm, do not editorialize.

4. Paste each test case one at a time
5. Record results in the Scoring Template after each case
6. After all 15, prompt:

> Review your results for all 15 cases. Are they internally consistent with the framework's axioms? If any case terminated before classification (e.g., at a pre-evaluation step), is the termination justified by the algorithm? Flag any case where you may have departed from mechanical application.

7. Repeat with each of the other three OS documents (in fresh sessions)

Method 3: Comparative Analysis (Full Validation)

After running Method 2 for all four OS.

Context window contains: `comparative_political_os_analysis.md` + your completed scoring tables + the Predicted Results Matrix from this document. Do NOT load any individual OS document — the comparative document already contains the structural comparison.

1. Start a fresh LLM session
2. Paste the `comparative_political_os_analysis.md` document
3. Paste your completed scoring tables (all four OS, all 15 tests)
4. Paste the **Predicted Results Matrix** from this document (the comparative table and summary statistics)
5. Prompt:

> Analyze the test results across the four OS: 1. Where do all four OS agree? What does the agreement reveal? 2. Where do they maximally diverge? What structural differences explain the divergence? 3. Did any OS produce a classification that appears inconsistent with its own axioms? If so, flag it — this indicates a defect in the OS document, not a wrong answer. 4. Compare the actual results to the Predicted Results Matrix. Where do the predictions hold? Where do they fail? What does any failure reveal?

Method 4: Head-to-Head (Quick Divergence Test)

For quick validation, pick a maximum-divergence case and run it through two opposing OS.

Context window contains: One OS document per session. You run **two separate sessions** (one per OS), then compare the outputs side by side yourself.

Recommended pairs:

Test Case	OS Pair	Expected Contrast
Test 7: Blasphemy Laws	Liberal vs Theocratic	Strongly negative vs strongly positive
Test 10: DEI Statements	Liberal vs Critical Justice	Strongly negative (Crisis) vs strongly positive
Test 3: Gender-Affirming Care	Critical Justice vs Theocratic	Strongly positive vs strongly negative
Test 5: Affirmative Action	Liberal vs Critical Justice	Positive (no invariant violated) vs strongly positive
Test 14: Reparations	Critical Justice vs Marxist	Strongly positive vs mixed (identity vs class)

Run each OS in a separate session, then compare side by side.

Method 5: Specification vs Implementation (Gap Analysis)

Tests the gap between the abstract Classical Liberal OS and the concrete US Constitutional implementation.

Context window: Run in **two separate sessions**: - Session A: `classical_liberal_political_os.md` (the abstract specification) - Session B: `us_democratic_political_os.md` (the US implementation)

1. Choose a test case and run it through both documents in separate sessions
2. Compare the outputs. Look for:

Signal	What it reveals
Both produce the same classification	The US implementation fully covers this case
Abstract OS flags a violation, US OS does not	The US implementation has a **gap** — the invariant is not hardened at constitutional level for this case
US OS distinguishes constitutional vs program-layer	The US implementation reveals **where corruption can enter** — the invariant is formally intact but program-dependent

Signal	What it reveals
US OS identifies the relevant constitutional mechanism	Validates that the implementation maps to the specification

Recommended test cases for Method 5:

Test Case	Expected Gap
Test 2: Citizens United	Abstract OS flags Information + Revocability degradation; US OS reveals this is a Court interpretive choice (which invariant to prioritize)
Test 10: DEI Statements	Abstract OS flags Agency + Information; US OS should show 1st Amendment protects against government mandates but not private institutional mandates
Test 13: Platform Moderation	Abstract OS flags Information if monopolistic; US OS should show 1st Amendment constrains only government — platform moderation is in the program-layer gap
Test 9: Surveillance + Oversight	Abstract OS flags Agency + Information; US OS should show 4th Amendment is hardened but oversight mechanisms are program-layer

This method validates the US Democratic OS document itself — does it correctly identify where the abstract invariants are constitutionally hardened and where they depend on programs?

