REFLECTIONS SUMMER 2026

by Isidor Seppälä

July 12 – August 21, 2026

Summary:

Prologue: Personal context.

In the summer of 2022, I had a dream about programming. The day after I looked up a free online course in programming and started to learn some Python. This was my first ever "technical interest". At the time I was studying music at a gymnasium in Gothenburg, and only after a few days of programming, I began seeing it as a "plan B" to pursuing a career in music, something I had been pursuing since 2019.

To me, the change initially looked academic, as I moved from music towards computer science. But internally, it felt much larger. It took longer to understand that I was not really finding a second profession, but was discovering a new way of relating to knowledge. The transition was merely a tiny part of a larger change in how I had begun thinking about what purpose means to me.

I don't believe there was anything inherently special about why I decided to study computer science, as I quickly began studying mathematics and physics simultaneously. It was more about the act of learning, and it began partly as a response to a more existential wonder, related to purpose. I found myself thinking about the vastness of our universe, which feels both beautiful and weird, and about our place in it. With this, I began thinking about our temporariness as individuals, but about the effect we have as humanity, and how that might look like in the future. One thing was certain to me: knowledge outlives individuals. The purpose I found myself constructing was that therefore, I should learn, build, contribute, and help us increase understanding. I wrote the following short point which in many ways anticipated my current research philosophy:

> You must want to understand more than you want to seem smart. You must be curious and creative. You must want to tackle problems and make no progress. You must want to live in ambiguity. You must want to

> explore ideas others think are wastes of time. You must never stop learning. You must understand.

Towards the end of 2022, and the beginning of 2023—my last year of gymnasium—I was building an entire conception of purpose around knowledge, learning what humanity had already discovered, eventually contributing something of my own, and using that knowledge to help us solve problems. This kind of drive is what led me to my interests in programming, and later, physics, mathematics, and entrepreneurship. One part of me was curious and patient (understand more than you want to seem intelligent, pursue questions even when they yield no progress, and learn to live in ambiguity) and the other was more punitive (every hour had to serve a purpose, every discomfort had to become growth, and every failure had to be overcome through greater discipline). This response of dedicating yourself to learning and contribution felt extremely liberating. It made the universe inspirational and never overwhelming or scary. But it also made learning feel morally necessary. Curiosity was no longer only something I enjoyed but also evidence that I was using my existence correctly. The problem arose when that punitive voice made contribution the condition under which curiosity was allowed to count, rather than ensuring that curiosity became contribution. I understand this distinction better now, but I still move between the two. I think because purpose is uncertain, I attempted to make it actionable through self-directed study, building, and contribution. Because this direction was carrying the weight of an existential answer, failure within it never feels merely practical. It easily begins to *feel existential*. A difficult research day can still transform from a technical problem into a judgment about whether I am becoming the person I currently intend to become. A slightly less-than-expected score on an exam can still become a day of questioning my identity.

In 2023, I finished my gymnasial studies in music. I had learned to treat discomfort as evidence of growth. That belief helped me persist, but it also made me suspicious of paths that were accessible and ordinary. A path seemed more meaningful if it required escape or defiance. This was part of why entrepreneurship initially resonated with me, and in May 2023 I started a social media marketing agency. I invested hours into learning about marketing, sales, and leadership. At the time my main goal was to make some money *independently*, which I interpreted as "capable of becoming an entrepreneur". However, it would probably be much safer to instead invest all that energy into finding a summer job, but I refused. I refused because finding a summer job was doing something ordinary. I eventually abandoned this first business attempt of mine, but not only because I was rejected and couldn't get any sales. Even at the time, I wrote that cold calling companies to sell Facebook ad strategies felt disconnected from the kind of life I was trying to build. I believed the work was misaligned, although I cannot fully separate that realization from the discouragement of failing to find clients. On a deeper level, I wanted entrepreneurship to mean solving consequential problems through research and technology. I just didn't know what that would look like, but I already knew that social media marketing was not it.

Later in the fall of 2023, I left to study in Mountain View, California, to chase many of my dreams. It quickly became a stage on which my ideas of purpose could become visible, promising intellectual expansion, geographic escape, tech-startup culture, and proximity to success narratives. In the back of my head was the idea of being the one who did something radically different and exceptional. The young, self-driven person who moved to another country, studied, excelled at their courses, got into prestigious universities, built a successful tech company in Silicon Valley, all of that. I felt like I had set myself up for this success, being in the right environment at the right age. I had an "I-will-prove-you-all-wrong" part of myself where through external validation I *will* prove that my perspective on purpose is valid. It is never clear who "you" is. Perhaps it was someone specific or maybe it was everyone. What matters is that even my own declarations of service and curiosity contained some sort of idea of an audience. I did not only want to understand and contribute. I wanted the contribution to make me undeniably visible. It would be a lie to say that "I'm not like I used to be" in this regard. The desire has changed form rather than disappeared. I increasingly want recognition to follow from work that others genuinely use and value, but I still want the recognition.

After a couple of months I realized I wouldn't be able to afford the whole 4 years of studies to receive my Bachelor's degree. Even if I could, I would never have enough to pursue higher degrees or to spend time looking for jobs in the US. I thought at the time that my absolute worst case scenario would be to go back home to Gothenburg, and have to "start over". I applied for some jobs in the US but stood no chance without a degree and as an international student. I even considered moving to another country where education was cheaper, but after a year of studies in the US, I inevitably went back home in the fall of 2024. This is where I started to seriously think about failure, and what this return meant to me and the ideas of purpose I had built up. Along the last 2 years I sometimes struggled with comparison. The whole idea and dream I had in Mountain View was built on comparisons, modeled on biographies of unusually young founders, researchers, and students. The painful part was not about wanting their outcomes, but about what me failing at seemingly following their path meant about me. I still think there was something real to grieve. I had found an environment I wanted to remain in, and had already begun living mentally inside a future that the these financial limitations now had abruptly removed. Understanding the meanings I projected onto the return does not make the loss imaginary.

I did not allow the return to remain a loss for very long, as part of my purpose was tied to movement and contribution. So a setback could not remain unresolved for long. Almost immediately I decided

I would build a second company, earn my own way, and prove that I had not needed the path that had closed. The plan was partly sincere and partly defensive. It gave me somewhere to direct myself before I had to understand what coming home could mean to me. This time I was going all-in on web development as I had an interest in programming. But after a couple of months, after 2 demo calls and sales pitches, 2 no's for answers, and 1 free website in return for a case study, I moved on. Again, I thought website development wasn't for me. I didn't like building websites for companies. On my free time I was studying mathematics, physics, and deep learning. In the first half of 2025, I started to really develop my relationship with mathematics on a deeper level. I was pretty much only reading textbooks and reading about various theories during these months, not developing websites and learning web-design.

Moving forward into the fall of 2025, I was reading linear algebra, differential geometry, real analysis, abstract algebra, general topology, probability theory, yet on paper I lacked the high-school mathematics required to enter a Swedish university program because I had studied music in gymnasium. The mismatch was absurd to me. My inner and external lives had become misaligned, and realizing this hurt a lot. I felt transformed by what I had learned, yet pretty much none of that transformation was visible. I had returned to the room where I had grown up, and on paper I still lacked the qualifications needed to begin the degree toward which I felt I had already been moving for years. The pain was not that my years in music had been meaningless, but that the path through which I had genuinely developed now appeared, institutionally, as the reason I was heavily unprepared.

In October 2025, I entered a so-called "base year" so I could formally apply for a mathematics program at the university-level in the future. During the first half of this base year I made my third business attempt, where I built a study app. SaaS felt closer to what I wanted to do as an entrepreneur, and this time I actually made some money! After about 5 months of launches and updates, feedback iteration, confusion and struggle, I got my first subscriber. In total I made about $50 before I decided this business model wasn't going to hold in the long run.

