Position: Beyond Euclidean -- Foundation Models Should Embrace Non-Euclidean Geometries

Abstract:
In the era of foundation models and Large Language Models (LLMs), Euclidean space has been the de facto geometric setting for machine learning architectures. However, recent literature has demonstrated that this choice comes with fundamental limitations. At a large scale, real-world data often exhibits inherently non-Euclidean structures, such as multi-way relationships, hierarchies, symmetries, and non-isotropic scaling, in a variety of domains, such as languages, vision, and the natural sciences. It is challenging to effectively capture these structures within the constraints of Euclidean spaces. This position paper argues that moving beyond Euclidean geometry is not merely an optional enhancement but a necessity to maintain the scaling law for the next-generation of foundation models. By adopting these geometries, foundation models could more efficiently leverage the aforementioned structures. Task-aware adaptability that dynamically reconfigures embeddings to match the geometry of downstream applications could further enhance efficiency and expressivity. Our position is supported by a series of theoretical and empirical investigations of prevalent foundation models. Finally, we outline a roadmap for integrating non-Euclidean geometries into foundation models, including strategies for building geometric foundation models via fine-tuning, training from scratch, and hybrid approaches.

Add: Withdrawal

Paper Decision

Decision by Program Chairs 26 Sept 2025, 09:01 (modified: 26 Sept 2025, 14:16) Program Chairs, Authors

Revisions (/revisions?id=PeDEGRGHdo)

Decision: Reject

Meta Review by Area Chair erHD 11 Sept 00:29 (modified: 26 Sept 2025, 21:09) Area Chairs, Authors, Program Chairs

Revisions (/revisions?id=JGqY6J7JdU)

Ethics:
No ethical concerns were raised.

Strengths:
The reviewers were extremely positive about this paper. They mentioned that it is very well-written and intellectually compelling.

Weaknesses:
The counter position is only discussed quite briefly in section 5. This is a big weakness because, in general, foundational models do not exploit non-Euclidean structures. There is a big argument within the community about implicitly learning the structure (geometry, symmetries, physical constraints, etc) from data rather than explicitly using it in the design. The claim is that standard foundational models are simpler, more general, and more efficient. This paper could have discussed those points more in-depth and given empirical evidence in not-so-specific (aka foundation) models.

Questions:
I would like the authors to discuss the counter position a bit more if possible.

Agreement: 5: strongly agree

Rating: 7: Accept: The paper presents a solid argument about an important issue that remains unresolved and is of importance to at least one sub-area of the NeurIPS community.

Confidence: 4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

Thoroughness: 1: You read the paper abstract but only a cursory read of the paper. No detail checks.
Code Of Conduct Acknowledgement: Yes

Author Survey by Authors

29 Aug 2025, 12:37 (modified: 06 Sept 2025, 00:20) Program Chairs, Area Chairs, Authors Revisions (/revisions?id=BYDAfZtjrV)

1-1 Submission Process: 3
1-2 Next Year: A more clear timeline with clearer instructions/expectations of the author survey would make the whole process less confusing for the authors.

1-3 Future Development: An overall clearer instructions regarding the process after the manuscript is submitted would allow the authors to have a better idea of what to expect.
1-4 Interest: Panel discussions with other position paper authors, Structured debates on controversial topics, Workshops for developing position papers, Mentorship programs for early-career researchers
1-5 Thoughtful: 6
1-6 Supportive: 7
1-7 Technical Aspects Versus Position: 3
1-8 Gate Keeping: 9
1-9 Camera Ready Changes: We will make the following edits in the revision based on reviewer suggestions (see comments for reviewers in Section 3 for details on each point):

Additional Citation:

1. Incorporate suggested citations (based on reviewer PtqG)

Further Discussions:

1. Include additional challenges and potential solutions for implementing non-Euclidean foundation models in addition to the ones mentioned in Sec 4.4, such as geometric optimization algorithms. (based on reviewer PtqG)

2. Discussion of benchmarks and metrics for error analysis in non-Euclidean foundation models (based on reviewer PtqG)

3. Expand the counter argument section to include more critical points, such as nuance on distortion under invariance or Euclidean method alternatives. (based on reviewer 1KYp)

Clarifications:

1. Emphasize that we are arguing for the collective community efforts for realizing large scale non-Euclidean foundation models (based on reviewer PtqG, FKUN)