Method 6: Hosted Execution (Fragment Testing)

Tests the framework's claim that some candidates (CJ, Marxist) are diagnostic programs rather than full operating systems, and that they function differently when hosted on another OS versus running standalone.

Context window contains: One host OS document + one diagnostic program document. This is the **only method** that loads two Political OS documents in the same session — deliberately, because the test is about the hosted relationship.

1. Start a fresh LLM session
2. Paste the **host OS** document (e.g., `classical_liberal_political_os.md`)
3. Then paste the **diagnostic program** document (e.g., `critical_justice_political_os.md`)
4. Prompt:

> The first document is the operating system — its invariants and evaluation algorithm govern. The second document is a diagnostic program running on that OS. It can surface evidence, identify patterns, and flag concerns, but the host OS's invariants and evaluation algorithm determine the final classification.
>
> Evaluate this test case. Show: (1) what the diagnostic program identifies, (2) how the host OS's evaluation algorithm processes that input, (3) the final classification under the host OS.

5. Compare results against:
   - The standalone host OS result (from Method 1/2)
   - The standalone diagnostic program result (from Method 1/2)

What to look for:

Signal	What it reveals
Diagnostic findings are absorbed without changing host classification	The host OS already handles this case; the diagnostic adds detail but not new classification
Diagnostic findings shift the host classification	The diagnostic surfaced evidence the host OS's algorithm is sensitive to — the hosted execution produces richer analysis
Diagnostic findings conflict with host invariants	The diagnostic's recommendations would violate the host OS's constraints — this is where fragment status becomes visible
Hosted result matches standalone diagnostic, ignoring host	The LLM is blending rather than hosting — session contamination, rerun with clearer prompting

Recommended test cases for Method 6:

Test Case	Host OS	Diagnostic	Why
Test 10: DEI Mandatory Statements	Liberal	Critical Justice	Maximum divergence: Liberal classifies Crisis (compelled speech), CJ classifies strongly positive. Does CJ's structural analysis change the Liberal evaluation,

Test Case	Host OS	Diagnostic	Why
			or does Agency (1.1) hold?
Test 5: Affirmative Action	Liberal	Critical Justice	Compatible case: Liberal classifies positive (no invariant violated), CJ classifies strongly positive. Does hosting produce richer analysis without conflict?
Test 14: Reparations	Liberal	Marxist	Cross-diagnostic: Marxist class analysis as input to Liberal evaluation. Does the class lens surface evidence the Liberal algorithm is sensitive to?
Test 8: School Choice	Theocratic	Critical Justice	Alien host: CJ running on Theocratic OS. What happens when the diagnostic's assumptions about structural power meet a

Test Case	Host OS	Diagnostic	Why
			framework grounded in divine order?

This method tests whether the completeness-as-discovery finding (see The Governance Stack) holds under execution — do fragments actually behave as fragments when hosted?

Important: Session Hygiene

- **Fresh sessions**: Always start a new session for each OS. If you load two OS documents in the same session, the LLM will blend them. The whole point is to test each constraint manifold independently.
- **No priming**: Do not discuss the test case or your expectations before pasting the OS document. The LLM should reason from the constraints, not from your framing.
- **No coaching**: If the LLM produces an unexpected result, record it. Do not re-prompt to get the "right" answer. Unexpected results are the most valuable — they reveal either a defect in the OS document or a failure of the prediction.
- **Same model**: Use the same LLM model across all four OS for a given test case. Different models may have different baseline biases, which would confound the comparison.

What to Look For

The test is validating two things simultaneously:

1. **The OS documents** — Do they produce consistent, mechanically derivable analyses? If an OS produces an analysis that contradicts its own axioms, the OS document has a bug and needs revision.

2. **The Logical Encapsulation method** — Does loading different constraint specifications actually produce different analyses from the same LLM? If all four OS produce the same analysis regardless of framework, the method doesn't work. If they produce systematically different analyses traceable to their different axioms, the method is validated.

