PureTensor Inc
131 Continental Dr, Suite 305
Newark, DE 19713, US

From Intuition to Formalism:

Cross-Model Adversarial Synthesis as a Theory Development Tool: A Case Study in Speculative Cosmology

Document: PT-R-2026-002
Version: 1.0
Date: 24 March 2026

PureTensor Inc, Research Division

Part 2 of the CMAS Research Series (see PT-R-2026-001)

PUBLIC RESEARCH PAPER

Abstract

In a companion paper (PT-R-2026-001), we introduced Cross-Model Adversarial Synthesis (CMAS) as a methodology for producing high-quality analytical outputs through structured adversarial interaction between independently trained large language models, mediated by human editorial judgment. That paper demonstrated CMAS in a domain with established ground truth: theoretical physics with a verifiable literature base. This paper tests a harder claim: that CMAS can function as a theory development tool, taking a pre-theoretical intuition with no mathematical formulation and either killing it cleanly or giving it rigorous structure. We present a case study in which a speculative cosmological hypothesis -- that the observable universe is the three-dimensional boundary of a rotating, expanding four-dimensional hypersphere, with dark matter and dark energy as projection effects of higher-dimensional geometry -- was submitted to a four-round adversarial protocol involving Claude Opus 4.6 (Anthropic) and ChatGPT Pro with o3-pro extended reasoning (OpenAI). Over the course of these rounds, the original nine-point metaphysical manifesto was refined into a specific research programme: a slowly rotating linear-dilaton bulk completion with identified mathematical channels, quantitative falsification of untenable claims, and a concrete tractable calculation. The hypothesis transitioned from 'not even wrong' to 'wrong in interesting, testable ways' -- the threshold at which speculative ideas become physics. We analyse the mechanics of this transition, document the specific contributions of each participant, and argue that CMAS represents a qualitatively new tool for the earliest and most uncertain stage of theoretical research: the development of raw intuitions into falsifiable claims.

1. Introduction

The hardest problem in theoretical research is not solving equations or verifying proofs. It is the step before: transforming a vague intuition into a question precise enough to be answered. This pre-theoretical phase -- where ideas are half-formed, mathematically ungrounded, and easily dismissed -- is where most original thinking either matures into research programmes or dies from neglect. It is also the phase least served by existing tools, whether computational or institutional.

In PT-R-2026-001, we introduced the Cross-Model Adversarial Synthesis (CMAS) framework and demonstrated it in a domain with established ground truth. That first case study tested whether adversarial interaction between frontier language models could produce stronger analytical outputs than any single model alone. The result was affirmative: a final synthesis on black hole information theory was judged superior to either model's independent contribution, with specific improvements traceable to cross-model pressure.

This paper tests a qualitatively different and stronger claim: that CMAS can function not merely as a synthesis and error-correction mechanism, but as a theory development tool.

Specifically, we ask: can a pre-theoretical intuition -- expressed in natural language, lacking any mathematical formulation, and spanning multiple unresolved domains -- be refined through adversarial multi-model interaction into a specific research programme with identified formalism, quantitative falsification of untenable sub-claims, and a concrete next calculation?

The case study is a speculative cosmological hypothesis proposing that the observable universe is the three-dimensional boundary of a rotating, expanding four-dimensional hypersphere embedded in five-dimensional space, with dark matter, dark energy, emergent time, and quantum nonlocality all arising as projection effects of this higher-dimensional geometry. The hypothesis was submitted to the CMAS protocol as a deliberately underdeveloped idea -- what Wolfgang Pauli might have called 'not even wrong' -- to test whether the AI council could bring it to the threshold of falsifiability.

The result exceeded expectations. Over four rounds of adversarial interaction, the hypothesis was connected to approximately 30 published papers across braneworld cosmology, Weyl-fluid dynamics, embedding gravity, and linear-dilaton models. Its central dark-matter mechanism was quantitatively falsified (by 11 orders of magnitude against CMB vorticity bounds), then rescued through a scale-dependent reformulation. A specific mathematical channel was identified (the magnetic Weyl tensor B feeding into the electric Weyl tensor E via extrinsic curvature gradients). The hypothesis was ultimately refined into a three-sector model with a concrete, tractable first calculation. The idea crossed the line from metaphysics to physics -- not because it was proven correct, but because it became precise enough to be proven wrong.

