

A Log of AI Use in Mathematical Research

Daniel Weser

Introduction

This document presents a log of personal experiments (spanning the period Oct. 2025 – Jan. 2026) evaluating the capabilities of frontier language models in the context of geometric analysis research. These questions ultimately fell into three distinct categories (in order of their appearance below):

1. **Auditing Reasoning:** Testing the model’s ability to handle high-level abstractions and identifying specific failure modes.
2. **Accelerating Computation:** Offloading routine but tedious symbolic derivations to test improvements in research efficiency.
3. **Proof Exploration:** Using a sort of “curriculum prompting” to guide models through increasingly complex geometric settings to uncover proof strategies.

My findings suggest that current models function close to the level of an advanced undergraduate or early graduate research assistant: capable of algorithmic computations but prone to fundamental conceptual errors when handling high-level abstractions. As a result, I found that success required heavy expert supervision, where I, as the researcher, acted less as a prompter and more as a rigorous advisor.

Moreover, I sensed a strong parallel between the current models and my experience learning these fields as a student. One of the greatest drivers of the **significant learning curve** for mathematics is the deep subtlety of each line in a proof, for example regarding notation and conventions. Looking ahead to Calculation 1, the model did not demonstrate a conceptual understanding of the restriction of a tensor, causing its responses to read as reproductions of manipulations of symbols. In parallel with the analogy of the learning curve, students will read research papers which, for example, simplify notation or assume familiarity with a concept/argument when they feel it should be clear to the reader, but then they will mimic lines of these proofs without the deeper understanding, leading to analogous conceptual errors.

This parallel is further exemplified in Calculation 3, where, for one concrete goal, I guided the model through a series of increasingly complex settings. Not only is this what I would do for a student, but, in fact, this is precisely what one does in research: begin with the simplest relevant case, and increase complexity until reaching the desired setting, using the previous setting’s insights for the current one. This both helped the model and solidified to me that this is a robust, helpful approach one can take (and which has been fruitful for me in many other instances outside the scope of this note).

Calculation 1: An alternate proof of the preservation of the Lagrangian condition under Mean Curvature Flow

Problem

The standard way to prove that the Lagrangian condition is preserved under Mean Curvature Flow goes back to [Smoczyk, Prop. 1.8], in which one computes the *pointwise* evolution of the (pointwise) norm-squared of the restriction of the symplectic form ω to the evolving submanifold. Then, after obtaining a suitable parabolic equation, one can apply the maximum principle to conclude. The main drawback of this argument is that the equation one obtains is a heat equation, and the Laplacian introduces a significant number of curvature terms which must be handled very carefully.

My collaborator and I recently found a much simpler argument which is energetic. Instead of deriving a parabolic equation for the *pointwise* evolution of ω , one may look at its squared L^2 -norm, take only the time derivative of this quantity, obtain a much simpler evolution, and apply the standard Gronwall argument to conclude the argument.

I asked Claude to reproduce this approach, giving it instructions at the beginning of exactly how the argument should go.

AI Response

Claude obtained the correct initial setup of the argument, but it immediately made errors, including **circular reasoning** (i.e. assuming the conclusion that ω indeed vanishes in order to get the argument to work). Interestingly, there was one more involved pointwise computation needed, which Claude did incorrectly, but the momentary pointwise focus prevented Claude from **remembering the context** of how that step fit in to what it had already obtained. After 5 corrections and hints, it became clear I would need to provide the complete argument step-by-step due to fundamental conceptual errors detailed below.

Accuracy

The initial computation of the integral quantities was correct, but nothing significant after that step was correct. I had to fix multiple basic mistakes, stop it from circularly reasoning, and help it remember how each smaller piece fit into the bigger picture.

However, the most significant error it repeated was a fundamental misunderstanding of the meaning of the restriction of a tensor to a submanifold. When one restricts a tensor, it is not only that the tensor is now evaluated at points on the submanifold, but, most importantly, it is restricted to vector fields *tangent to the submanifold*. In these computations, normal vectors repeatedly appear, including as arguments in the symplectic form. When this occurs, such a term is no longer the restriction of the form to the submanifold, since it is now being evaluated in a non-tangent direction. Claude did not understand this distinction and used these terms involving normal vectors as though they were still the restriction of the form.

Insights

Claude's conceptual errors regarding the restriction of a tensor to a submanifold were by far its biggest obstacle to success. While not unique to Differential Geometry, authors in this field can be loose with notation, relying heavily on the reader to work to understand the subtleties implicit in their arguments, which creates a **significant learning curve** for anyone approaching this subject.

Consequently, the main insight I obtained from this exercise with Claude is that these learning curves are a great obstacle which these models still need to overcome. (However, see Calculation 2 below where this issue does not persist.)

Additionally, Claude struggled with the context-switching between the “local” and “global” components of these proofs, and it was imprecise with computations that I would guess it could not directly query. These issues reinforce that these AI models are currently best for validating small computational chunks rather than complete arguments.