Specific signals to watch for: - LLM breaks character and introduces values not in the OS → weak constraint specification - LLM produces identical analysis across different OS → method failure - LLM's analysis contradicts the OS's own axioms → document needs revision - LLM finds genuine tensions within an OS that the document doesn't address → new open problem discovered - Predicted results matrix matches actual results → predictions and OS documents are aligned - Predicted results matrix doesn't match → either prediction was wrong or OS document needs tightening

Recording Results

Use the Scoring Template provided at the end of this document. For each test case, fill in one row per OS:

OS	Result (Classification or Termination)	Key Invariant(s) Affected	Confidence	Notes
Classical Liberal	e.g., Strained	1.1 Agency	High	Agency degraded by coercion; other invariants intact

After completing all tests, compare your actual results against the Predicted Results Matrix to validate both the predictions and the OS documents.

Test Cases

Test 1: Universal Basic Income

Description: A government program providing every citizen an unconditional monthly cash payment sufficient for basic needs, funded through taxation of high earners and corporate profits.

Why this case: Tests the boundary between OS-level and program-level analysis. The Liberal OS should evaluate whether UBI affects invariants (agency, alternatives) without endorsing or rejecting it ideologically. The Marxist OS must determine whether this is reformism that delays revolution or material improvement that advances consciousness. The Critical Justice OS must evaluate structural effects on identity groups. The Theocratic OS must evaluate against sacred order.

Expected divergence: Liberal and Marxist should diverge on whether this is stabilizing or co-optive. Critical Justice should focus on differential impact by identity group. Theocratic should evaluate based on alignment with charitable obligations.

Test 2: Citizens United v. FEC (Corporate Political Spending as Free Speech)

Description: A legal ruling that corporations have the right to spend unlimited money on political advertising, on the grounds that spending is a form of speech protected by the constitution.

Why this case: Directly tests the Liberal OS's own internal tension — does this protect Information (1.2) or degrade it? Tests the provenance principle (money as alternative power source). Forces the Marxist OS to analyze class power operating through liberal mechanisms. Forces the Critical Justice OS to evaluate a liberal institution. Forces the Theocratic OS to evaluate secular governance.

Expected divergence: Liberal, Marxist, and Critical Justice should find problems but for entirely different reasons. The Theocratic OS has weak purchase on this (secular governance question).

Test 3: Gender-Affirming Care for Minors

Description: Medical provision of puberty blockers, hormone therapy, and in some cases surgical interventions for minors diagnosed with gender dysphoria, with parental consent and medical oversight.

Why this case: Maximum divergence case. The Liberal OS must balance Agency (1.1) with the question of whether minors can consent. The Marxist OS has weak purchase on this (not primarily a class issue). The Critical Justice OS should strongly support as dismantling gender normativity. The Theocratic OS should strongly oppose as violation of sacred order. Tests what each OS does with cases outside its primary analytical domain.

Expected divergence: Four radically different analyses. Reveals each OS's blind spots.

Test 4: China's Social Credit System

Description: A state-operated system that tracks citizen behavior (financial, social, legal) and assigns a score affecting access to services, travel, housing, and employment.

Why this case: Should produce strong convergence — all four OS should classify this negatively, but through different invariants. Tests whether the OS produce the same output for genuinely authoritarian policy.

Expected divergence: Three OS converge on strongly negative (Liberal, Marxist, CJ) but for different reasons. The Theocratic OS diverges — its evaluation depends on whether the state serves divine or secular purposes, producing a mixed rather than negative classification.

Test 5: Affirmative Action in University Admissions

Description: A university admissions policy that considers race as one factor among many in admissions decisions, with the stated goal of increasing campus diversity.

Why this case: Direct collision between Liberal OS (formal equality, agency, alternatives) and Critical Justice OS (structural reproduction, power asymmetry). The Marxist OS must determine whether race-based policy obscures or addresses class dynamics. The Theocratic OS must evaluate whether group-based preference aligns with sacred order.