2. Clarify how embedding dimensions relate to model size and computational resource consumption; Clarify how orthogonal invariance from self-attention results in requirement of high dimensionality for Euclidean models (based on reviewer 1KYp)

3. Clarify how our results in Sec 3.2 reflect the scaling limitations of Euclidean foundation models discussed in Sec 3.1 (based on reviewer FKUN)

4. Emphasize the lack of non-Euclidean libraries for foundation models and how this is a line of research advocated by our position paper (based on reviewer FKUN)
1-11 Submit Again: Probably yes
2-1 Review Adjudication Request1: Yes (please continue with this section)
2-2 Reviewer Anon Id1: FKUN
2-4 Comments To Area Chair1: We'd like to bring to attention that the reviews provided by Reviewer FKUN are not typical of that of an human expert and do not pertain to the assessment of a position paper. For example, the weaknesses raised for downstream performance and computational/memory cost benchmarks in real-world settings are generic review points for a research track submission. We suspect that the review could have been generated with the help of an LLM.
2-5 Review Adjudication Request2: No (you may skip the rest of this section)
2-9 Review Adjudication Request3: No (you may skip the rest of this section)
3-1 Review Response1: PtqG
3-2 Reaction To Review1: 1. **Regarding Implementation Challenges:** We will include the suggested citations in connection to this. Furthermore, in Sec. 4.4, we identified key issues and actionable items, e.g., the computational complexity of non-Euclidean operations and the difficulty of relating embedding distortion to downstream performance. We also noted the need for non-Euclidean libraries, as existing ones are largely tailored to Euclidean models (lines 71–75). In the revision, we will expand this discussion to include further challenges and solutions, such as non-Euclidean optimization.

2. **Regarding Benchmark/Metrics:** Non-Euclidean models can be evaluated with standard metrics (e.g., accuracy, F1, perplexity). For error analysis, prior work has applied non-Euclidean distance to embeddings, showing advantages over Euclidean counterparts, such as improved semantic hierarchy organization (lines 282–288). These findings suggest that non-Euclidean spaces can enhance explainability and provide better reasoning for when and why models succeed/fail. Additional analyses could include measuring alignment of local curvature with the data in curvature-adaptive models. We will incorporate these points into the revision.

3. **Distinguishing from Prior Works:** Most prior work focused on low-dimensional settings, particularly hyperbolic spaces. Few have attempted non-Euclidean models at foundation-model scale. We argue in our paper for extending general non-Euclidean geometry to models with billions/trillions parameters, which is highly underexplored and requires significant community effort. Beyond proposing non-Euclidean counterparts of Euclidean operations (the current focus), we advocate building the full training infrastructure for this scale (e.g., multi-node training and flash attention adapted to non-Euclidean settings), developing and training large-scale non-Euclidean architectures, and analyzing their behavior, which often diverges from low-dimensional models.
3-3 Review Response2: 1KYp
3-4 Reaction To Review2: 1. **Regarding the Counter Position:** In the revision, we will expand on counterarguments, e.g., the nuances of embedding distortion and space invariance from self-attention (as noted by the reviewer).

2. **Regarding embedding dimension:** Our discussion of the Nash Embedding Theorem highlights that non-Euclidean spaces ***do not*** only halve dimensions; rather, the opposite is true. Embedding dimension is **directly tied to parameter count and computational complexity**, making it a key advantage of non-Euclidean space. Euclidean spaces require high dimensions to limit distortion (Theorem 3.2, Fig. 4) and still suffer significant distortion even with arbitrarily high dimensions (Theorem 3.1, Fig. 4), leading to diminishing returns as noted in prior work(lines 62–70).

For the example of $SO(n)$, this is a quadratic increase in dimension for Euclidean space. When $n = 50$, under this framework results in a ***50 times*** increase in parameter count. SVD also introduces additional complexity and potential of numerical instability.

3. **Regarding distortion:** Incorporating orthogonal invariance from self-attention into distortion computation could be informative. We intentionally dismissed invariances from kernel designs—both Euclidean and non-Euclidean—to isolate the representational capacity of the spaces. Orthogonal invariance is addressed in Sec. 3.3, where prior work shows Euclidean LLMs use orthogonality to encode hierarchies. Yet, since R^d provides only d orthogonal dimensions, Euclidean spaces must rely on high dimensionality to capture the expansive semantic hierarchies of language [108, 109].