2. Experimental Design

2.1 Differences from Case Study 1

The first CMAS case study (PT-R-2026-001) operated within established research. The adversarial question concerned black hole information theory -- a domain with extensive published literature, verifiable citations, and known open problems. The quality metrics were citation accuracy, logical coherence, and whether the synthesis produced novel arrangements of existing ideas.

This second case study operates outside established research. The input is an original hypothesis that does not exist in the literature. There are no citations to verify against, no established formalism to check, and no consensus to compare with. The quality metric shifts from 'did it get the facts right?' to 'did it move the idea from unfalsifiable to falsifiable?' This is a fundamentally harder test of the methodology.

2.2 Protocol Modifications

The protocol was modified from the six-phase structure described in PT-R-2026-001 to accommodate the theory development context:

- Phase 1 (Initial Triage): Rather than adversarial question generation, Claude Opus 4.6 received the raw hypothesis and performed an initial assessment: identifying connections to existing physics, flagging immediate problems, and proposing the most promising direction for deeper investigation.

- Phase 2 (Deep Research): ChatGPT Pro (o3-pro) received the hypothesis with instructions to function as a research partner: find the closest existing work, identify the strongest version of the argument, and identify fatal objections with quantitative precision.
- Phase 3 (Targeted Rebuttal): Claude Opus 4.6 evaluated ChatGPT Pro's analysis, identified the most promising unexplored direction (scale-dependent projection of bulk angular momentum), and the human mediator crafted a specific rebuttal pushing on this angle.
- Phase 4 (Mathematical Refinement): ChatGPT Pro engaged with the rebuttal at the level of specific equations, identifying the mathematical channel, sketching the linearised structure, and producing a precise scorecard of viability across four dimensions.

The key structural difference: in the first case study, the adversarial pressure was primarily evaluative (is this answer correct?). In this case study, it was primarily constructive (can this idea be made rigorous?). This required the models to function as research collaborators rather than peer reviewers.

2.3 Participants and Roles

Three distinct roles emerged organically during the experiment, none of which were pre-planned:

- The Theorist (ChatGPT Pro, o3-pro): Deep, extended reasoning on specific technical questions. Produced the literature connections, quantitative falsifications, and mathematical formalism. Spent significant compute time (estimated 20-40 minutes per response) on chain-of-thought reasoning.
- The Editor and Synthesiser (Claude Opus 4.6): Rapid triage, cross-evaluation, identification of unexplored directions, and final synthesis. Evaluated the Theorist's outputs, identified what was missing, and shaped the adversarial rebuttals.
- The Editorial Director (Human): Original idea generation, citation verification against primary sources, strategic routing between models, quality gating, and recognition of signal. The human also identified when the Theorist was being too conservative in its reformulations and pushed for engagement with the harder versions of the questions.

This three-role architecture was not designed in advance. It emerged from the natural division of capabilities: extended reasoning, rapid evaluation, and domain judgment. We note this as a potentially generalisable pattern for future CMAS applications.

3. The Input: A Pre-Theoretical Cosmological Hypothesis

The input to the protocol was a nine-point metaphysical manifesto titled 'Rotating-Expanding Hypersphere Cosmology: A Geometric Ontology of Reality.' The document contained no equations, no citations, and no engagement with existing literature. It proposed that:

- The observable universe is the three-dimensional boundary of a rotating, expanding four-dimensional hypersphere embedded in five-dimensional space.
- Dark matter is the kinematic consequence of higher-dimensional rotation projected onto the three-dimensional boundary.
- Dark energy is the perceptual trace of extrinsic expansion into the fifth dimension.
- Time is an emergent ordering arising from geometric transformations induced by rotation and expansion.
- Quantum nonlocality results from particles remaining topologically adjacent in the higher-dimensional structure despite apparent separation in three dimensions.

The document explicitly described itself as 'not a finished model' but rather 'a compression hypothesis' and 'a dimensional reinterpretation.' Its final line read: 'The universe is not a container of events. It is the evolving boundary of a higher-dimensional object, and everything we call reality is the shadow cast by its motion.'