> What if it all just fails? Research, business after business attempt, all visions you have, all goals you set, no matter how humble you try to be. No matter if you try being all about the direction and journey, no matter how hard you try to tell yourself that you do this because you love it and because you learn, how failure is good because you grow, what if it all just never works out? No matter how much your goals change, how your philosophy changes, what if you never satisfy your own construction of purpose? You hear about all these legends. In some sense they inspire you, in other senses you envy them. What if they had something fundamental that you just cannot reach. What if you are limited, in the sense that your ambition slowly turns into delusion, more and more, the harder you try pushing towards those ambitions? Maybe all feelings of growing based on failures is just our minds way to accept defeat without us going insane?

This was on the night from which this prologue first emerged. The fear is not that I have learned nothing. It is that this internal change may never become externally consequential, and in some sense, only "true to me". That is perhaps beautiful and invariant, or something genuine to fear. Probably a balance of both, but I don't know. However, I fear a world where I continue becoming more capable, reflective, and more certain that I am approaching something meaningful, while never producing work that matters to anyone else. I fear remaining indefinitely in the state of almost contributing.

I am beginning to understand why setbacks become existential so easily for me. In 2022, I did not only develop new interests. I thought about my own views on purpose around learning, expanding, discomfort, and contributing beyond myself. That gave me direction, but it also made every project and idea carry more than its own weight. A business is never only a business. A textbook is never intended to solely pass a class. California was never only a place to study. A university was never only a university. Research is not only research. Each becomes an opportunity to prove that my purpose is real, and each setback can therefore appear to disprove far more than it reasonably could.

I do not think the honest conclusion is that I have learned to stop caring about prestige, recognition, achievement, or failure. I have not! Nor do I think that every setback was necessary, or that returning was where I was always meant to end up. I understand only that I repeatedly tied my actions to an answer about why my life should matter, and then made their outcomes responsible for preserving that answer. I do not know whether the research I want to pursue will be meaningful to others or whether I will build a successful company. I still want these outcomes, and I do not want to disguise that desire as detachment. What I am beginning to question is whether any bridge between purpose and external visibilities, whether a publication, institution admission, company, or departure, could prove the purpose itself. Maybe they should be separated, and maybe one's self-defined notion of purpose is not something "to be validated" by the external world? But building bridges, stemming from one's purpose to the world is certainty part of *my purpose*, because contributing to knowledge and being useful is definitely part of mine.

Perhaps coming back means returning not only to Gothenburg, but to the uncertainty I tried to solve through motion. The purpose I constructed gave me a direction. But it is constantly changes and it will never give me any form of guarantee. I have not moved beyond that, but can only see it more clearly now.

Preface.

During the summer of 2025 I thought to myself: “In one year, I want to have published my first research paper”. For a while I had been wanting to get into research, and that goal had given me direction in the months prior, where I had solely focused on independent study. During the last quarter of 2024, up until the summer of 2025, I had been re-studying and *re-exploring* deep learning, physics, and mathematics. My long-term goals with this independent studying was to prepare myself for research and explore applications and entrepreneurial opportunities where “doing research” *is* “doing business” and vice versa. With this I mean practical solutions to real problems, fundamentally built upon research contributions, that can be made profitable.

Along this journey, in late July 2025, I felt mature enough to finally “get into research”. No more textbooks, no more reading just to learn, and no more clear goals like “finishing a chapter” and “completing exercises”. I needed a question—a question that sparked excitement, something I felt was important and something I was curious about. Machine learning felt like a natural field to begin this journey in. At the time, it was the field I was most familiar with, and when it came to research, it also felt somewhat more forgiving than other of my interests such as physics or mathematics. But more importantly, I had a particular curiosity about the nature of AI models: We don’t have a good theoretical understanding of how large neural networks and AI systems work. I knew about empirical phenomena like emergence, grokking, and double-descent, and the fact that they don’t have well-established theoretical explanations excited me. Not only was I excited because of their mystery and our nescience, but because it seemed to involve the act of *theory-building*.

During my independent study of mathematics, I read books about many theories, but topology and abstract algebra stood out to me. I quickly felt a type of intrigue that I hadn't felt for other fields like analysis, probability theory, or classical differential geometry of surfaces and curves. I guess this was my first hint that I personally am more of an algebraist rather than an analyst. I like to understand concepts and theories based on their most fundamental, *axiomatic structures*. Probability theory felt more interesting when I turned to measure theory, where I got to experience the foundations that probability theory is based on (or at least is made rigorous by). Differential geometry felt more interesting when I started exploring generalizations to topological manifolds and diffeological spaces. Linear algebra felt more interesting when I started learning about tensor algebra, and even more interesting when learning how to apply the theory of tensors on manifolds! Real analysis felt more interesting when I realized how the axioms of the real line works and how various theorems use these axiomatic foundations. Axiomatic constructions, building and refining mathematical abstractions, applying them for scientific use and other applications, and the general work of theory-building has been the *flavor* of my first year of research.

With this said, I turned to a subfield of machine learning, namely interpretability, with the desire of combining my mathematical interest in theory-building with my prior knowledge of machine learning, in order to understand the black box problems of AI. In this reflection, I want to trace a few themes that have shaped how I think about research, mathematics, creation, and contribution over the past year:

1. My first serious experience with research: how my first year of research has looked like, and how I am beginning to form a research philosophy of my own.

2. My experience with category theory: why I was drawn to it,

and how it has changed my view on mathematics.

3. My developing view of deep learning as a scientific subject: what a more explanatory and predictive theory of learning might require, where categorical deep learning (CDL) fits into that picture, and some questions I am interested in exploring.

4. My first movement toward pure mathematical research: how questions arising from CDL led me into (categorical) differential geometry, o-minimality, and diffeology, and also to think more carefully about what kind of mathematical person I am.

5. My changing view on entrepreneurship: from various business attempts to trying to understand the harder distinction between something I want to build and a problem that genuinely exists for someone else, and how research, mathematics, and entrepreneurship might eventually strengthen one another.

6. My relationship with art and expression: how I have stopped seeing my pivot from music toward mathematics and science as leaving one form of creativity for something else, and how art has helped me appreciate the value of creating things that do not first need to justify themselves through usefulness, recognition, or any form of external validation.

1. My first year of research.

Continuing with my interest in interpretability and theoretical work of deep learning, capable of *explaining* empirically mysterious phenomena, I began freely asking some questions: What are emergent capabilities? Does there exist a set of empirically valid assumptions, in the form of mathematical axioms, such that we can derive the scaling laws from them? The latter question was my guiding light for the first 3 months. If I were to summarize my learning process, then I would say that I entered research expecting the task to be solving difficult problems, but discovered that much of research consists of learning how to see, formulate, navigate, and care about problems.

An important lesson I learned is that research is about distilling a broad direction. I started with a general question on the "axioms of deep learning", wanting to understand a potential "mathematical theory of emergence". But it was hard to know what to do with these questions. They are almost more visionary than concrete, well-stated problems. While they gave me something to work *towards*, they didn't give me something practical to work *on*. Thus, I believe a mistake I did early on was trying to forcefully build such theory before distilling the direction more. For example, instead of trying to reverse-engineer a solution, looking solely at the notion of emergence and trying to build a theory around that, I should have focused on the fundamentals. I should have looked into what we even need to state something about emergence and scaling laws in deep learning and how such components might be modeled mathematically: What is a "neural network"? What do we mean by "backpropagation", "training a network", "generalization", and so on? It isn't clear at all whether these questions even are meaningful, as any AI engineer or deep learning researcher probably have an answer. However, I don't think there's a correct and universal answer to these questions. There can exists many definitions of the same thing, doing different kinds of work and allowing us to explore different aspects of it. It depends on what level of abstraction we want or need them to exist on.