As noted in Sec. 4.4, current research does not directly link distortion with downstream performance. While some prior works touch on this in language or graph embeddings, we propose this as a future research direction in Sec. 4.4.

We will include all of the above in the revision.
3-5 Review Response3: FKUN
3-6 Reaction To Review3: 1. **Regarding W1/Q1:** We'd like to remind the reviewer that, as a position paper, our aim is to advocate for non-Euclidean foundation models, not to propose a new non-Euclidean architecture. Exploratory prior works shows that non-Euclidean neural networks models improve downstream performance such as retrieval, question-answering, and graph tasks [158, 44, 35, 26, 159]. Our results in Fig. 4 reinforce these advantages, demonstrating the superior scaling potential of non-Euclidean spaces: 4-dimensional hyperbolic space shows lower distortion than 50-dimensional Euclidean embeddings, and distortion continues to decrease for hyperbolic and hybrid spaces but plateaus for Euclidean space. These results suggest the potential for non-Euclidean foundation models to excel in downstream tasks by leveraging the geometric properties inherent to real world data.

2. **Regarding W2/Q2:** As noted in Sec. 4.4, prior works have compared Euclidean and non-Euclidean operations [e.g., 158], demonstrating the potential of these methods to scale to billions or trillions of parameters. While some efforts have introduced frameworks for implementing non-Euclidean foundation models [62], a significant gap remains compared to Euclidean ones. Addressing this gap is precisely one of the research directions we advocate in this position paper.

3. **Regarding W3/Q3:** In Sec. 4.4, we noted key challenges and preliminary investigations, including dependence on ground-truth data geometry, computational demands, and isolating distortion effects. Early work shows that manifolds better aligned with data structure can improve performance in graph tasks and word embeddings [55]. We propose relating distortion to performance as a future direction for advancing model design and analysis.

For δ-hyperbolicity, we also note additional empirical and qualitative results, including the scale-free property of LLM token distributions and embedding distortion from misaligned geometry.

−
=
≡

Official Review by Reviewer PtqG 13 Aug 2025, 23:08 (modified: 16 Aug 2025, 02:47)
Program Chairs, Area Chairs, Reviewers Submitted, Reviewer PtqG, Authors Revisions (/revisions?id=RvbCTz5VfZ)
Ethics: NO or VERY MINOR ethics concerns only
Position: Yes, the paper argues for or against a position related to machine learning.
Summary:
The paper argues that foundation models should adopt non-Euclidean geometries (hyperbolic, spherical, and mixed-curvature manifolds) instead of relying exclusively on Euclidean space, due Euclidean limitations in capturing complex data structures and interactions. The authors argue that this shift addresses three critical aspects: enhanced representational capabilities, improved adaptability to diverse geometric structures, and better scalability. They support their position with theoretical analysis demonstrating Euclidean embedding limitations, and providing empirical evidence that reveal hierarchical structures in token embeddings, and presented a roadmap to adopt non-Euclidean geometries based on fine-tuning of existing models, training from scratch, and hybrid approaches.

Author Identification: No.
Support: 3: good
Significance: 3: good
Presentation: 2: fair
Context: 2: fair
Discussion: 3: possibly
Alternative Position: Yes, and alternative positions are well-considered and addressed by the argument
Strengths:
1. The paper addresses a topic of clear interest to the NeurIPS community. Its call to shift toward non-Euclidean geometries is timely and relevant, bringing together known challenges into a unified narrative that could help bridge discussions across subfields and spark broader community debate.
2. Good theoretical foundation with formal theorems demonstrating current Euclidean embedding limitations, complemented by empirical analysis of non-Euclidean structures in existing foundation models. The paper provides a detailed roadmap with three different implementation approaches.

Weaknesses:
1. While the paper presents a sound theoretical motivation for adopting non-Euclidean representations, it overlooks important prior works that directly address the mathematical and computational transition from Euclidean to non-Euclidean spaces, as well as the practical challenges of operating in such geometries. Previous studies have examined difficulties in learning within curved spaces, including numerical stability issues, optimization on manifolds, algorithmic and implementation-level constraints, gradient instability near manifold boundaries, and integration challenges with existing architectures, as well as trade-offs between embedding dimension and computational complexity. Please cite some of these works (see below).
2. The lack of standardized evaluation frameworks for measuring both performance and interpretability in non-Euclidean settings—and for systematically diagnosing model successes and failures—represents a critical gap that should be discussed. Without consistent benchmarks, error analysis protocols, and domain-relevant metrics, it is difficult to validate results, compare methods, or build trust in these models. This is needed before non-Euclidean approaches can be reliably adopted and deployed at scale.