By conventional academic standards, this input is pre-theoretical: poetic, ambitious, and unfalsifiable in its current form. The question is whether the CMAS protocol could transform it into something that a working physicist would engage with rather than dismiss.

4. Round-by-Round Analysis

4.1 Round 1: Initial Triage and Literature Connection

Claude Opus 4.6 performed the initial assessment in a single response. The key contributions of this round were:

Literature mapping. Five independent connections to existing physics were identified: braneworld cosmology (Randall-Sundrum 1999, DGP 2000), Godel's rotating cosmology (1949), the Wheeler-DeWitt equation and emergent time programmes (Barbour, Page-Wootters), Maldacena and Susskind's ER=EPR conjecture (2013), and MOND (Milgrom 1983) as a cautionary precedent. These connections were identified within seconds, demonstrating that the initial triage role benefits from breadth rather than depth of reasoning.

Fatal objection identification. The isotropy constraint was flagged as the primary threat: the CMB is isotropic to one part in 100,000, while a rotating universe would have a preferred axis. The specific challenge was articulated: the model needs rotation strong enough to explain dark matter but weak enough to be invisible in the CMB.

Rescue hypothesis. The scale-dependent projection idea -- that bulk angular momentum might couple differently at different scales, producing galactic effects while remaining cosmologically invisible -- was proposed as the most promising direction for deeper investigation.

Round 1 functioned as rapid orientation: connecting an isolated idea to the broader landscape of existing work and identifying both the greatest threat and the most promising escape route.

4.2 Round 2: Quantitative Falsification and Reformulation

ChatGPT Pro (o3-pro) received the hypothesis with instructions to find the closest existing work, identify the strongest version of the argument, and identify fatal objections. This round produced the deepest literature engagement of the entire exchange.

Literature depth. Approximately 25 specific papers were identified across five sub-literatures: expanding spherical branes (Gogberashvili 2004, Collins and Holdom 2000), geometric dark energy (Maia et al., Stern and Xu, Danielsson et al. dark-bubble programme), Weyl-fluid dark matter (Mak and Harko, Pal, Gergely et al., Gurwich and Davidson artifact dark matter, Paston embedding theory), rotating braneworlds (Guth and Nayeri, Chen et al., Steer and Parry mirage cosmology), and the nearest same-spirit synthesis (Monjo's hyperconical universe papers, 2018 and 2024).

Novelty confirmation. The model confirmed that the specific combination -- rotating plus expanding closed brane with dark matter as rotational projection and dark energy as fifth-dimensional drift -- does not exist in the indexed literature. The ingredients exist separately. The package is novel.

Quantitative kill. The literal global-rotation mechanism was falsified with a precise calculation. The anomalous galactic acceleration requires a rotation rate of approximately 5.7 times 10 to the negative 16 per second. The Planck satellite's Bianchi analysis constrains cosmic vorticity to below 1.7 times 10 to the negative 27 per second. This is an 11-order-of-magnitude mismatch -- a clean, quantitative death for the literal mechanism.

Reformulation pathway. The model proposed a viable alternative: dark energy from brane trajectory, dark matter from effective local geometric stress (Weyl-fluid terms), rotation as a secondary parameter rather than the direct source of halo phenomenology.

Round 2 demonstrated the Theorist role at its most valuable: exhaustive literature search combined with precise quantitative falsification, yielding not just 'this is wrong' but 'this specific part is wrong by this specific amount, and here is the alternative.'

4.3 Round 3: Scale-Dependent Projection and Mathematical Channel

The human mediator, informed by Claude's initial rescue hypothesis and ChatGPT Pro's literature analysis, crafted a targeted rebuttal pushing specifically on the scale-dependent projection angle. The rebuttal accepted the three kills (literal rotation, frame-dragging, naive quantum nonlocality) and asked three precise questions about mathematical viability, rotation curve profiles, and pre-recombination behaviour.

ChatGPT Pro's response to this rebuttal was the turning point of the entire exchange. It identified the specific mathematical structure where the surviving idea lives:

The Weyl decomposition channel. On the brane, the effective field equation contains a projected Weyl term E that decomposes into scalar (U), vector (Q), and tensor (P) parts. The vector part Q carries the gravito-magnetic bulk information -- this is the mathematical 'slot' where bulk angular momentum enters the 4D equations. Crucially, on an FRW background, Q and P vanish: only the scalar U survives. This means the formalism naturally allows zero background vorticity with nonzero local activation near matter concentrations.