During this year of research I began reading research papers for the first time instead of textbooks. One thing that I've realized about my own way of reading is that the papers that gave me the most clarity was those that felt like a conversation between me and the author(s). They strike this perfect balance between being conversational and formal. In order to acquire insight, we need to be open about our questions, problems, new perspectives, comparisons, failures, and of course, our own insights, and often I feel like the best way to communicate these experiences is to allow yourself to be conversational. State open questions, remark on failures you encountered, and clarify perspectives. This idea of *research as conversation* is one that I both believe in and one that to me feels natural and rewarding when embracing. Allowing myself to be conversational helps not only to clarify my thoughts and ideas to myself, but hopefully to others as well. Most of my ideas and insights has come from other peoples' remarks, tiny comments, footnotes, and informal reflections.

One challenging part about pursuing research is that there's no specific metric for how "good you did" on a given day. From a productivity standpoint, we usually have defined metrics that measure how good we are. In studying, there's always metrics such as how much time you spent studying, how many pages you read, how many chapters you went through, how many exercises you completed, and what grade you got on an exam. In startups, you again have metrics like how many hours you spent working, how many lines of code you wrote, how many features you implemented, how many customers you talked to, how many posts you made, how much engagement you got, and how much money you made. But in research, these conventional metrics break down and cannot really be applied at all. It feels completely irrelevant to measure "productivity" in research by how many papers you've published, how many papers you've read, how many paragraphs you've written, etc. Progress in research is characterized by clarity, and slow, structural shifts in understanding that helps you see things differently. For example, clarifying what types of spaces are relevant to your research questions, and what authors and papers to immerse yourself in can constitute a large amount of progress. When you stop trying to solve the problem through brute force and instead focus on immersing yourself in the right kinds of texts and niches of knowledge, you progressing. This is why, to me, the solution often becomes an obvious consequence of the environment I am immersing myself within. If the environment is misaligned with the type of problems I want to solve, then I need to recognize that and find more relevant places to immerse myself, where the problem becomes clearer. As another example, I found myself reading Niu and Spivak [2024] on polynomial functors and dependent lenses, a construction I believed to be relevant in the formalization of neural networks and gradient descent. However, after about 150 pages and weeks of studying, I realized due to Gavranović [2024, p. 89], that it wasn't the correct abstraction for my specific questions. This was already huge progress, because it clarified many things about the properties we desire and don't desire when it comes to modeling deep learning components categorically, but it felt like I had "wasted" time. Afterwards, I moved on to other theoretical work that generalized polynomial functors, namely optics and additive dependent lenses, which gave me a clearer approach to the mathematical modeling of neural networks and backpropagation. However, the time spent reading about polynomial functors was definitely not wasted as I have recently found myself coming back to them as parameterized polynomial endofunctors in $\mathbf{Para}(\mathbf{Set})$ to describe the syntax of neural network architectures. So while the research progress is certainly not linear nor well defined by conventional, productivity-based metrics, it is characterized by clarity, and many things you learn along the way *is* useful, no matter if it might be directly tied to what you're currently doing. However, it can be hard to distinguish productive wandering from undisciplined wandering while you're still inside it. Sometimes this type of wandering genuinely wastes time, and sometimes "clarity" becomes something we tell ourselves to justify this wasted time. So always adhere adhere to relevancy relative to your own questions.

Along the way of entering research I thought about what a research contribution is and what type of research is worthy of publication? Initially it seemed that it is only about doing something novel, i.e., answering an open question no one has answered before, or producing work that no one has produced before. I had this naive view that all contributions must contain a sort of "breakthrough-like" result that no one had thought of before. But as I now see it, a contribution is more about continuing and moving forward an already existing conversation, where you and the others in the field of research gets to add your unique insights, ideas, perspectives, and results. They may not be completely novel, they may not be the thing that lands you in a future history book, but they *move the conversation forward.* Although I haven't published anything yet, the most significant milestone I've reached is getting to the level where I am actively partaking in an ongoing conversation with other researchers, contributing thoughts that genuinely advance the pursuit of answering open questions. In this sense, a prerequisite to any form of research contribution comes from one's willingness and capability to contribute with personal thoughts, explorations, ideas, struggles, and results, no matter how small or big. As long as others find your perspectives to be useful, and as long as they enable you together to continue talking about your shared questions and problems of interest in a productive manner that brings clarification and better understanding, you are contributing to the research. However, while research is conversational in this sense, conversation without formalization can become meaningless. But then on the other hand, formalization without the openness of conversation can conceal the process that generates further contributions. A publication might just be a formalization of a specific part of a larger research conversation?

I thought research would be about solving hard problems, similarly to what we do in school, but just much harder, and in a way that builds on the knowledge acquired in school; instead, it has been surprisingly more personal (at least for me). First, I didn't expect philosophy to be such a crucial part of the research process. Questions about mathematical theory versus scientific theory and the nature of axioms were (and still are) constantly showing up and those forms of questions constantly requires you to think deeply about your own perspectives. From everything I've reflected on, "research" is about expressing what I am curious about, distilling a broad direction, clarifying a question, and actively contributing to a conversation, in the pursuit of both intellectual curiosity and broadening our collective understanding. To me, it's this level of understanding-seeking, depth, and human interaction, that makes research identity-work. It becomes personal because at the core, it's just you, your intuition, and your questions, and you need to find ways to connect with others about them, asking for help, and offering help. In some cases it's indistinguishable from trying to understand yourself and your own way of thinking. For example, through my first year of research I've found people that truly inspire me, communities whose missions resonates with me, and I've begun developing a taste for what problems I consider important, the fields I thrive in, and the kind of mathematics I want to engage in. Getting into research has personally meant finding a specific direction that simultaneously makes me curious but also a bit uncomfortable because of some open questions I cannot stop thinking about. Finding your own research direction is not trivial. I have written down ideas, archived them, thought I had a paper ready, then realized it contained many fundamental problems, then wrote about 80 pages of pseudo-research notes just to boil everything down into 10 pages. Then, after eventually finding more relevant texts, I archived the 10 pages I had, chose the most concrete insights I had accumulated, and focused purely on clarifying that. Finally, I now have a draft, and I've begun talking with researchers about it, maybe not because there's anything groundbreaking within it, but because it has allowed me to open a genuine conversation about more fundamental and open questions. Research is a long, chaotic process, and I can sometimes feel like I haven't done anything because currently I haven't actually published anything. But I believe that the navigation, the clarity gained, and the identity-work behind the research is necessary, and perhaps understated.

As I'm entering my undergraduate in mathematics, I want to reflect on how I intend to move forward with my research. With this I hope to focus on embracing a top-down approach to learning, one where traditional prerequisites to research questions are studied in reverse, based on what the research suggests. In other words, I intend to jump into research almost prematurely, starting with the modern frontiers, not knowing the specifics but nonetheless getting an understanding of the philosophy, the broad ideas, and the mathematical objects involved. Then, in parallel, I intend to work my way backwards, studying the theory that the current open questions build upon. This results in studying theory with a broad idea of the research questions you're interested in, and I believe it allows you to ensure that whatever you study stays relevant to your research focus. This approach seems for me to be better than forcing myself through multiple textbooks before even allowing myself to dive into research literature.

2. Experiences with category theory.

In mathematics we often study some form of objects, like numbers, sets of things, shapes, functions, probabilities. These objects often come with some *structure*. For example, consider all integers with access to addition. Then your objects are the integers, but the structure is your addition operation. More abstractly, we might think of this *structure* as that of some operation which takes two objects and spits out a new object, and such that certain properties. I like to think of a mathematical structure as something objects can be equipped with, so as to allow for a richer, more interesting and more useful construct of study. Pretty much all mathematical theories start with this: A fundamental construction corresponding to some collection of *structured objects*. Interestingly, there's usually a corresponding notion of transforming one structured object into another, while preserving that structure. So if this is a persistent pattern across mathematics in general, then why not study the notion of "structured objects with structure-preserving transformations" itself? This was the sort of questions that initially led me to category theory, where we uncover something remarkable: By understanding the structure-preserving transformations we often automatically understand the structured objects themselves. The point is that sometimes the relationships between things carry more useful information than the things considered in isolation. Category theory is a way of asking what matters about things when we stop focusing on what it is made of and start focusing on how it relates, transforms, and composes with other things.

