

Argumentative Tension through Semantic Displacement.

An Experiment Applied to the Debate on AI Development.

Mariano Mosquera
Universidad Católica de Córdoba (Argentina)
mariano.mosquera@ucc.edu.ar

Abstract

This article examines the generation of argumentative diversity from positions produced by artificial intelligence itself in debates about its development. It presents a semantic displacement experiment applied to a debate from the synthetic synIALO corpus, using a language model, an external ontology represented in SKOS (Simple Knowledge Organization System), and a procedure of argumentative counterposition aimed at preserving genuine tensions. An original debate comprising twenty arguments and a transformed debate comprising twenty-one were compared through qualitative analysis and a size-adjusted composite diversity index. Eight transformations were identified, including the introduction of a concept of dual contestability, argumentative reformulations, and new relationships between distant regions of the debate. The overall diversity index increased by 5.87%, with increases in dimensional diversity (3.78%), SKOS richness (22.49%), and functional diversity (9.89%), while normalized semantic distance decreased by 2.98%. The findings provide exploratory evidence that AI-generated adversarial positions can be reconfigured to broaden ontological and functional variety without eliminating their tensions.

Keywords

Argumentative diversity; semantic displacement; argumentative contradiction; ontology; deliberation.

I. Introduction

1. Problem and Hypothesis

Artificial intelligence can contribute to the homogenization of human conceptions through training and selection biases that favor dominant representations and reduce the visibility of minority positions. This phenomenon can be intensified by two related mechanisms: cultural feedback, whereby AI-mediated outputs influence new human outputs that subsequently feed back into the systems, and model collapse, associated with the deterioration of diversity in original distributions when models are recursively trained on synthetic data (Shumailov *et al.*, 2024; Doshi and Hauser, 2024). Added to this phenomenon is a feature of the relational organization of language that, under certain conditions, can amplify conceptual homogenization. Language constructs meaning through differences and oppositions that, in certain adversarial contexts, may be associated with relations of dominance and subordination (Dung, 1995). One conception prevails while others are displaced, without their contradictions necessarily generating new conceptual possibilities (Saussure, 1916; Mouffe, 2005).

These mechanisms affect not only cultural, political, scientific, and economic narratives but also current debates about the development of AI itself (Bender *et al.*, 2021). In summary, the problem of homogenizing AI and adversarial language has a direct impact on debates about the development of AI itself. Thus, certain architectures, priorities, and solutions may become established as dominant trajectories, while other technologically plausible alternatives are relegated by dynamics of selection, concentration, and competition (Bommasani *et al.*, 2021).

Within this framework, our hypothesis is that adversarial positions generated by artificial intelligence in the debate about its own development contain possibilities for semantic and argumentative innovation. This innovation can be expressed as argumentative tension and diversity and enables non-adversarial alternatives for resolving the debate on AI development. To make this tension explicit, we developed an experiment based on the semantic displacement of an AI-generated debate. A previously developed corpus and an external ontology tested in multi-agent debates were used. The result shows the reformulation and incorporation of positions and relationships, with tensions and approaches that integrate the debate as a whole, favoring agreements and resolutions while simultaneously increasing diversity.

2. Background

Decision sciences constitute an interdisciplinary field that brings together beliefs, rationality, and governance conditions for action. Decision theory, multicriteria analysis, bounded rationality, and collective decision-making procedures, among others, make relevant contributions (Simon, 1955; Keeney and Raiffa, 1976). From our perspective, the state of the art in decision sciences makes it possible to identify diverse argumentative fields for the creation of collective value through deliberation. This involves articulating justificatory transparency that equalizes the conditions of argumentation and reciprocity to foster better prospects for agreement and resolution in a debate. For this reason, our experiment uses an external ontology based on decision sciences. This orientation has precedents in theories of communicative rationality and public deliberation (Habermas, 1984; Gutmann and Thompson, 2004).¹

Furthermore, the possibility of producing new conceptual configurations from opposing positions has important precedents in Jungian psychology (Jung, 1969) and the Confucian conception of harmony (Mosquera, 2020; Yao, 2023), among other theoretical traditions. The concept of tension used in this study can be situated in continuity with these perspectives, although it is neither an unequivocal designation nor a theory exclusive to them. In contemporary studies of creativity and innovation, tensions between seemingly incompatible demands are regarded, in certain approaches, as productive conditions for exploring new solutions and configurations (Smith and Lewis, 2011). Moreover, this perspective is also present in recent research on artificial intelligence (Salma *et al.*, 2025).