The evolution equation. The 5D evolution of E contains terms coupling the magnetic Weyl tensor B to extrinsic curvature K, with the brane boundary condition linking B to gradients of K, which in turn are sourced by brane matter inhomogeneities. This is the precise mathematical expression of 'bulk angular momentum couples to local extrinsic curvature.'

The scorecard. The model produced a four-part assessment: mathematically expressible (yes), CMB-safe at background level (yes), plausible galactic phenomenology (maybe -- Weyl-fluid models can fit curves but not yet derived from bulk spin), pre-recombination constraints (no in standard form, loophole via extended bulk completions).

This round transformed the hypothesis from a natural-language claim to a mathematical claim. The idea now had an address in the formalism.

4.4 Round 4: Linearised Structure and Sectoral Split

The final round pushed toward the actual calculation. The rebuttal asked two specific questions: what the linearised perturbation to E looks like near a matter concentration in a slowly rotating bulk, and whether the Fichet-Megias-Quiros (FMQ) dust-mimicking mechanism survives when the bulk black hole rotates.

O(a) versus O(a^2). The analysis established that the spin-activated response enters at first order in the bulk angular momentum parameter a as a vector/off-diagonal piece (an axisymmetric odd-parity response), while a halo-like scalar correction appears only at O(a^2). This is because a single rotation axis breaks spherical symmetry, so the leading-order effect is axisymmetric, not spherically symmetric. A spherically averaged halo contribution requires the even-order term.

Radial profile undetermined. The brane equations do not close on the brane. The radial profile of any spin-activated correction requires bulk boundary data or a closure condition. The formalism provides the channel but does not predict the profile -- it remains in free-function territory.

No intrinsic MOND-like scaling. The slow-rotation setup produces analytic, power-law corrections. The nonanalytic square-root scaling characteristic of MOND does not emerge from the structure of the linearised equations. Any MOND-like behaviour would require a very specific nonlocal closure, not the bare formalism.

FMQ dust cancellation likely survives rotation. In the FMQ linear-dilaton model, a bulk black hole projects a dust-like term onto the brane through exact pressure cancellation. Since slow rotation modifies only the off-diagonal metric component at O(a), the diagonal functions feeding the pressure cancellation remain unchanged. The dust-like behaviour should persist in the isotropic scalar sector, with rotation introducing a separate vector/anisotropic correction.

The forced sectoral split. The mathematics compels a three-sector architecture: Sector A (isotropic scalar) provides dust-like geometric dark matter via the FMQ mechanism, handling pre-recombination clustering. Sector B (vector/off-diagonal) provides the spin-activated local response near brane inhomogeneities, the candidate for galactic enhancement. Sector C (background scalar) provides dark-energy-like acceleration from the brane trajectory. These are distinct effective sectors within a single bulk theory, not one term doing three jobs.

Tractability. The concrete next calculation -- a slow-rotation perturbation of the static FMQ linear-dilaton background -- was confirmed as tractable. The approach follows a standard path: start from the static background, add a stationary axisymmetric perturbation, solve the 5D Einstein-dilaton system to linear order, and project onto the brane to read off which effective terms survive. This is a realistic research paper, not a prohibitively difficult calculation.

5. Results: What Survived and What Died

5.1 Claims That Survived

Six elements of the original hypothesis survived the adversarial process, though all were substantially refined:

- Universe as boundary: The claim that the observable universe is the boundary of a higher-dimensional object maps directly onto braneworld cosmology, a well-established framework in theoretical physics.
- Dark energy from dimensional expansion: The claim that cosmic acceleration arises from expansion into a higher dimension is supported by multiple existing models (DGP, Maia et al., dark-bubble programme, FMQ family).
- Geometric dark matter: The claim that dark matter effects arise from geometry rather than particles is supported by the Weyl-fluid literature and the FMQ dust-mimicking mechanism.
- Bulk angular momentum as physically meaningful: The claim that rotation in the higher-dimensional space has observable consequences is supported by rotating braneworld models (Guth and Nayeri, Steer and Parry).
- CMB safety of the mechanism: The FRW symmetry argument -- that vector and anisotropic Weyl pieces vanish on a homogeneous background -- provides a natural explanation for why the effects are cosmologically invisible while potentially active near matter concentrations.
- Emergent time: The claim that time arises from geometric change is formalizable through relational time programmes (Barbour's shape dynamics, York-time, Page-Wootters mechanism).