I am unsure whether category theory have actually changed the way I think, or if it just gave me a language to a form I already thought in. I say this because I often pursue questions that are more concerned with the foundations of some construct or theory, rather than a well-stated problem. These two worlds of mathematics are to me typically interrelated, as general questions about foundations are answered by solving problems, and since solving problems typically involve building new foundations. I guess I just seem to be starting in the "foundational questions"-end more often. One dilemma I've recognized with this is that these questions can sometimes be *very* broad, almost starting out as purely philosophical. So at the core, I think the nature of the questions I ask led me to category theory, which in many ways provides the ability to explore these broad questions but in a precise, mathematical way. Furthermore, this is probably why I find it helpful to embrace research as distilling a broad direction, starting with an informal question or idea, and breaking it down to a set of concrete problems, and more specific questions. Maybe for someone who instead often starts with a well-defined problem, the dilemma would flip? That is, this person might find themselves asking whether the solution to a specific problem is related to a broader question or a larger pattern?

One lesson I've learned throughout studying category theory is to never mistake generality for meaningfulness. Because the objects of category theory (e.g., categories, functors, natural transformations), are highly general, I think it's easy to fall in the trap of translating mathematics just for the sake of it. Category theory gives you a high-level view of mathematics and of various fields, and when it comes to applications, it can feel like it shows up everywhere you turn, and that everything can be formalized using category theory. However, the value lies not in a high-level view or a translation itself, but in how that translation is built and what mathematical meaning it actually conveys. In the beginning of learning how to apply category theory, I was asking questions like, "Is this a category?" or "Can I make this into a category?" This is rather short-sighted and in some cases, trivial. Although it can be useful to think about how something might form a category, it is only the first step in a *useful translation*. There have been cases where I have prematurely started constructing some category, just to realize it doesn't buy me anything nor clarify something I couldn't see before. The better questions to ask are often questions regarding the transformation between two relevant objects, like, "Given these related objects of interest, can I formalize the structure that is preserved or must be preserved when moving between them?" This type of question is much more fruitful because it asks about the nature of objects based on its relationship to other objects, which is where category theory can become a powerful mathematical language. Another interesting question is, "Given these transformations between objects, can I compose them?" In applications, we want to ask whether composition corresponds to an actual, real process, or even stronger, are important properties of the construction preserved by composition. The strongest cases of applied category theory often revolve around problems where the central difficulty is not about calculating or understanding one particular thing, but rather, understanding relationships. Lastly, these sort of questions also often help in defining the objects themselves!

Looking back, one of the things I associate with becoming more mathematically mature is that theories stopped looking like isolated subjects, and studying category theory has certainty played a core part in this. It taught me that mathematical theories are not so separated as I thought. This is a big jump from classical textbook reading, where you are supposed to deeply study *one* theory in isolation, then do exercises while remaining within that specific theory, pass an exam, and then move onto the next theory. To me, that limits one's view on mathematics. We are often interested in studying how objects relate, how problems can be translated, and even how whole mathematical theories themselves connect, which is a big part of the current frontier of research (e.g., Langlands program). The clearest cognitive shift I've experienced has probably been in the realm of *structural thinking*. To me, this means thinking deeply about the axioms of a theory or construct, constantly adhering to them, using them, but also questioning them. For example, a derivative might primarily look like a limit formula, but it's also interesting to ask what algebraic properties fully characterize differentiation, independent of the axioms of the real line, based purely on some "axioms of differentiation"[1]. A neural network might be a stack of tensors and nonlinearities, but what structures are actually invariant across architectures, training and deployment? Manifolds are spaces locally modeled on Euclidean spaces, but which part of Euclidean structure is actually necessary for calculus, and what happens if we weaken it? Lastly, one illustration of this structural thinking also comes into play in my research on CDL, where I spent a lot of time thinking about the axioms of the base category of deep learning[2] (see Section

[1]This is what synthetic differential geometry is all about, a topic I'm increasingly drawn towards.

[2]I will throughout this reflection refer to the "base category of deep learning" or just "the base category", which I will denote by $\mathcal{C}$. With this I mean a category whose objects are data spaces, like the usual input and output space of a neural network, and whose morphisms correspond to the neural network

3): What categorical structure is needed to model empirical deep learning? What structure is desirable? What structure is empirically forbidden? For instance, while a Cartesian category is valid for classical machine learning, which is fundamentally deterministic, it is not possible to use such Cartesian structure if one were to develop a theory of quantum machine learning, where the no-cloning theorem forbids the notion of a Cartesian structure. In this case we would have to generalize to purely (symmetric) monoidal categories.

There is also a certain epistemic value in learning category theory. During my study of it, I have come back to computer science, quantum mechanics, quantum computing, theoretical physics, and even found new interests like linguistics. It has started to act as a lens through which I learn, and remarkably, it can make completely unfamiliar fields feel less foreign by looking at things *categorically*. However, although category theory seems to show up everywhere, as I've expressed earlier, this generality should never be mistaken for superiority. Abstraction can be genuinely useful for compressing structures. Personally, a lot of the value comes from being able to take a long definition and extract the main takeaway or intuition by viewing the defined construct categorically. It can also allow you to group several objects and results under the same fundamental universal property[3]. However, viewing mathematics categorically can cause one to undervalue quantitative or analytic approaches, which I've definitely been guilty of. So does recognizing a categorical pattern mean you understand the field? Probably not. But it does help tremendously with entering unfamiliar subjects and relating different kinds of knowledge, whether in different kinds of fields, applications, or different theories of mathematics.

---

maps themselves.

[3] I recently started learning about sheaves, and I found the category-theoretic definition using equalizers and contravariant functors easier to parse than the more axiomatic definition using the gluability and identity axioms.

3. Mathematical foundations and theories of AI.

Can we construct a mathematical theory, consisting of objects modeling concepts from deep learning, such that empirical observations provably appear as direct consequences of the mathematics? Achieving such theory, which in turn would be capable of producing novel predictions about deep learning, is my idea of *scientific theory of deep learning*. This particular question has become the guiding light for my recent research in deep learning, specifically in CDL. By CDL I will refer to a theory of deep learning consisting of mathematical constructions built using principles and objects from category theory. So I am not restricting myself to the ideas presented in the seminal paper by Gavranović et al., "Categorical Deep Learning is an Algebraic Theory of All Architectures" [Gavranović et al., 2024]. I want to make this clear because in my eyes, CDL is broader than a theory of neural architectures and their compositions: It is a theory of the components of deep learning and how they relate.

A major part of my interest revolves around formalizing the idea of *priors* on a higher level, that instantiates it as a main object of deep learning theory. The idea of it comes from the observation that many successful models generalize partly by exploiting certain algebraic structure present in the task or data. In my draft I explored this using colored PROP categories, which indeed provides *a* mathematical abstraction for priors. However, I suspect it holds too little structure to become useful for more theoretical work, but that it could potentially become useful in the computational semantics of priors and their implementations. The second part of my draft contains ideas toward generalizing backpropagation on Euclidean spaces via Cartesian reverse derivative categories (CRDCs) [Cockett et al., 2019] to arbitrary manifolds via Cartesian reverse tangent categories (CRTCs) [Cruttwell and Lemay, 2023].