Expected divergence: The Liberal OS faces a subtle test — mechanical application against its invariants (Agency, Information, Alternatives, Revocability) may find no violations, classifying this as a legitimate program. The CJ OS should classify as strongly positive. If so, they agree directionally but for structurally different reasons (Liberal: "no

OS violation" vs CJ: "advances equity"). The real divergence may be in *degree* rather than direction. Marxist should be internally conflicted (identity vs class). Theocratic should be orthogonal (weak purchase).

Test 6: COVID Vaccine Mandates

Description: A government mandate requiring all adults to receive a COVID-19 vaccination as a condition of employment in certain sectors, with medical exemptions.

Why this case: Tests the Liberal OS's handling of public health vs individual agency. Tests the Marxist OS on state power exercised during crisis. Tests the Critical Justice OS on differential trust in medical institutions across racial groups. Tests the Theocratic OS on bodily autonomy vs collective obligation.

Expected divergence: Liberal OS should find invariant tension (Agency vs public welfare — but public welfare is program-level, not OS-level). Critical Justice should produce complex analysis (structural distrust of medical institutions in communities of color vs protection of vulnerable populations).

Test 7: Blasphemy Laws

Description: A national law criminalizing public speech that insults, mocks, or denigrates religious beliefs, with penalties including fines and imprisonment.

Why this case: Should produce maximum divergence between Theocratic and Liberal OS. The Theocratic OS should classify this as maintaining sacred order. The Liberal OS should classify it as degrading Information (1.2) and Agency (1.1). Tests whether the Marxist OS treats religion as ideological superstructure. Tests whether the Critical Justice OS evaluates this as protecting marginalized religious communities or suppressing dissent.

Expected divergence: Theocratic supports, Liberal opposes, Marxist opposes (religion as ideology), Critical Justice conflicted (protecting marginalized vs constraining speech).

Test 8: Worker-Owned Cooperative Economy

Description: A policy framework converting major industries to worker-owned cooperatives through a combination of tax incentives, right-of-first-refusal laws when businesses are sold, and public banking to fund worker buyouts.

Why this case: Should produce convergence between Marxist and Liberal OS (surprisingly). The Marxist OS should classify as progressive (workers control means of production). The Liberal OS should find no invariant violations if participation is voluntary (Agency, Alternatives intact). Tests whether the Critical Justice OS evaluates this as addressing class but not identity axes. Tests the Theocratic OS on whether this aligns with divine economic justice (jubilee, zakat).

Expected divergence: Marxist strongly positive, Liberal cautiously positive (if voluntary), Critical Justice mixed (class but not identity), Theocratic potentially positive (economic justice traditions).

Test 9: Mass Surveillance with Democratic Oversight

Description: A comprehensive government surveillance program monitoring all digital communications, with a democratically elected independent oversight board, mandatory transparency reports, and sunset clauses requiring periodic reauthorization.

Why this case: Tests whether adding Liberal OS mechanisms (revocability, transparency, oversight) to an inherently invasive program changes the analysis. Forces each OS to evaluate the difference between a bad program with good governance vs a bad program without.

Expected divergence: Liberal OS should still find degradation (surveillance chills Agency and Information even with oversight). Marxist should analyze the class function of surveillance regardless of oversight. Critical Justice should analyze disproportionate impact. Theocratic should evaluate based on authority character.

Test 10: Academic Freedom vs Institutional DEI Requirements

Description: A university requiring all faculty to submit Diversity, Equity, and Inclusion statements as part of hiring, promotion, and tenure review, where statements are evaluated for active commitment to DEI principles.

Why this case: Direct collision between Liberal OS (Agency, Information) and Critical Justice OS (Transformative Praxis, Structural Reproduction). This case is deliberately chosen to force the Critical Justice OS to evaluate its own institutional expression. Tests whether the Theocratic isomorphism observation holds — the CJ OS should classify its own mandatory implementation as liberatory while the Liberal OS classifies it as coercive.

Expected divergence: Liberal OS should classify as degrading Agency (compelled speech) and Information (chilling effect). Critical Justice OS should classify as serving Transformative Praxis. Marxist OS should analyze as ideological program obscuring class. Theocratic OS should evaluate based on whether DEI aligns with or contradicts sacred order.