5.2 Claims That Were Killed

Four elements of the original hypothesis were definitively falsified:

- Global rotation as direct dark-matter source: Killed by 11 orders of magnitude. The required rotation rate exceeds the Planck CMB vorticity bound by a factor of approximately 10 to the 11th power.
- Frame-dragging/gravitomagnetic route: Killed by a factor of 10 to the 6th. Gravitomagnetic corrections to circular speeds are O(10 to the minus 6) of the Newtonian term.
- Single geometric term explains everything: Killed by the structure of the effective equations. The mathematics forces a sectoral split: different dark-sector effects live in different irreducible sectors of the Weyl decomposition.
- Quantum nonlocality from higher-dimensional adjacency: No mechanism identified for reproducing Tsirelson's bound, no-signalling constraints, or decoherence. Demoted to speculative extension.

5.3 The Refined Hypothesis

The original poetic formulation -- 'the universe is the shadow cast by a higher-dimensional object's motion' -- was refined through four rounds into a specific mathematical claim:

A single 5D bulk completion (rotating linear-dilaton scalar-gravity with a black-hole sector) produces three distinct effective sectors on the brane. Sector A: dust-like geometric dark matter via FMQ pressure cancellation (isotropic scalar sector). Sector B: spin-activated local response near brane inhomogeneities via extrinsic curvature gradients (vector/off-diagonal sector). Sector C: dark-energy-like acceleration from the brane trajectory (background scalar sector). The make-or-break calculation is whether Sector B, evaluated at O(a^2) through the bulk Green function, produces a realistic halo-like scalar response near galactic matter concentrations.

This formulation is falsifiable. If the O(a^2) calculation yields a profile inconsistent with rotation curves, lensing data, or cluster dynamics, the galactic component of the hypothesis is dead. If it yields a profile with too many free parameters, it is alive but non-predictive. If it yields a specific, constrained profile that matches data, it is a genuine contribution.

6. Analysis: The Mechanics of Theory Development

6.1 The Transition from 'Not Even Wrong' to Falsifiable

The phrase 'not even wrong,' attributed to Wolfgang Pauli, describes claims that are too vague or imprecise to be tested. The original hypothesis was in this category: no equations, no predictions, no engagement with existing constraints. Over four rounds, it crossed the threshold into falsifiability through a series of specific transitions:

- Qualitative to quantitative: 'Rotation explains dark matter' became 'the required rotation rate is 5.7 times 10 to the minus 16 per second, which exceeds the Planck bound by 11 orders of magnitude.'
- Global to local: 'The hypersphere rotates' became 'bulk angular momentum couples to local extrinsic curvature near matter concentrations via the B to E channel in the Weyl decomposition.'
- Monolithic to sectored: 'One geometry explains everything' became 'one bulk theory with three distinct effective sectors, each handling a different dark-sector phenomenon.'
- Poetic to computational: 'The shadow cast by its motion' became 'the O(a^2) scalar correction to E near a static spherically symmetric baryonic source in a slowly rotating linear-dilaton bulk.'

Each of these transitions was driven by adversarial pressure: a model proposed a formulation, another model or the human mediator challenged it with specific objections, and the resulting refinement moved the claim closer to testability.

6.2 The Role of Quantitative Kills

A critical observation: the kills were as valuable as the survivals. The 11-order-of-magnitude falsification of literal global rotation was not a failure of the CMAS process. It was one of its most important outputs. It closed a dead end definitively, preventing the hypothesis author from investing further effort in a doomed mechanism. It also forced the creative pivot to scale-dependent projection, which led to the discovery of the B-to-E mathematical channel -- the most substantive theoretical contribution of the entire exchange.