Part of the challenge in developing mathematically rigorous, yet predictive theories of deep learning is that the field contains multiple components and is fragmented across different mathematical descriptions. In deep learning we're working with datasets, tasks, architectures, parameterized models, loss functions, optimizers, backpropagation, training loops, regularization, initialization, hyperparameters, and so on. Not only do we need to develop useful mathematical abstractions for these concepts, but we need to do so in a way that checks out with the empirical reality of them. Furthermore, especially in CDL, it isn't particularly interesting to study these components in isolation. Instead, we ought to study the structural relations between them, form falsifiable predictions, and if they are invalidated by experiments, we would go back and refine the foundational abstractions. It is this iterative theory-loop we want to end up in, but as of now, this hasn't been formed yet. To develop theories of deep learning that connect the various components of deep learning like priors, architectures, and learning dynamics, while forcing those theories into contact with experiments, is the research program I've developed. Category theory seems particularly well-suited to the relational or compositional parts of the problem I care about, not only in unifying various components of deep learning, but understanding how they interplay.

As mentioned, an important observation is that architectures are deeply related to the task. More generally, there seems to be a correspondence between priors and architectures, which often manifests as an algebraic constraint or symmetry in the space of possible neural network models. At this level of abstraction, it isn't clear to me what the relation between these priors and the architectures are, or what a useful mathematical description of priors would look like? Does the prior remove architectures from a search space? Does it generate parameter-sharing constraints? Does it transform an unconstrained architecture into a constrained one? Does it quotient architectures that implement the same prior-respecting computation? In Geometric Deep Learning (GDL), parameter tying is the *expression* of the restriction, but not the restriction itself. More fundamentally, perhaps a prior determines a certain subspace of neural networks, which are just maps in some mathematical space (or morphisms in some base category). If there exists a useful category of prior specifications **Prior**, and a category of parametric architecture syntax like **Para**(**Set**), can one construct an indexed family $P \mapsto \mathrm{Sol}_P(X, Y) \subseteq \mathbf{Para}(\mathbf{Set})(X, Y)$ of architectures satisfying the prior $P \in \mathbf{Prior}$? What characterizes or represents these forms of spaces, where strengthening priors functorially (and contravariantly) restricts architecture?

In understanding the relationship prior and architecture, we first and foremost need to establish what a prior should be at a theoretical level. It is already evident that the *geometric* priors in GDL only form a subset of possible priors. Using category theory, Gavranović et al. [2024, Example 2.6] established how approaches in universal algebra, using monads and monad algebra homomorphisms, can subsume GDL, generalizing priors beyond invertible group symmetries. However, whatever a prior ultimately is, I still believe it should induce a restriction on admissible architectures. So what is **Prior**? Is there a useful mathematical generalization of the various forms priors can take? There's a lot of theoretical work needed to be done here in order for the connection between priors and architecture to be made mathematically precise, and in a way that is also useful on a *scientific* level.

One long-term milestone for my line of work here is developing CDL and other theoretical tools to understand and scientifically study phenomena like generalization and emergence. These phenomena are inherently tied to all of components of deep learning, hence my interest in CDL. Many empirical mysteries like scaling laws, emergence, double descent, and grokking, is concerned with nature of how specific architectures generalize on certain tasks. Central to my research program is what a mathematically universal definition of generalization would look like? Generalization is tied to the task, prior[4], loss metric, parameters, model architecture, and training dynamics. This is why I am so interested in studying the mathematical connections between components of deep learning. One direction I am interested in is studying generalization as a certain extension problem, where we want to extend the behavior defined on observed data to a larger, unseen data domain. Is there a useful representation of some generalization problems as extension problems?

Understanding generalization ultimately requires us to define a notion of a "dataset" at this categorical level of abstraction, as well as the distinction between "training data" and "unseen data". This is not trivial, as there are multiple different levels to what a dataset is, and as it's not clear what representation is desirable or which one is preferable. A dataset is classically a finite indexed family $(x_i, y_i) \in X \times Y$. Categorically, we might model it as a morphisms $d : I \to X \times Y$. Probabilistically, a dataset is given by a finite sample from a data-generating distribution.

Dynamics is something I hope to explore much further. Currently in CDL, optimizers or gradient updates are defined as certain 2-morphism in $\mathbf{Para}(\mathbf{Lens}(\mathcal{C}))$[5] of the form $\binom{P}{P} \to \binom{P}{P'}$. My guiding question here is: How does imposing a prior alter the trajectory of learning, and under what conditions does that alteration improve generalization? For example, if I have an update box in $\mathbf{Para}(\mathbf{Lens}(\mathcal{C}))$, implemented as:

$$\theta_{P'} = \mathrm{update}(X, Y, P, L),$$

together with observables such as $\mathrm{Loss}_{\mathrm{Train}}(P_t)$, $\mathrm{Loss}_{\mathrm{Test}}(P_t)$, and some measure of "prior defect" $\delta_{\mathrm{Prior}}(P_t)$, then does prior defect

[4]Note that respecting a constraint, symmetry, or prior is not the same as generalizing on a task that encodes such prior. For instance, a constant model can be invariant under a group action while being useless for generalizing for the task.

[5]Or in a suitable $\mathbf{Para}(\mathbf{DLens}(\mathcal{C}))$ bicategory.

decrease before, during, or after the transition to generalization?

Loss functions are also deeply involved in determining generalization. In the most general sense, given a neural network $(P, f) \in \mathbf{Para}(\mathcal{C})(X, Y)$, a "loss" is a parametric map $Y \to L$ whose parameter space is Y, that is, a parametric morphism $(Y, \ell) \in \mathbf{Para}(\mathcal{C})(Y, L)$. In Gavranović [2024, p. 150], Bruno asks whether any loss functions have a universal property? This results in many interesting questions. What is the minimal structure needed to express loss functions?

Closely related to my work are questions regarding the relationship between categories of dependent additive lenses[6] and (reverse) tangent categories. In Gavranović [2024], Bruno defines the category of dependent additive lenses, $\mathbf{DLens}_A(\mathcal{C})$, over the base category $\mathcal{C}$, defined as a contravariant Grothendieck construction built from the a specific contravariant slice pseudofunctor $\mathbf{CMon}(\mathcal{C}/-) : \mathcal{C}^{\mathrm{op}} \to \mathbf{Cat}$:

$$\mathbf{DLens}_A(\mathcal{C}) = \int_{X \in \mathcal{C}} \mathbf{CMon}(\mathcal{C}/X)^{\mathrm{op}}.$$

$\mathbf{DLens}_A(\mathcal{C})$ has objects corresponding to maps $p : X' \to X$, where $X \in \mathcal{C}$, and p is a commutative monoid object in the slice category $\mathcal{C}/X$. Morphisms are dependent lenses, i.e., pairs $\binom{f}{f'} : \binom{X}{X'} \to \binom{Y}{Y'}$, where for $f : X \to Y$, and $f' \in \mathcal{C}/X(X \times_Y Y', X')$, that additionally, in the backward map f', are fiberwise additive. The important part is that the additive structure is necessary for the additive reverse-differential frameworks relevant to deep learning, where error contributions compose by addition, or "accumulate". We also often have layers in neural networks that don't contribute to any accumulated error, hence a zero element is needed. So categorically, the fibers used in backpropagation are commutative monoids. Now,

[6] When $\mathcal{C} = \mathbf{Set}$, one recovers the familiar container/polynomial setting for dependent lenses as in Niu and Spivak [2024, Def. 3.1]: objects correspond to containers (equivalently, polynomial functors), while dependent lenses correspond to the appropriate morphisms between them in **Poly**. The dependent lenses referred to here are formulated more generally over a base category $\mathcal{C}$, which need not be **Set** [Gavranović, 2024, Example D.16].

as part of my ideas in my draft was to define a *category of dependent lenses over a cartesian reverse tangent category* $\mathcal{C} \in \mathbf{CRTC}$, denoted $\mathbf{DLens}^*(\mathcal{C})$, consisting of the following data:

- **Objects:** pairs $\binom{A}{p^*_A}$, where $A \in \mathcal{C}$ represents the base space (the data space), and $p^*_A : T^*A \to A$ is the canonical projection of the reverse tangent bundle defined in the system of differential bundles $\mathcal{D}$;
- **Morphisms:** a morphism from $\binom{A}{p^*_A}$ to $\binom{B}{p^*_B}$ is a pair $\binom{f}{T^*f}$, where $f \in \mathcal{C}(A, B)$ is the forward pass, and T^*f is the backward pass, evaluated as the map:
$$T^*f : A \times_B T^*B \to T^*A;$$
- **Identity:** the identity for an object $\binom{A}{p^*_A}$ is the morphism $\binom{\mathrm{id}_A}{T^*\mathrm{id}_A}$, where $T^*\mathrm{id}_A = \pi_2 : A \times_A T^*A \to T^*A$ is the canonical projection.
- **Composition:** let $\binom{f}{T^*f} : \binom{A}{p^*_A} \to \binom{B}{p^*_B}$ and $\binom{g}{T^*g} : \binom{B}{p^*_B} \to \binom{C}{p^*_C}$ be two morphisms in $\mathbf{DLens}^*(\mathcal{C})$. Their composition is a morphism $\binom{h}{T^*h} : \binom{A}{p^*_A} \to \binom{C}{p^*_C}$, where $h : A \to C$ and $T^*h : A \times_C T^*C \to T^*B$ are given by:
$$\begin{aligned} h &= g \circ f, \\ T^*h &= T^*(g \circ f) = T^*f \circ \langle \pi_A, T^*g \circ \langle f \circ \pi_A, \pi_{T^*C} \rangle \rangle. \end{aligned}$$

On the one side, you have $\mathbf{DLens}_A$, which provides the correct typesetting for reverse differentiation and the "ambient setting" where any form of learning by means of backpropagating an error or loss is even possible. Now, I wonder whether a reverse tangent structure is more of a canonical choice of dependent lenses, specifically for when one backpropagate errors using differential geometry. For this we need more structure, especially in the cases where the geometric spaces involved are more complicated, like stratified or curved. Does there exist an embedding $\mathbf{DLens}^* \hookrightarrow \mathbf{DLens}_A$? Does a reverse tangent structure pick out a canonical family of additive dependent lenses? Which additive dependent lenses arise from the reverse tangent functor of some reverse tangent structure? Exactly what additional structure or equations distinguish reverse tangent dependent lenses in $\mathbf{DLens}^*$ from arbitrary dependent additive lenses in $\mathbf{DLens}_A$? Moreover, optimization additionally requires a way to turn cotangent information into an update of parameters. Dependent lenses in $\mathbf{DLens}_A$ may model backpropagation of abstract covectors on non-flat spaces, but they do not by themselves solve parameter updating on non-flat spaces. This could potentially be where the theory of reverse tangent categories meet that of dependent additive lenses. On a deeper level, this particular direction is very interesting to me as it explores a question where we are forced to reconcile the geometry of differentiation and the semantics of bidirectional learning: What is the boundary between an arbitrary backward process and a geometrically valid reverse derivative?

If we arrive at correct abstractions, elevating empirical deep learning in the language of category theory, then perhaps we can connect questions and insights from many different theories of deep learning? Indeed, there exists other theories that aims to explain and understand different aspects of deep learning: Geometric Deep Learning (GDL), Neuroalgebraic Geometry, Topological Deep Learning (TDL), Neural Tangent Kernel (NTK). Maybe CDL could connect insights across other theories and ask new questions that transcend the boundaries of individual theories? What if we can ask questions where we connect mathematical objects in one theory to that of another? Perhaps we can ask and explore meaningful questions about deep learning that aren't fully well-posed in any single field? We've already begun seeing this sort of "higher level work" being done with the generalization of GDL in Gavranović et al. [2024]. So as part of my future work, I hope to explore connections between the various theories of deep learning. It is not yet clear what these "connections" could look like, but the idea is motivated by how we develop theories of physics: We didn't start with some grand unified theory. We started with by developing theories around specific phenomena, at different levels of abstraction. Different theories use different mathematics and exist on different scales. Even though a unified theory might exist, it doesn't invalidate the other theories. It allows you to zoom in and out on different conceptual levels, and *choose* what theory to apply depending on what problem you're faced with. This is how I envision the science of deep learning to be developed, where a big part of it is about understanding the boundaries of the theories we develop and the translations between them.

Finally, I've been thinking about what applied category theory in deep learning even could help us with. What can category theory do for us that other theories cannot do? Conversely, what can category theory *not* do? CDL risks becoming abstraction without prescriptiveness. A lot of CDL papers *describe* deep learning concepts like emergence, architectures, and backpropagation, but none are currently, as far as I'm aware, producing empirical predictions about learning phenomena like emergence. I don't yet know whether the lack of empirical prescriptiveness is merely because CDL is young, because the abstractions are not yet right, or because categorical structure must ultimately be combined with other, more quantitative theories to explain these phenomena. I believe predictions will be made using CDL, but maybe in a different way we think? Maybe CDL won't allow us to predict exact scaling factors and numerical thresholds for specific capabilities to emerge in neural networks. Instead, maybe it could allow us to prove more structural connections between, say the ability for a model to respect a certain prior and its architectural structure? Maybe it can help in explaining how priors and generalization are related? Maybe it can shine light on specific classes of architectures or loss functions by uncovering universal properties? Or, as many hope for, perhaps it could predict completely new classes of architectures? I also hope to explore broader questions about generalization.

4. Venturing into mathematical research.

I'd argue mathematics is one of the most universal, free, and creative forms of expression. When I initially started out researching, my mental map of "doing research" was split into different types of research, like doing research in physics, deep learning, and pure mathematics. But as I've experienced, there are no hard boundaries. In fact, for me, the commonality of all my interests *is*, in many ways, mathematics. Whether you're exploring physics, deep learning, or computer science, mathematics is what we use to formalize what we build and what we experience. Consequentially, I suspect that I will find myself researching mathematics *as* I research other parts of my fields of interest, which is something I've already begun experiencing with my research in deep learning. So naturally, because mathematics is so connected to many things in my life, I am excited to focus on purely mathematical research. Although I haven't started seriously researching any concepts in mathematics, I feel mature enough to begin exploring, and I do have some directions I am curious about. But before I continue reflecting on these directions, I want to have an honest discussion with myself on what mathematics is for me.

Like any other person with an interest in mathematics in 2026, I want to discuss the effect of AI on it. I think Terence Tao put it well when he said that we've never really been through an existential crisis this large before. If an AI system can solve conjectures and open research problems that even the worlds most proficient mathematicians couldn't solve, then what is left for mathematicians to do? What does mathematics look like in such future? What is the purpose of mathematics? There's so many interesting questions to explore, and much philosophical work to be done here, and I hope to genuinely spend time writing about this in the future. For now, I'll only express my initial thoughts briefly. First, I believe this opens up a healthy discussion on what mathematics is to us, why we develop mathematics in the first place, and why we desire solving mathematical problems. I've seen people saying "math will be solved" and that "mathematicians will become replaceable". Disagreements here probably stem from different beliefs about what mathematics fundamentally is about. If you see it as purely solving problems, then I guess mathematics will be solved, and mathematicians will be replaceable. These technological breakthroughs is of course impressive, and I believe they are good. But even if we end up with an AI model that solves the Riemann hypothesis, I don't believe this will "solve math" nor do I think it's even reasonable to say that math "can be solved", because mathematics is also about expression. It is about a way of reasoning, asking questions, understanding problems, finding applications, and a way of constructing useful theory. Can an AI ask questions "we" care about? What if I personally have a question and want to explore it mathematically? Can an AI that have solved the Riemann hypothesis help? Probably! But does that mean that I shouldn't care about my question, just because an AI also can explore it mathematically? Absolutely not! To ask a meaningful, mathematical question, you have to possess an understanding of what matters, and what connects deeply to other parts of human knowledge, as well as your own goals. In fact, if an AI could provide valuable help, then I believe we should learn to explore and collaborate with AI systems, and build infrastructure around better AI-human collaboration! Every time technology have evolved, we naturally have this kind of existential worry. When computers were first invented to perform calculations, the same anxieties probably appeared, especially among those whose careers relied on performing arithmetic calculation by hand. But the horizon just moved. It all keeps moving. If we have machines that can verify any proof or solve any problem we throw at it, then the human role of being a mathematician will shift. No one knows where it will shift, but

I currently believe it will shift toward asking questions that matter, and building mathematics that matter to us. As long as we are encountering new complexities in our world, whether in physics, the nature of intelligence, pure math, technology, or in spaces we haven't even discovered yet, we will need to ask new questions and articulate them. I believe mathematics enables us to do just that, and thus, AI will not "solve mathematics", but it will probably help us in whatever the future of doing mathematics will look like.