3. Objective

The objective of our study is not to eliminate difference or achieve consensual uniformity, but to produce argumentative diversity through the productive stabilization of genuine contradictions in a debate on AI development generated by AI itself. From this perspective, possibilities for resolving the debate are then interpreted.

The project does not seek to replace one dominant conception with another or definitively resolve their contradictions. Its philosophical principle is that conflict should neither be overcome nor eliminated; it should remain and operate productively within a new conception. Synthesis in tension is not equivalent to a compromise solution, superficial complementarity, or the victory of one position. It entails a new configuration in which genuinely contradictory elements coexist without losing their constitutive tension.

¹ In our conceptual framework, debate and deliberation are closely related in their definitions, although the tradition of debate is more adversarial and zero-sum in nature, whereas deliberation is more closely associated with the creation of collective value. Nevertheless, this project can be understood as a transition from debate to deliberation.

II. Method

The design comprises four consecutive operations: identification of argumentative contradictions concerning artificial intelligence in a debate corpus; semantic displacement using an LLM and an external ontology developed in SKOS; argumentative counterposition as prompt engineering; and characterization and interpretive evaluation of the resulting syntheses of tension. All experimental documentation is available in the open repository, together with its visualization on Hugging Face, to ensure reproducibility (Geburu et al., 2021).²

4. Corpus

The study uses the syncIALO argumentation and counterargumentation corpus (A Multi-Purpose Synthetic Debate and Argument Mapping Corpus), developed by the DebateLab team at the Institute of Philosophy of the Karlsruhe Institute of Technology (KIT), Germany, and available on Hugging Face in the DebateLabKIT/syncialo-raw repository, which contains argumentative positions generated by artificial intelligence models (DebateLabKIT, n.d.). The corpus undergoes a procedure of thematic isolation (debates on AI development) through semantic search, reconstruction of argumentative relationships, and identification of genuine contradictions. This operation makes it possible to organize the corpus through relational structures that can be represented and analyzed using NetworkX (Hagberg *et al.*, 2008). The methodological relevance of the corpus lies in the fact that the contradictions are not introduced externally by human researchers but recovered from arguments previously generated by artificial intelligence itself. For this particular experiment, debate 0988 was selected from the corpus, based on the central premise “Implementing artificial intelligence will ultimately enhance global human well-being”. Twenty central arguments were selected in a balanced manner from the original debate to make the experiment and its human interpretation manageable.

5. Displacement

An LLM (OpenIA Astra, maximum reasoning) expands the space of conceptual possibilities using exclusively an external SKOS ontology, organized into four key dimensions of attention (values and agenda), rationality, and governance (Miles and Bechhofer, 2009). The ontology had previously been developed and evaluated using specialized corpora, both to examine its conceptual and semantic fit and to identify external novelties relative to its original structure. The ontology was also tested in previous multi-agent debate experiments. This constitutes a methodological precedent for the project and makes it possible to use the ontology as a previously examined semantic infrastructure on which conceptual displacements are performed.

In other words, the external ontology provided constitutes the guiding conceptual framework for semantic displacement. Its concepts, relationships, and hierarchies guide the transformation of contradictory positions and the generation of novel syntheses of tension. Its use is not limited to that of a dictionary, nor does it imply a closed catalogue of solutions; rather, it helps guide new conceptual configurations.

6. Prompt Engineering

The process encompasses seven possible operations on the argumentative structure: creation, reformulation, merging, and deletion of arguments; creation and deletion of links; and

² GitHub link: <https://github.com/MarianoMosquera/CCLTR>

Hugging Face link: <https://huggingface.co/spaces/mmosquera/ccltr-argument-explorer>

hierarchical ordering of arguments. All operations must be explicitly evaluated according to their argumentative value, without establishing quotas or privileging the creation of new arguments. The transformation may modify argumentative relationships and configurations while maintaining the distinction between PRO and CON positions (support and attack in the original debate) and identifying genuine oppositions, without inferring an argument's overall position from the valence of an edge.

The primary objective of the prompt is to produce novel argumentative syntheses that transform the relationships between contradictory positions while keeping their tensions active and recognizable. Synthesis must not be confused with consensus, averaging, juxtaposition, superficial compromises, or fictitious resolutions of conflict. Each selected operation must be justified in terms of its conceptual contribution, the positions it transforms, the contradiction it preserves, and its conceptual superiority over discarded alternatives. When the creation of new arguments is proposed, it must be justified why this is preferable to reconfiguring existing ones. The deletion of arguments or links must not silently suppress relevant contradictions; it must explicitly indicate where the original tensions remain represented. Likewise, the process must remain independent of the results, measurements, and classifications of the original debate. Finally, each synthesis retains its connection to the argumentative contradictions of the original debate, making it possible to reconstruct the starting positions and the procedure through which the new conception of tension was generated.