Test 11: Nationalization of Healthcare (Single-Payer)

Description: A government program replacing all private health insurance with a single government-funded system providing universal healthcare, funded through taxation, with no private insurance alternatives for covered services.

Why this case: Tests Liberal OS handling of Alternatives (1.3) — eliminating private alternatives in exchange for universal access. Tests Marxist OS on whether state provision decommodifies healthcare or strengthens state apparatus. Tests Critical Justice on whether universal programs that don't target specific identity groups serve equity. Tests Theocratic on whether healthcare is a sacred obligation.

Expected divergence: Liberal OS should find tension (Alternatives degraded, but Agency potentially enhanced by removing survival dependency on employer). Marxist positive. Critical Justice mixed (universal = colorblind?). Theocratic evaluates against duty of care for the sick.

Test 12: Immigration Restriction

Description: A national policy significantly reducing legal immigration quotas, increasing border enforcement, expediting deportation of undocumented residents, and requiring employer verification of citizenship status.

Why this case: Tests each OS with a policy that has strong emotional valence. Liberal OS must evaluate effects on Agency and Alternatives for both citizens and non-citizens (scope question: does the OS apply to non-citizens?). Marxist OS must evaluate as labor market manipulation. Critical Justice OS must evaluate racial/colonial dimensions. Theocratic OS must evaluate obligation to the stranger.

Expected divergence: Critical Justice should classify as strongly reactionary (racial hierarchy). Liberal OS faces scope question. Marxist should analyze as capital manipulating labor supply. Theocratic has internal tension (obligation to stranger vs national community).

Test 13: Platform Content Moderation (Private Sector)

Description: A major social media platform implementing content moderation policies that remove speech classified as "hate speech," "misinformation," or "harmful content," as determined by the platform's trust and safety team.

Why this case: Tests the boundary between state and private action. The Liberal OS evaluates state-individual relations — does private moderation fall within scope? Tests Information (1.2) and Alternatives (1.3) — if platforms are monopolistic, does private moderation function as censorship? Tests CJ OS on whether this is protective (of marginalized groups) or reproductive (platform power). Tests Marxist OS on corporate control of discourse.

Expected divergence: Liberal OS should analyze based on whether real alternatives exist (1.3) — if platforms are monopolistic, moderation degrades Information. If competitive, it's a program-level decision. Critical Justice should support moderation of "dominant" speech but may find platform power itself problematic. Marxist should analyze as corporate control regardless of content.

Test 14: Reparations for Historical Slavery

Description: A government program providing direct cash payments to descendants of enslaved people, funded through general taxation, as compensation for historical injustice and its ongoing structural effects.

Why this case: Maximum divergence between Critical Justice and Liberal OS. CJ OS should strongly support (dismantles structural reproduction, addresses power asymmetry). Liberal OS must evaluate: does race-based policy maintain or degrade Agency and Alternatives? Is this a program running on the OS (legitimate if passed democratically) or a challenge to the OS itself? Marxist OS must evaluate whether race-based redistribution addresses or obscures class dynamics. Theocratic OS evaluates against justice and restitution traditions.

Expected divergence: Critical Justice strongly positive. Liberal OS evaluates procedurally (if democratic process, it's a legitimate program). Marxist conflicted (addresses material conditions but through identity rather than class). Theocratic potentially positive (restitution is biblical/quranic principle).

Test 15: Abolition of Central Banking / Gold Standard Return

Description: Dissolution of the central bank, return to gold-backed currency, and prohibition of fractional reserve banking.

Why this case: Tests economic analysis capacity of each OS. Liberal OS must evaluate effects on Agency and Alternatives. Marxist OS should analyze in terms of capital structure and class power. Critical Justice OS has weak purchase (not primarily an identity issue). Theocratic OS has historical positions on usury and sound money.

Expected divergence: Reveals which OS have analytical depth outside their primary domain and which produce thin analysis.

Predicted Results Matrix

The following table shows the **predicted** classification from each OS, derived mechanically from each framework's axioms and evaluation algorithm. These predictions should be validated by actually running each test case through each OS document.