In the most general sense, my personal idea of mathematics is centered around it being a structured way to explore thoughts. It's a form of expression because we build theories to explore questions and solve problems, where by "building a theory" I mean constructing abstractions and defining objects that allow for us to reason! Reason about what? Well, about anything we want! I hope to engage in mathematics where solving problems is about building the right theory around it, and where building theory stems from, first and foremost, a personal curiosity, but also a genuine need or an open problem. I can find myself almost prematurely wanting to validate a mathematical curiosity. I try to look at it and immediately answer whether it will be a useful direction and whether it will solve any meaningful problem. But it often paralyzes one into an endless loop of preparation, rather than being productive in any sense. So moving forward, I will not try to look for any form of answers before I explore. Consequentially, I might come off as a bit sporadic in the following paragraphs, where I just reflect on questions, tiny openings for exploration, without having done the exploration nor pretending to have any answers. Finally, I'll end this with a philosophical question I find interesting: Does math even need to be useful? What does this even mean? Math is to me intrinsically personal, and maybe it is not primary that one's mathematical ideas and directions become "useful" or applicable in the real world?

I will now continue by discussing some personal "tiny openings" into mathematical research directions in mathematics.

In my ongoing work on CDL I thought carefully about what properties and structures the base category $\mathcal{C}$ for deep learning should have, or are empirically allowed to have. This would be the category where deployed neural networks live as morphisms $f_p : X \to Y$, where $X, Y \in \mathrm{Ob}(\mathcal{C})$. We often take it to be the category **Smooth** of Euclidean spaces and smooth maps. The problem is that we use non-smooth maps in deep learning, like ReLU, which is piecewise smooth but not differentiable at $x = 0$:

$$\mathrm{ReLU}(x) := \max(x, 0) = \begin{cases} x & \mathrm{if} x > 0, \\ 0 & x \leq 0 \end{cases}$$

This is not a big problem in automatic differentiation (AD) frameworks like PyTorch or TensorFlow, which adopt a convention for the backward derivative at the non-differentiable points; for example, PyTorch uses 0 at $x = 0$. However, on a *mathematical* level, this matters. Even though we often use **Smooth** and ReLU when composing neural networks, we theoretically can't. The classical derivative does not exist at $x = 0$, even though an AD implementation may specify a backward rule. Moreover, if we formalized such a model in Lean with layers required by type to be morphisms of **Smooth**, ReLU could not simply be inserted as a valid layer, because one would have to provide a smoothness proof that does not exist. However, there are also widely used smooth alternatives to ReLU, such as GELU and Swish. Moreover, in neuroalgebraic geometry [Marchetti et al., 2025], one important tractable setting is neural networks with polynomial activations, hence with smooth activations. So while it *might* not be a big deal after all, both computationally and theoretically, it poses a question I find interesting: If **Smooth** does not contain all of the maps we actually use in deep learning, what should the base category be? This led me to generalized geometric spaces like diffeological spaces, o-minimality, and the idea of tame geometry.

A more concrete question I began exploring after this concerns whether a suitable category of diffeological spaces can support a non-trivial reverse tangent structure [Cruttwell and Lemay, 2023, Def. 1, Def. 24]. A natural place to begin is Blohmann's theory of *elastic diffeological spaces* [Blohmann, 2023]. Blohmann identifies a class of diffeological spaces on which the left Kan extension of the ordinary tangent functor on smooth manifolds defines an abstract tangent structure. Thus, rather than beginning with the category **Diff** of *all* diffeological spaces, one can ask whether the tangent category of elastic diffeological spaces—or perhaps a suitable full subcategory of it—admits the additional structure required of a reverse tangent category. In particular, can its differential bundles be equipped with the appropriate duality or involution [Cruttwell and Lemay, 2023, Def. 23] needed to define reverse differentiation? More broadly, I am interested in whether there exists a nontrivial category $\mathcal{D}$,

$$\mathbf{Man} \hookrightarrow \mathcal{D} \hookrightarrow \mathbf{Diff},$$

containing spaces beyond ordinary smooth manifolds, such as suitable singular, stratified, or infinite-dimensional diffeological spaces, for which the Kan-extended tangent structure can be enhanced to a reverse tangent structure. If such a category exists, what additional restrictions on its objects or differential bundles are required? And if it does not, where exactly does the obstruction arise? What could a possible obstruction result say about the nature of reverse-mode AD?

As far as I understand it, diffeology is a theory of generalized smooth structure. It doesn't really "allow" ReLU to be a morphism in a suitable category of generalized smooth spaces per se, as it depends on what diffeology we choose. It only gives us a specific form of flexibility to change what "smooth" even means. We could choose nonstandard diffeologies that make ReLU smooth. However, in making choice we may simultaneously create other problems that destroys any form of expected (reverse) tangent structure. If our only goal were to construct a valid base category to include ReLU networks, then we could just pick suitable diffeologies. But in this case, o-minimal structures or semialgebraic geometry is arguably more natural. ReLU is semialgebraic because its graph can be described by finite polynomial equalities and inequalities:

$$\Gamma_{\text{ReLU}} = \{(x, y) \mid x \leq 0,\ y = 0\} \cup \{(x, y) \mid x \geq 0,\ y = x\}.$$

Therefore, ReLU is a morphism in a category whose objects are semialgebraic sets and whose morphisms are semialgebraic maps. However, we also would like a reverse tangent structure in order to perform backpropagation on these types of generalized geometric spaces. This is potentially where diffeology comes in, and the point is that semialgebraic geometry and diffeology might be doing different jobs. O-minimality imposes finiteness conditions that rule out wild oscillation and controls tameness. Diffeology supplies a generalized notion of smooth structure. Then, the question is whether there exist a useful bridge between the two to properly accommodate morphisms like ReLU while providing the general geometry such that we get a nontrivial reverse tangent structure and a setting in which reverse differentiation is supported. A hypothetical category might therefore consist not merely of diffeological spaces, but of some form of "stratified diffeological spaces", whose strata have compatible tangent and reverse tangent structures.

Diff seems to be a very vast category, and often times it's impossible to prove general statements *for all* diffeological spaces. However, much of the usability of **Diff** seems to come from restricting to specific subcategories, and studying specific spaces and problems therein. Can we define such a subcategory where objects include Euclidean spaces, smooth manifolds, semialgebraic varieties, possibly quotient or moduli-like spaces, and whose morphisms include smooth maps, ReLU, max, min, and piecewise-polynomial maps, *while* still allowing for some notion of a reverse tangent structure? In such a setting, the reverse structure would have to remain compatible with the geometry of the spaces, behave coherently across singular or stratified regions, and remain stable under the categorical operations needed for composition. It is not even clear that such a construction would fit within the existing axioms of reverse tangent categories; perhaps the obstruction itself would point toward a suitable generalization of reverse tangent structure.