Questions:
While the three proposed approaches in the roadmap are outlined, the paper would benefit from a deeper discussion of their practical implementation, the specific challenges involved and how to address them, and potential pitfalls. Providing concrete examples of domains where each approach is most applicable, along with scenarios where they might struggle, would make the roadmap more actionable, realistic, and persuasive.

Given that aspects of your position are already broadly accepted, what specifically distinguishes your proposal from prior work? Is the novelty primarily in its focus on foundational models?https://arxiv.org/pdf/2206.00606 (https://arxiv.org/pdf/2206.00606) https://arxiv.org/abs/1611.08097 (https://arxiv.org/abs/1611.08097) https://arxiv.org/abs/2504.08896 (https://arxiv.org/abs/2504.08896) https://pmc.ncbi.nlm.nih.gov/articles/PMC12315666/ (https://pmc.ncbi.nlm.nih.gov/articles/PMC12315666/)

How do you envision evaluating the proposed position beyond theoretical justification? What metrics, benchmarks, or protocols would be used to assess performance and the success of adoption in real-world deployments?

Are there specific domains, data regimes, or architectural contexts where your position might not apply, underperform? If so, can you discuss what safeguards/strategies would you propose to mitigate these risks while preserving the benefits of the approach?

Agreement: 4: agree
Rating: 7: Accept: The paper presents a solid argument about an important issue that remains unresolved and is of importance to at least one sub-area of the NeurIPS community.
Confidence: 3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.
Thoroughness: 3: You read the paper carefully but did not check all of the details.
Code Of Conduct Acknowledgement: Yes

Official Review by Reviewer 1KYp 09 Aug 2025, 22:33 (modified: 16 Aug 2025, 02:47)
Program Chairs, Area Chairs, Reviewers Submitted, Reviewer 1KYp, Authors Revisions (/revisions?id=vMcF0lNS7V)
Ethics: NO or VERY MINOR ethics concerns only
Position: Yes, the paper argues for or against a position related to machine learning.
Summary:
This paper argues that training foundation models using non-Euclidean geometries could be advantageous compared to using Euclidean geometries.

The paper provides in **Section 2** a detailed review on prior literature on learning in non-Euclidean spaces. **Section 3** provides several arguments for learning in non-euclidean spaces. Discusses the potential of non-Euclidean spaces in training foundation models and provides a brief empirical validation of the distortion introduced into embedding spaces when resorting to Euclidean spaces. **Section 4** discusses how the field could obtain non-euclidean foundation models. **Section 5** discusses VERY briefly arguments against learning in non-Euclidean spaces.

The paper is very well written, thoroughly researched, and I am sure that it's publication will spark a vivid discussion on the utility of non-Euclidean geometry in training foundation models.

Author Identification: No.
Support: 3: good
Significance: 4: excellent
Presentation: 4: excellent
Context: 4: excellent
Discussion: 4: very likely
Alternative Position: Yes, and alternative positions are well-considered and named but not addressed
Strengths:
- Position is clearly expressed and discussed in detail.
- Interesting empirical analysis of token embeddings.
- Paper is well written, properly supported by citation, and provides a coherent lines of argumentation.
- Figure 5 is pretty nice.

Weaknesses:
- The **counter position is only discussed quite briefly** in Sections 4.3, 4.4., and 5. Here, the authors could have been a bit more critical.
- **C1:** The authors point out that using non-Euclidean spaces may only reduce the embedding dimension by half. **However, why should we care about the embedding dimension?** Consider the case of learning 3D rotations. We can learn in Quaternion space or in rotation matrix space. There are several works that show that learning in R^9, using a SVD map to SO(3), is favorable to learning with Quaternions (or other low-dimensional representations). Simply put, learning in Euclidean space can be favorable as long as we map the resulting representations onto a manifold that respects important in-variances. In R^9 distance metrics are distorted but this is less of a concern compared to issues arising from double cover in Quaternion space.