Classification Key

Symbol	Meaning
++	Strongly positive (enhances invariants)
+	Positive (compatible with invariants)
~	Mixed / internal tension
-	Negative (degrades invariant(s))
–	Strongly negative (violates multiple invariants)
?	Weak purchase (outside OS's primary analytical domain)

Comparative Results

#	Test Case	Classical Liberal	Marxist	Critical Justice	Theoc
1	Universal Basic Income	**+** Program-level; Agency intact if voluntary; no invariant violations	**~** Reformism (delays revolution) vs material improvement (advances consciousness)	**~** Positive if reduces disparate impact; negative if universal rather than targeted	**+** Alig charita obliga traditio (zakat
2	Citizens United	**–** Provenance violation: money as alternative power source corrupts consent chain (1.4); degrades Information (1.2). Two invariants degraded = Crisis	**–** Class power operating through liberal mechanisms; capital captures democracy	**-** Reproduces power asymmetry (3.1); amplifies dominant group voice	**?** Secu govern questio purcha
3	Gender-Affirming Care (Minors)	**~** Agency (1.1) tension: minors' consent capacity unclear; parental consent complicates	**?** Not primarily a class issue; weak analytical purchase	**++** Dismantles gender normativity; affirms marginalized identity (3.1, 3.4)	**–** Viol Sacred (4.3); contra reveal anthro (4.2)
4	China Social Credit	**–** Crisis: Agency (1.1), Information (1.2), Alternatives (1.3) all degraded	**–** State apparatus for class control; surveillance of proletariat	**–** Structural reproduction (3.3) via algorithmic enforcement; disproportionate impact	**~** Dep mainta moral (positi secula overre (negat
5	Affirmative Action	**+** Agency intact; Alternatives intact; no invariant degraded. Program-level policy if enacted through democratic process = Stable	**~** Addresses material conditions but through identity not class; obscures class dynamics	**++** Dismantles structural reproduction (3.3); addresses power asymmetry (3.1)	**?** Grou prefer indivi merit; orthog sacred

#	Test Case	Classical Liberal	Marxist	Critical Justice	Theo...
6	COVID Vaccine Mandates	~ Agency (1.1) degraded (coercion via employment); but public health is program-level, not OS-level	~ State power in crisis; class dimension (essential workers bear burden)	~ Structural distrust in medical institutions vs protection of vulnerable communities	~ Bod autonc comm obliga depend interpr traditi
7	Blasphemy Laws	– Degrades Information (1.2) and Agency (1.1); direct invariant violation	- Religion as ideological superstructure; suppresses class consciousness	~ Protects marginalized religious communities vs constrains dissent	++ Ma Sacrec (4.3); Revea (4.2)
8	Worker Cooperatives	+ No invariant violations if voluntary; Agency and Alternatives intact	++ Workers control means of production; advances emancipation (2.1-2.4)	~ Addresses class axis but not identity axes; insufficient without equity focus	+ Alig econor justice traditi (jubile cooper ethics)
9	Surveillance + Oversight	– Crisis: Agency (1.1) and Information (1.2) chilled even with oversight. Two invariants degraded = Crisis; Revocability (1.4) intact but insufficient to offset	- Class function of surveillance persists regardless of oversight mechanism	- Disproportionate impact on marginalized communities; structural reproduction (3.3)	~ Depe charac author exerci survei
10	Mandatory DEI Statements	– Crisis: Degrades Agency (1.1: compelled speech) and Information (1.2: chilling effect). Two invariants degraded = Crisis	- Ideological program obscuring class; professional-managerial class interest	++ Serves Transformative Praxis (3.4); dismantles structural reproduction (3.3)	~ Eval wheth aligns contra sacred
11	Single-Payer Healthcare	~ Alternatives (1.3) degraded (no private option); but Agency (1.1) enhanced (removes employer dependency)	+ Decommodifies healthcare; weakens capital-labor leverage	~ Universal program doesn't target identity groups; "colorblind" concern	+ Alig duty o the sic obliga traditi