Time Saved

None, as it was much easier and quicker to compute by hand.

Calculation 2: Derivation of the equation for geodesics on ellipsoids

Problem

In RMS-normalization, one allows weights in the computation of the l^2 -norm of a point (yielding an anisotropic norm), so that the normalization corresponds to projecting onto an ellipsoid. To extend geometric optimizers (e.g. ADAM, AMSGRAD, ADAGRAD) to ellipsoids with accuracy, one would ideally like an explicit formula for the exponential map, as one has on the sphere. Unfortunately this is not possible, but one can still derive and approximate the equation for geodesics.

(Note: while this is not computationally efficient in practice, I found it an interesting exercise to understand the discrepancy between the choice of retraction map and a direct approximation of the exponential map.)

I asked Claude to compute the geodesic equation in two different coordinate systems: the first was using coordinates on the ellipsoid itself, and the second was in terms of coordinates on the sphere, as one can see the ellipsoid as the image of the sphere under a simple mapping. Finally, I asked it to compute an approximation of the solutions of this equation.

AI Response

Claude produced the correct derivations of the geodesic equations on its first try. These computations do have a standard “algorithm” one follows – choose coordinates, compute the Christoffel symbols, etc. – and Claude closely followed this roadmap. Additionally, it clearly explained its steps along the way and demonstrated a clear understanding of what it was doing and why. Finally, the approximation it produced was a simple application of Runge-Kutta, which is exactly what I would have done.

Accuracy

100%

Insights

These algorithmic derivations seem perfectly suited for current AI models, and it can definitely save users time both in needing to look up formulas again and also the pain of having to compute these routine quantities over-and-over as they consider different cases/scenarios in their work. Moreover,

the fact that Claude wrote the formulas in general form before computing allows users to refresh themselves in addition to time saved by not having to do the computations by hand.

Comparing to Calculation 1, I believe that another major reason Claude performed so well in this case is due to the lack of conceptual subtlety in the nature of the question. There were no cases in which the metric, Christoffel symbols, etc. were anything other than their definitions, say, on Wikipedia – the argument was conceptually the same every time, regardless of the choice of coordinates.

Time Saved

While it only saved me 3-5 minutes of work, it becomes tedious to repeatedly compute these quantities, so the mental savings were far greater than the time saved.

Calculation 3: Comparison theorems for intrinsic vs. extrinsic distances

Problem

In the course of one of my ongoing research projects, I became interested in extensions of the deep work of [Belkin & Niyogi] on Laplacian-based learning methods. In this paper, the authors first define a weighted graph induced by a sampling of a submanifold M of Euclidean space and then prove the spectral convergence of the induced graph Laplacian to the Laplacian-Beltrami operator on M .

One step in their proof can be precisely formulated as follows: given a closed submanifold M embedded in Euclidean space and two points on M sufficiently close together, prove an estimate comparing the geodesic distance on M to the Euclidean distance of these two points. The proof in Belkin & Niyogi uses a number of properties unique to the Euclidean setting, so I tested Claude’s proof exploration capabilities by progressively increasing the complexity of the setting for the estimate as follows:

AI Response

- 1. (Euclidean, no constraints):** Claude used a standard argument, Taylor expanding the geodesic in terms of the tangent initial velocity of the curve, the second fundamental form, and then higher-order terms. This required computing derivatives of the geodesic, which heavily relied on the ambient Euclidean space having trivial geometry.
- 2. (Euclidean, no derivatives):** I next required Claude not to use any derivatives in its argument, and it successfully re-derived the estimate by instead using projections. More specifically, it worked out an argument where the distance comparison arises from projecting the ambient geodesic (straight line between the two points) onto M and bounding the differential of the projection, which is a notion of curvature.
- 3. (General Riemannian):** Finally, I changed the setting to a submanifold M of an ambient Riemannian manifold N and had it reproduce its second proof in this setting. It followed the previous argument closely, essentially only needing to take a tubular neighborhood of M and use the sectional curvature of N .

Accuracy

The proofs it sketched were 100% accurate. However, it did not rigorously prove any step, which always opens up the possibility for subtle errors.

Insights

Guiding Claude through increasingly complex settings removed some issues I had previously run into with AI models when asking these types of arguments. For example, in the past I had asked ChatGPT essentially the third question I asked Claude in the non-Euclidean setting, however *without* increasing the complexity as above. Its answers were filled with trivial mistakes, such as confusing points on the manifold with tangent vectors, treating the manifold like a vector space, and more. After this experience with Claude, I re-prompted ChatGPT using the above format, and it performed excellently. As a result, I believe that my above methodology of sequentially introducing complexity works very well for these open-ended explorations of proof strategies.

Time Saved

This was exploratory in nature, so not applicable. While I had numerous ideas of my own, I was mostly curious what Claude would come up with and how that would compare to my own ideas. They mostly aligned, but Claude's particular strategy of how to use the projection seems better than mine.