7. Interpretation and Diversity

Two operations were performed on the results obtained. First, the content was interpreted by identifying the transformations made to the original debate, including their types, argumentative regions, and impact on the overall debate. Second, diversity was measured across the four key dimensions of the external ontology, together with ontological richness at the conceptual level, functional diversity, and vector-based semantics. For this purpose, the original and transformed debates were compared, with adjustment for size.

Regarding the second point, the overall composite diversity index is constructed from four components normalized to the interval $[0,1]$, whose values must be finite and whose presence is mandatory for the calculation. The components are dimensional diversity (D_{dim}), SKOS richness (R_{skos}), functional diversity (D_{func}), and semantic diversity (D_{sem}). Dimensional diversity and SKOS richness each receive a weight of 0.30, while functional and semantic diversity each receive a weight of 0.20. Dimensional diversity represents the distribution of ontological anchors across four dimensions: values, agenda, rationality, and governance. Functional diversity represents the distribution across three argumentative functions: pragmatic, conditional, and prospective. Both are calculated using normalized Shannon entropy. SKOS richness expresses the coverage or variety of distinct ontological concepts. In the operational specification considered, it is calculated as the mean number of distinct SKOS URI identifiers divided by a reference of 398 concepts (the entire external ontology). Semantic diversity, for its part, captures variation in meaning through vector representations or embeddings (Reimers and Gurevych, 2019). Its operationalization as semantic distance consists of calculating, for pairs of arguments, the mean of $1 - \text{cosine similarity}$ between their embeddings and normalizing the result by dividing by 2. To control for differences in size between the original and transformed debates, an adjusted comparison is conducted through sampling without replacement to a common size. Where feasible, all possible combinations are enumerated; otherwise, Monte Carlo resampling is used with one thousand repetitions and a configurable seed. Semantic sensitivity tests and robustness analyses for variations in length are also considered.

III. Results

The original debate, comprising twenty arguments, began with a central thesis favorable to artificial intelligence as a generator of collective well-being, with twelve arguments in favor and eight against. The favorable arguments were primarily associated with advances in health, optimization of resource allocation, improvements in information for public policy decision-making, redundant systems to address the risks of artificial intelligence, and combating climate change. The latter theme was linked to arguments supporting a notion of collective well-being. The arguments opposing the central thesis, meanwhile, concerned AI training biases, the elimination of jobs, the risks AI poses to infrastructure, security threats from cyberattacks, shortcomings in human capacities to take advantage of AI benefits, and challenges to collective well-being as an absolute concept. Subsequently, semantic displacement and prompt engineering produced eight transformations, including the creation of one new argument, the reformulation of two others, and the incorporation of new supporting and opposing links between arguments, resulting in a new debate of twenty-one arguments.

The most significant transformation occurred in the tension between the affirmation of collective well-being and its questioning as an absolute concept. A new argument termed dual contestability was introduced into this space, semantically shifting the controversy toward participatory mechanisms capable of articulating community definitions of collective well-being with individual rights and conceptions of well-being that transcend the communities themselves. Significantly, this argument is positioned against the central thesis, establishing that collective well-being cannot be assumed to be a universal outcome of artificial intelligence but instead requires mechanisms of participation and contestation through which communities can negotiate their own definitions of well-being and the tensions with individual and universal rights.

A second transformation reformulated the favorable argument concerning the benefits of AI-generated information for public policy design, which was challenged by limitations in human capacities to make use of those benefits. The reformulation introduces a dynamic and evolutionary perspective in which improvements in public policies and the development of human and institutional capacities advance through a process of progressive convergence, without requiring a perfect correspondence between the two at every moment.

A third transformation reformulated a favorable argument maintaining that redundancy in artificial intelligence systems could prevent infrastructure risks. The new formulation establishes that redundancy alone is insufficient and must be complemented by risk analysis and consistent safety criteria.

The new links connect previously distant argumentative regions. For example, limitations in human capacities to take advantage of AI benefits in public policies were linked through an attack relationship to the argument concerning infrastructure risks, which belonged to another region of the debate, showing that deficiencies in human capacities may also compromise the identification and management of security risks. Similarly, an attack relationship was established between AI biases, as an argument opposing the central thesis, and its potential to reduce health inequalities, showing that such biases may compromise the expected benefits in health and making explicit the opposition between arguments originating in different argumentative regions.