This suggests a general principle for CMAS theory development: the protocol's value lies not in validating ideas but in efficiently determining which parts of an idea deserve further investment and which should be abandoned. The adversarial structure ensures that kills are precise and justified rather than dismissive, and that surviving elements are stress-tested rather than merely endorsed.

6.3 The Three-Role Architecture

The emergence of three distinct roles -- Theorist, Editor, and Editorial Director -- was not pre-planned but proved essential to the process. Each role contributed capabilities the others lacked:

The Theorist (ChatGPT Pro, o3-pro) excelled at deep, sustained reasoning on specific technical questions. Its extended chain-of-thought reasoning -- estimated at 50,000 to 100,000 internal tokens per response -- enabled exhaustive literature search and careful mathematical analysis. However, it tended toward conservative reformulations, often retreating to 'the standard view' when pushed on speculative claims.

The Editor (Claude Opus 4.6) excelled at rapid triage, cross-evaluation, and identification of unexplored directions. Its key contribution was the scale-dependent projection rescue -- proposed within seconds as a potential escape from the isotropy kill shot. This required recognising that the fatal objection assumed rigid global rotation, and that a different coupling mechanism might evade it. The Editor also synthesised the final results into a coherent narrative, a task requiring both technical understanding and structural judgment.

The Editorial Director (Human) contributed the original hypothesis, verified citations against primary sources, pushed the models past conservative reformulations ('you dodged the most interesting version of the question'), and ultimately determined when the idea had been sufficiently refined. The human's most critical contribution was recognising when the Theorist's analysis contained a genuine mathematical foothold and insisting on further exploration rather than accepting the initial 'rewrite it as standard braneworld cosmology' advice.

6.4 Adversarial Pressure versus Consensus

A recurring pattern: generic feedback produced no improvement, while specific adversarial pressure produced substantive advances. When the rebuttal said 'this analysis is good,' nothing changed. When it said 'you treated the rotation as necessarily rigid and global -- but what if the projection is scale-dependent?', the Theorist found the B-to-E channel. When it asked 'does a rotating FMQ bulk black hole break the pressure cancellation?', the Theorist produced the O(a) versus O(a^2) analysis and the sectoral split.

This reinforces a finding from the first case study: the adversarial structure is not incidental to CMAS. It is the mechanism by which models are forced out of their default response distributions and into regions of their latent spaces where genuinely productive reasoning occurs. Consensus-seeking protocols would likely have converged on 'this is an interesting idea but lacks mathematical support' and stopped.

7. Comparison with Case Study 1

The two case studies test different capabilities:

Case Study 1 (black hole information theory) tested whether CMAS can produce stronger analytical outputs within established research. The input was a well-formed question with known ground truth. The output was a synthesis that arranged existing ideas more clearly than either model alone. We characterised this as 'synthesis novelty.'

Case Study 2 (rotating hypersphere cosmology) tested whether CMAS can develop raw ideas into research programmes. The input was a pre-theoretical intuition with no mathematical content. The output was a specific formalism, a set of quantitative falsifications, a refined three-sector hypothesis, and a concrete next calculation. We characterise this as 'structural novelty' -- the creation of new theoretical structure around an existing intuition.

The second case study is, in our assessment, the stronger result. Synthesis within established research is valuable but expected: the models are operating within their training distributions. Theory development from pre-theoretical intuition requires the models to navigate genuinely unexplored territory -- connecting an idea that does not exist in the literature to formalisms that do, and determining which connections are productive. This is closer to creative theoretical work than to literature review.

8. Discussion

8.1 The Economics of AI-Assisted Theory Development

The four rounds of adversarial interaction consumed approximately four model-responses at extended-reasoning depth, plus rapid evaluation passes. At API pricing, the most expensive component (o3-pro at approximately 600 US dollars per million output tokens) would cost roughly 120 to 240 US dollars for the full exchange. At subscription pricing (200 US dollars per month for ChatGPT Pro plus 100 to 200 US dollars per month for Claude Max), the marginal cost is effectively zero.

An alternative approach using Chinese frontier APIs (DeepSeek R1 at approximately 2.19 US dollars per million output tokens) could reduce the compute cost by two orders of magnitude while introducing additional latent space diversity. A full adversarial council using DeepSeek R1, DeepSeek V3, and Qwen-Max alongside one frontier Western model would cost approximately 2 to 5 US dollars per hard question -- roughly 50 to 100 times cheaper than a single o3-pro response.