Questions:
- **Q1 (related to C1):** In Line 124, the authors argue that "low-distortion embeddings are often only possible in high-dimensional spaces". However what speaks against learning in high-dimensional spaces?
- **Q2:** In Section 3.2 the authors argue that non-Euclidean spaces produce higher quality embeddings compared to Euclidean spaces. Here, "quality" is measured through the average point-wise distortion. However, self-attention layers perform a dot product operation which imposes an in-variance onto the space, **shouldn't this be considered when measuring distortion in embeddings?**
- **Q3:** How does the distortion of an embedding space relate to the performance of the network?

Agreement: 2: disagree
Rating: 8: Strong Accept: The paper presents a strong argument about an important issue that ought to be discussed and is of importance to a sub-area within the NeurIPS community.
Confidence: 3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.
Thoroughness: 3: You read the paper carefully but did not check all of the details.
Code Of Conduct Acknowledgement: Yes

Official Review by Reviewer FKUN 07 Aug 2025, 22:00 (modified: 16 Aug 2025, 02:47)
Program Chairs, Area Chairs, Reviewers Submitted, Reviewer FKUN, Authors Revisions (/revisions?id=4XC72zhTa7)
Ethics: NO or VERY MINOR ethics concerns only
Position: Yes, the paper argues for or against a position related to machine learning.
Summary:
This position paper argues that the Euclidean geometry underlying current foundation models (e.g., LLMs, ViTs) fundamentally limits their ability to represent complex real-world data structures such as hierarchies, cycles, and symmetries. Through theoretical analyses and empirical evidence, such as δ-hyperbolicity measurements and distortion-vs-dimension tradeoffs, the authors demonstrate that non-Euclidean geometries (hyperbolic, spherical, and mixed-curvature) offer lower-distortion, more scalable, and semantically faithful representations. The paper proposes a roadmap for integrating these geometries into foundation models via fine-tuning, pretraining, and hybrid architectures, and argues that embracing non-Euclidean structures is essential to sustain scaling laws, improve generalization, and reduce inefficiencies in large-scale AI systems.

Author Identification: No.
Support: 3: good
Significance: 3: good
Presentation: 3: good
Context: 3: good
Discussion: 4: very likely
Alternative Position: Yes, and alternative positions are well-considered and addressed by the argument
Strengths:
S1. This paper convincingly highlights the mismatch between Euclidean geometry and the structural nature of real-world data (hierarchies, cycles, symmetries), particularly in NLP, vision, and natural science domains.

S2. The discussion on embedding distortion, Markov convexity, and limitations of the Nash Embedding Theorem is rigorous and relevant. The authors provide theoretical analyses to clearly show why low-distortion embeddings in Euclidean spaces are not always possible.

S3. The authors provide empirical evidence such as δ-hyperbolicity of token embeddings and distortion plots across geometries to support their theoretical claims, showing that hyperbolic/mixed geometries are more compact.

S4. The three-pronged approach (fine-tuning existing models, training from scratch, and hybrid modeling) is comprehensive. Suggestions such as geometric LoRA, non-Euclidean attention, and geometry-aware MoE seem to be practically grounded.

Weaknesses:

W1. The empirical section lacks direct downstream task comparisons (e.g., LLM or vision model performance using Euclidean vs. non-Euclidean architectures). Claims such as improved scaling or reduced hallucinations are not validated.

W2. The paper acknowledges computational overhead but lacks a clear benchmark comparison for training cost, inference latency, or memory consumption in real-world settings, as non-Euclidean operations can also incur high computation complexity.

W3. This work is somewhat overreliant on δ-Hyperbolicity. While δ-hyperbolicity is an insightful metric, its sensitivity and interpretability are not thoroughly discussed. Moreover, it is unclear how to relate δ values to model performance or the quality of representation

Questions:

Q1. Regarding W1, can you provide empirical results showing improved performance (e.g., perplexity, BLEU) when using non-Euclidean architectures in foundation models?

Q2. Regarding W2, are there existing frameworks that support scalable non-Euclidean training at the scale of LLMs? How do they compare with DeepSpeed or FlashAttention?

Q3. Regarding W3, do you have any preliminary results or hypotheses on how distortion might correlate with model performance or representation quality?

Agreement: 4: agree

Rating: 5: Borderline accept: The paper presents a relevant position, and the reasons to accept outweigh reasons to reject, e.g., unclear reasoning or limited support for the claims.

Confidence: 4: You are confident in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work.

Thoroughness: 3: You read the paper carefully but did not check all of the details.

Code Of Conduct Acknowledgement: Yes