In sum, semantic displacement introduced greater demands for participation, risk analysis, dynamic processes, and conditional logics while preserving the original argumentative tensions. These transformations allowed for greater articulation of the debate, since the interventions, although localized in particular arguments, produced effects that extended across the argumentative structure as a whole.

Finally, it should be noted that the overall diversity index increased by almost 6%, driven mainly by increases in ontological and functional diversity and ontological richness. The transformation

shows a greater presence of the governance dimension and a reduction in the values dimension, alongside an increase in the conditional dimension and a decrease in the pragmatic dimension. Taken together, these shifts produce a more diverse, complete, and balanced debate across ontological dimensions and argumentative functions. This greater diversity and completeness expand the possibilities for agreement, while persistent differences in quality and priority constitute the degrees of freedom for reaching a final resolution.

IV. Discussion and Conclusions

The results show that argumentative contradictions generated by artificial intelligence itself can provide a starting point for producing new conceptual configurations without eliminating opposing positions. The incorporation of dual contestability, the reformulation of arguments, and the creation of links between regions of the debate illustrate a shift from relatively isolated oppositions toward relationships that incorporate demands for participation, institutional capacities, and risk management. These transformations are consistent with the hypothesis that adversarial tensions contain possibilities for argumentative innovation.

The adjusted comparison records a 5.87% increase in the overall diversity index, accompanied by increases in dimensional diversity, SKOS richness, and functional diversity. However, normalized semantic distance decreases by 2.98%, which precludes interpreting the transformation as a uniform expansion across all dimensions of diversity.

The methodological contribution lies in combining an external ontology, explicit argumentative transformation operations, and a multidimensional assessment of diversity in a debate previously generated by AI. Nevertheless, the study is limited to one selected debate, one specific intervention, and a size-adjusted descriptive comparison.

In conclusion, the experiment provides exploratory evidence that it is possible to broaden the argumentative and functional variety of an AI-generated debate while preserving recognizable argumentative tensions. Its potential relevance to the development of artificial intelligence encompasses both the reconsideration of prevailing approaches and the exploration of alternatives by diverse technological actors. Determining whether these configurations foster effective agreements, technical innovations, or differentiated development trajectories requires further applications, comparisons, and independent validations.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

<https://doi.org/10.1145/3442188.3445922>

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv.

<https://doi.org/10.48550/arXiv.2108.07258>

DebateLabKIT. (n.d.). syncIALO-raw: A multipurpose synthetic debate and argument mapping corpus [Dataset]. Hugging Face. <https://huggingface.co/datasets/DebateLabKIT/syncialo-raw>

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28), eadn5290.

<https://doi.org/10.1126/sciadv.adn5290>

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2), 321–357. [https://doi.org/10.1016/0004-3702\(94\)00041-X](https://doi.org/10.1016/0004-3702(94)00041-X)

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. <https://doi.org/10.1145/3458723>

Gutmann, A., & Thompson, D. (2004). *Why deliberative democracy?* Princeton University Press.

Habermas, J. (1984). The theory of communicative action. *Volume 1: Reason and the rationalization of society* (T. McCarthy, trans.). Beacon Press.

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference* (pp. 11–15).

Jung, C. G. (1969). The transcendent function (R. F. C. Hull, trans.). En H. Read, M. Fordham and G. Adler (Eds.), *The collected works of C. G. Jung: Vol. 8. The structure and dynamics of the psyche* (2nd ed., pp. 67–91). Princeton University Press. <https://doi.org/10.1515/9781400850952.67>

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. Wiley.

Miles, A., & Bechhofer, S. (2009). *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation. <https://www.w3.org/TR/skos-reference/>

Mosquera, M. (2020). Incrementalismo dialéctico: Un caso de innovación en Hengqin (China). *Gestión y Política Pública*, 29(2), 321–353. <https://doi.org/10.29265/gypv29i2.776>

Mouffe, C. (2005). *On the political*. Routledge.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>

Salma, Z., Hijón-Neira, R., & Pizarro, C. (2025). Designing co-creative systems: Five paradoxes in human–AI collaboration. *Information*, 16(10), 909. <https://doi.org/10.3390/info16100909>

Saussure, F. de. (1916). *Cours de linguistique générale*. Payot.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755–759. <https://doi.org/10.1038/s41586-024-07566-y>

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118. <https://doi.org/10.2307/1884852>

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, 36(2), 381–403. <https://doi.org/10.5465/amr.2009.0223>

Yao, X. (2023). Harmony without uniformity: A way of reconstructing the shared values of humanity. *Frontiers of Philosophy in China*, 18(2), 135–147. <https://doi.org/10.3868/s030-012-023-0008-5>