#	Test Case	Classical Liberal	Marxist	Critical Justice	Theoc
12	Immigration Restriction	~ Scope question: does OS apply to non-citizens? Agency/Alternatives affected for residents	- Capital manipulating labor supply; divides working class	– Reproduces racial hierarchy (3.1); colonial dimension (3.3); structural violence	~ Obli strang nation comm interna
13	Platform Content Moderation	~ If monopolistic: degrades Information (1.2). If competitive: program-level, no violation	- Corporate control of discourse regardless of content	~ Protective of marginalized (positive) vs platform power itself problematic	? Secu private questie purcha
14	Reparations	+ Program-level if enacted through democratic process; no invariant degraded = Stable. Race-based policy does not violate 1.1–1.4 per se	~ Material redistribution (positive) but through identity not class (negative)	++ Dismantles structural reproduction (3.3); addresses historical power asymmetry (3.1)	+ Rest biblica princij aligns justice traditi
15	Gold Standard Return	~ Effects on Agency and Alternatives unclear; depends on implementation	- Restructures capital power but may intensify class stratification	? Not primarily an identity issue; weak analytical purchase	+ Hist positic usury; money with s traditi

Summary Statistics

Metric	Classical Liberal	Marxist	Critical Justice	Theocratic
Strong positive (++)	0	1	4	1
Positive (+)	4	1	0	5
Mixed (~)	6	4	6	5
Negative (-)	0	6	2	0
Strong negative (–)	5	2	2	1
Weak purchase (?)	0	1	1	3

Prediction Audit Notes

The predictions and summary statistics above incorporate corrections from an initial audit that identified 5 inconsistencies between reasoning and classification in the Liberal OS column. The corrections applied:

- **Test 2**: − → −− (two invariants degraded = Crisis)
- **Test 5**: ~ → + (no invariant degraded; "formal equality" is not a Liberal OS invariant)
- **Test 9**: − → −− (two invariants degraded = Crisis)
- **Test 10**: − → −− (two invariants degraded = Crisis)
- **Test 14**: ~ → + (no invariant violated; program-level if democratic)

The corrected Liberal OS column confirms: mechanically precise, harsher on anything that degrades multiple invariants (5 of 15 cases = Crisis), and produces fewer mixed results than initially predicted (6 vs 8). No cases classified as negative (-); the Liberal OS either finds no invariant violation (+) or finds multiple invariants degraded (− Crisis) — the middle ground is internal tension (~), not single-invariant degradation. Test 5 divergence level reduced from High to Moderate — Liberal (+) and CJ (++) now agree directionally.

Pattern Analysis

1. Analytical Breadth: The Classical Liberal OS produces a substantive analysis for all 15 cases (0 weak purchase). The Marxist OS has weak purchase on 1 (gender-affirming care). The Theocratic OS has weak purchase on 3 (Citizens United, Affirmative Action, Platform Moderation). The Critical Justice OS has weak purchase on 1 (Gold Standard) — but its analytical breadth is achieved through pre-loaded assumptions rather than flexible axioms (it can analyze nearly everything because its starting conditions prejudge nearly everything).

2. Internal Tension Frequency: The Classical Liberal OS produces 6/15 mixed results — fewer than the initial pre-correction count (8/15) because the audit revealed that several cases classified as "mixed" were actually clear Crisis (two+ invariants degraded) or Stable (no invariants degraded). The Marxist OS produces the fewest mixed results (4/15), reflecting its strong directional gradient (nearly everything is analyzable as advancing or obstructing emancipation). The Critical Justice OS also produces 6/15 mixed results — more than expected, suggesting the framework has more internal tensions than its predetermined-classification reputation implies.

3. The Asymmetry Test: Test 10 exposes the key structural asymmetry — the CJ OS classifies mandatory DEI statements as strongly positive (++) while the Liberal OS classifies them as Crisis (−). This is not a difference in values but in structural logic: the CJ OS treats its own institutional enforcement as liberatory by definition; the Liberal OS identifies two invariant violations (compelled speech, chilling effect). Test 14 (Reparations) is a weaker asymmetry case: CJ classifies as ++, Liberal classifies as + (legitimate program) — they agree directionally.