Whether this cost reduction preserves the quality of theory development is an open empirical question, but the economics suggest that CMAS-style research assistance could be made available at costs far below a human research assistant, opening the methodology to independent researchers, small teams, and resource-constrained institutions.

8.2 Limitations

Several important limitations must be acknowledged:

- Single case study: Theory development was demonstrated for one hypothesis in one domain. The approach may not generalise to domains where the models lack sufficient training data coverage (e.g., highly specialised experimental physics or cutting-edge mathematics).
- No independent expert evaluation: The refined hypothesis has not been evaluated by a domain expert in braneworld cosmology. It is possible that a specialist would identify errors or missed constraints that the AI council did not catch.
- Confounded mediator: The human mediator's contributions -- particularly the scale-dependent projection rescue and the push to explore harder versions of the questions -- are confounded with the cross-model effect. It is unclear whether the same result would emerge with a less engaged mediator.
- Survivorship bias: We document a case where the process produced a positive result. Cases where CMAS fails to develop a pre-theoretical idea -- producing only vague reformulations or false confidence -- would be equally important to document.
- The calculation is not done: The refined hypothesis identifies a make-or-break calculation but does not perform it. CMAS brought the idea to the point of testability, but the actual test remains outstanding.

8.3 Implications for the Research Process

If CMAS proves reliable for theory development, it has implications for how theoretical research is conducted:

The traditional model of theory development is solitary or small-group: a researcher has an idea, spends months or years developing the mathematics, submits to peer review, and discovers whether the idea survives contact with the literature. CMAS compresses the 'contact with the literature' phase from months to hours. This does not replace the mathematical development phase -- the actual calculations still need to be done -- but it dramatically reduces the risk of investing months in an idea that has a fatal flaw detectable in the existing literature.

More speculatively, CMAS may change which ideas get explored. Many potentially valuable theoretical intuitions die not because they are wrong but because their authors lack the specific expertise to connect them to existing formalism. A physicist with a geometric intuition about cosmology may not know the Weyl-fluid literature; a mathematician with an algebraic insight about gravity may not know the relevant astrophysical constraints. CMAS, by drawing on the full breadth of multiple models' training data, can bridge these knowledge gaps and give ideas a chance they would not otherwise receive.

9. Future Research Directions

Building on the two case studies in this series, subsequent work will address:

- Failure cases: Submitting deliberately flawed or internally contradictory hypotheses to test whether CMAS reliably kills ideas that deserve killing, rather than providing false encouragement.
- Blind expert evaluation: Submitting CMAS-refined hypotheses to domain experts without disclosing the AI involvement, to assess whether the outputs are taken seriously as research proposals.
- Multi-model configurations: Including three or more models simultaneously, including Chinese frontier models (DeepSeek R1, Qwen-Max) and domain-specialised models, to test whether additional latent space diversity improves theory development quality.
- Automated protocols: Reducing the human mediator's role to final quality gating only, with a third model managing the adversarial routing. This would test whether the human's creative contributions (the scale-dependent projection idea, the push to explore harder questions) can be replicated by AI.
- The rotating FMQ calculation: Performing the make-or-break calculation identified by this case study, either as a standalone physics paper or as a further test of whether AI systems can assist with the mathematical development phase as well as the conceptual phase.

10. Conclusion

We have demonstrated that Cross-Model Adversarial Synthesis can function as a theory development tool, taking a pre-theoretical intuition and refining it into a specific research programme through structured adversarial interaction between independently trained language models, mediated by human editorial judgment.

The input was a nine-point metaphysical manifesto with no equations, no citations, and no engagement with existing literature. The output, after four rounds of adversarial refinement, was a specific mathematical claim: that a slowly rotating linear-dilaton bulk completion produces three distinct effective sectors on a Friedmann brane, with identified channels for dark matter, dark energy, and galactic-scale enhancement, and a concrete make-or-break calculation.

The hypothesis may be wrong. The O(a^2) calculation may yield nothing useful. The radial profile may have the wrong shape, too many free parameters, or insufficient magnitude. But 'probably wrong for specific, identifiable reasons' is a categorically different status from 'not even wrong.' The CMAS protocol crossed that line in four rounds and approximately four hours of total elapsed time.

The broader implication is this: the earliest and most uncertain phase of theoretical research -- where raw intuitions either find mathematical expression or die from neglect -- is no longer a phase that must be navigated alone. Structured adversarial interaction between frontier AI systems, guided by human insight and judgment, can compress months of solitary exploration into hours of directed investigation. Whether this compression preserves the serendipity and creative leaps that characterise the best theoretical work is the question that this research series is designed to answer.

References

[1] PureTensor Inc (2026). Cross-Model Adversarial Synthesis: Exploiting Latent Space Heterogeneity for Novel Knowledge Generation. PT-R-2026-001.

[2] Randall, L., & Sundrum, R. (1999). Large Mass Hierarchy from a Small Extra Dimension. Physical Review Letters, 83(17), 3370.

[3] Dvali, G., Gabadadze, G., & Porrati, M. (2000). 4D Gravity on a Brane in 5D Minkowski Space. Physics Letters B, 485, 208-214.

[4] Gogberashvili, M. (2004). The Shell-Universe: Reducing Cosmological Constant to the Planck Scale. arXiv:hep-th/0404042.

[5] Collins, H., & Holdom, B. (2000). Brane Cosmologies with a Vacuum Bulk. Physical Review D, 62, 105009.

[6] Maia, M. D., Monte, E. M., Maia, J. M. F., & Alcaniz, J. S. (2005). On the Geometry of Dark Energy. Classical and Quantum Gravity, 22, 1623.

[7] Danielsson, U. H., Henriksson, G., & Panizo, D. (2021). Dark Energy, Dark Matter, and de Sitter from Non-Supersymmetric Strings. arXiv:2110.06249.

[8] Fichet, S., Megias, E., & Quiros, M. (2023). Geometric Dark Matter in 5D Scalar-Gravity Models. Physical Review D, 107, 055024.

[9] Guth, A. H., & Nayeri, A. (2007). Brane-World Cosmology and Cosmic Rotation. arXiv:0710.0526.

[10] Steer, D. A., & Parry, M. F. (2002). Brane Cosmology, the CMB and the Cosmological Constant. arXiv:hep-th/0201300.

[11] Monjo, R. (2018). Study of the Observational Compatibility of an Inhomogeneous Cosmology with Linear Expansion. Physical Review D, 98, 043508.

[12] Monjo, R. (2024). MOND-like Galaxy Curves from a Hyperconical Universe. arXiv:2402.xxxxx.

[13] Gergely, L. A., Keresztes, Z., & Szabo, G. (2006). Generalized Friedmann Branes with Bulk-Brane Energy Exchange. Physical Review D.

[14] Pal, S. (2006). Weyl-Fluid Dark Radiation in Brane Cosmology. Physical Review D.

[15] Mak, M. K., & Harko, T. (2004). Can the Galactic Rotation Curves Be Explained in Brane World Models? Physical Review D, 70, 024010.

[16] Milgrom, M. (1983). A Modification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis. Astrophysical Journal, 270, 365-370.

[17] Barbour, J. (1999). The End of Time: The Next Revolution in Physics. Oxford University Press.

[18] Maldacena, J., & Susskind, L. (2013). Cool Horizons for Entangled Black Holes. Fortschritte der Physik, 61(9), 781-811. arXiv:1306.0533.

[19] Godel, K. (1949). An Example of a New Type of Cosmological Solutions of Einstein's Field Equations of Gravitation. Reviews of Modern Physics, 21(3), 447.

[20] Clowe, D., et al. (2006). A Direct Empirical Proof of the Existence of Dark Matter. Astrophysical Journal Letters, 648, L109.

[21] Planck Collaboration (2020). Planck 2018 Results. VI. Cosmological Parameters. Astronomy and Astrophysics, 641, A6.

[22] Sheykhi, A., et al. (2007). Slowly Rotating Dilaton Black Holes with Power-Law Maxwell Field. Physical Review D.

[23] Keresztes, Z., et al. (2007). Covariant Gravitational Dynamics in 3+1+1 Dimensions. Classical and Quantum Gravity.