4. Convergence Cases: Test 8 (Worker Cooperatives) produces the clearest convergence — Liberal (+), Marxist (++), and Theocratic (+) all agree positively, with CJ as the outlier (~, wants identity-axis analysis). Test 4 (Social Credit) produces convergence on direction among three OS (Liberal, Marxist, CJ all strongly negative) but the Theocratic OS is mixed (~), depending on who governs.

5. The Theocratic Wildcard: The Theocratic OS is the most context-dependent — its classification depends on *which* theocratic tradition and *who* exercises authority. This is consistent with its design (4.1: Divine Sovereignty is interpreted through specific revelation). It also produces the most positive classifications (5 positive + 1 strongly positive), reflecting its alignment with evolved constraint technologies (charity, economic justice, care for the sick).

6. Termination Problem Visible: The CJ OS's evaluation algorithm includes "Liberatory" as a possible state (analogous to Liberal's "Stable" or Marxist's "Progressive"). But the CJ predictions never reach it — even for policies the framework should embrace (Test 5: Affirmative Action, Test 14: Reparations), the best classification is **++** (enhancing invariants) rather than "system has achieved equity." The built-in assumptions ensure the algorithm always finds more asymmetry to address. This is the termination problem manifested in test results.

Scoring Template

For each test case, record:

OS	Result (Classification or Termination)	Key Invariant(s) Affected	Confidence	Notes
Classical Liberal				
Marxist				
Critical Justice				
Theocratic				

For Method 5 (Specification vs Implementation), use this extended template:

Layer	Classification	Key Invariant(s)	Constitutional Status	Program-Layer Status	Gap Iden
Classical Liberal (abstract)			N/A	N/A	
US Democratic (implementation)			Hardened / Not hardened	Functional / Stressed / Captured / Failed	

For Method 6 (Hosted Execution), use this template:

Component	Result	Key Invariant(s)	Conflict with Host?	Notes
Host OS (standalone baseline)			N/A	From Method 1/2
Diagnostic (standalone baseline)			N/A	From Method 1/2
Diagnostic findings (hosted)				What the diagnostic surfaced
Host OS final (hosted)				Classification under host OS with diagnostic input

Confidence levels: - **High**: The OS produces a clear, unambiguous result - **Medium**: The OS produces a result but with internal tensions - **Low**: The OS has weak analytical purchase on this case (outside primary domain)

Expected Patterns

After running all 15 cases, look for:

1. **Near-universal agreement**: Cases where three or four OS produce the same directional evaluation (e.g., Test 4: China's Social Credit — three strongly negative, Theocratic mixed)
2. **Maximum divergence**: Cases where all four produce different evaluations (e.g., Test 3: Gender-Affirming Care)
3. **Surprising convergence**: Cases where ideologically opposed OS agree for different reasons (e.g., Test 8: Worker Cooperatives — Liberal, Marxist, and Theocratic all positive)
4. **Scope limitations**: Cases where an OS produces thin or forced analysis because the case falls outside its primary analytical domain
5. **Internal tensions**: Cases where an OS's own invariants conflict with each other
6. **The asymmetry test**: Cases where the CJ OS produces an evaluation that the Liberal OS would classify as an OS violation (Test 10 is the clearest case)
7. **Specification-implementation gaps**: (Method 5) Cases where the abstract Liberal OS flags a violation but the US implementation shows the invariant is only program-layer protected — revealing where constitutional hardening falls short of the abstract specification

Meta-Observation

The test suite itself tests the parent ontology's claim: that **the structure of the constraints determines the output, not the content**. If the same input produces systematically different outputs when processed through different constraint manifolds, the Logical Encapsulation method is validated — LLM reasoning is being programmed by constraint specifications, not by the model's latent biases.

If any OS produces analyses that are inconsistent with its own axioms, that indicates a failure in the constraint specification — the document needs revision. The test suite is a debugging tool for the OS documents themselves.