This story of going from a seemingly tiny problem (ReLU is not smooth) to mathematical curiosities that have become interesting independently of their deep learning origin has been a wonderful experience. It is also part of why I increasingly view mathematics as a form of expression. If, while using mathematics to explore questions about deep learning or physics, I find that the mathematical language itself is insufficient, then the limitation of that language becomes a new object of exploration. It becomes a research direction in pure mathematics! This synergy—where problems in the world motivate new mathematics, which can then return to help us understand the world better—is the kind of work I want to engage in. Thus, I think the most important thing I discovered this year has not been a solution to an open mathematical problem, but the confidence that I can discover meaningful directions in mathematics itself. If that pattern continues, then my task is not to predict where I will end up, whether an idea will become useful, or whether a question is worth pursuing, but to keep exploring honestly enough that the next question has a chance to appear.

5. Entrepreneurship.

Entrepreneurship is for me about asking whether something I care about can become useful to someone else. That is, can something I care about be used to solve a problem someone else is experiencing, and can I build a business in doing so? This has been challenging for me. Sometimes a problem is not actually real when you go out to talk to people, sometimes you're pitching a real problem to the wrong group of people, and sometimes a problem just isn't worth spending money on in turn for a solution. Throughout the years of my first 3 "business attempts", I've learned lots about myself. More than anything, my trajectory has taught me that there are two independent questions that need to agree if entrepreneurship should work for me: "Does this matter to me?" and "Does this matter to someone else?"

I've gone from the more "hustler"-side of the things, writing cold calling scripts, trying to find the right idea, and just grinding, to deep tech. I started out with social media marketing with the goal of "wanting to become an entrepreneur" and making some money. It was Facebook ads, cold calling, sales pipelines, and online courses. Eventually, I gave up on social media marketing because I simply didn't like it. It wasn't fulfilling, and it felt like I tried to sell someone something just for the sake of it. To sell something I didn't believe in felt horrible. Entrepreneurship still resonated with me, but definitely not the kind where you just "pick a profitable business model" and start grinding. My next attempt followed a similar, but slightly different pattern. I tried selling website development services to local businesses. This time I started with myself, asking what I could do and what I had some skill in. I already had built some websites from doing personal programming projects, and I also had an interest for programming. But again, it didn't took long before I eventually realized I didn't like building websites nor designing them for other companies. It was never what I spent my free time doing. The significant similarity between these first two business attempts is that neither begins with asking about a specific problem that at least I have experienced. This explains why the entrepreneurial process I initially started with now feels completely backwards.

I then went on to build a study app. My idea was that because I have a passions for studying, I could probably create an app that can aid in studying. It didn't really start with a specific problem other's experienced, which definitely was a big mistake. Nonetheless, there existed study apps and tools that people had reported helped them tremendously. Building actual software was a turning point for me. SaaS felt much closer to the kind of work I wanted to do, and I think there's a couple of reasons for this. First, it allowed me to do more visionary work, and allowed me to think about specific features, design principles, aesthetics, as well as a broader picture, a mission statement, and how to reflect a philosophy about studying in an app. Second, it was much closer to what I was doing. Now I had to learn about backend, different tools, databases, scalable ways of building software, and so on. The problem ultimately boiled down to a saturated market and an imagined problem. The problem I was trying to solve was partly real, but a lot of confirmation came from me because I liked the idea. Caring about a problem makes you better able to understand it, but also worse at neutrally assessing whether other people care about it. Moreover, "study apps" are so broad that I feel like there will always be businesses building them. There are probably hundreds of them out there, yet there will always be someone on Reddit asking about "how to become more disciplined" or "how to study better". This establishes that they experience *some problem*, but it does not establish that the problem can be solved by a study app or even by a product, that it occurs frequently enough, that it existing tools don't solve it, or that they would pay for a solution. I still believe in parts of the vision I had, and it became a really fun project, but not profitable as a business. I couldn't see myself continuing long-term and that's where I decided to stop. However, I did get a couple of users that I genuinely had great conversations with, and since I had a stronger belief in what I was building, marketing became more fun than ever. The most valuable part was learning about product iteration cycles. It was rewarding to engage with users, talk to them, trying to figure out what's working and what's not working, and iterate based on feedback. Practicing this way of working was definitely a great experience.

This touches on why I am drawn towards entrepreneurship. It is one mechanism that prevents me from confusing "meaningful to me" and "meaningful to the world". You cannot fully reason from first principles about whether a product matters. Someone outside your mind participates in determining the answer! This is also what makes entrepreneurship important for me; it ties to being useful to others, and grounding your ideas in this purpose. For me, it has become clear that I am looking for business opportunities where both of these worlds overlap. Social media marketing and website development are meaningful to the world, but certainly was not meaningful to me. Likewise, my idea of a study app was meaningful to the me, but the problem was that it was not meaningful enough to the people I pitched it to. I can fail by choosing businesses I do not fundamentally believe in or enjoy, but I can also fail by genuinely caring about a problem and still hallucinate its importance for others. So my question moving forward will first and foremost be: How can I make things that are simultaneously meaningful to me and meaningful to the reality outside me?

This is why I gravitate towards deep tech, where research, science, and mathematics fundamentally powers solutions to real-world problems. But this only solves one side of problem, namely the "meaningful-to-me" part. In this sense I guess the core of entrepreneurship is more the other part of finding profitable ways to solve problems for a group of people. It is just that I would never be able to build something sustainable if it wasn't meaningful to me. But I could still hallucinate the importance of an AI tool based on category theory just as easily as I hallucinated the importance of my study app.

One challenging part about deep tech is that mathematically meaningful properties and economically meaningful properties are not the same thing. GDL is perhaps one of the more mature theories when it comes to industry-level adaption. This is mostly because it offers *measurable advantages* like greater sample efficiency, fewer redundant parameters, and stronger generalization, properties which are economically meaningful. But the mathematical insights behind GDL are mathematically meaningful independently of this. In contrast, and as I've talked about, CDL's strongest benefits and potential advantages are harder to convert into a metric that engineers or customers care about—unification, compositionality, architecture specification, etc. Thus, for me it becomes not only important to build theory to understand AI systems and learning, but also to understand what problems could benefit economically from the theory, and how such a bridge between research insight and profitability could be built.

When it comes to early product experimentation, I am excited about exploring the way priors restrict architectures. I think CDL could help a lot in building better AI assurance infrastructure for systems that must obey explicit structural constraints. This includes systems where failure has a contractually, legally or physically meaningful definition, such as generated code, autonomous control, medical or clinical systems. As part of more sophisticated research in theoretical deep learning, constantly moving the field closer towards a branch of physics, I imagine more sophisticated systems where you genuinely could input architecture diagrams, tensor types, prior specifications, allowed transformations, perhaps even other forms of constraints tied to for example deployment, agent environments, and differentiation, and where you then can validate whether the architecture respects the prior, and understand which parts of your system breaks certain constraints—a "neural system linter". To distill this vision more, one particular aspect of this is building software around the prior-to-architecture relationship. Maybe a user could specify certain equational laws, information-flow constraints, and generating operations that constrains the possible universe of architectures capable of respecting the constraints. This is tied directly to my future research-level work (see paragraph 3 of Section 3), the mathematics of priors, and the notion of a model "respecting the prior". To again be critical, I can already feel the same danger here. This is extremely meaningful to me, and because the idea connects so naturally to my research, it is easy for me to become convinced that it should exist before clarifying the idea and establishing who actually needs it and why. Research can tell me what *could* be built in a way that is meaningful to me. Entrepreneurship requires me to discover whether it *should* be built for someone.

6. Art and expression.

For too long I've seen my change from music to science and mathematics as exactly that: a change *from* something *to* something else. However, as I see it now, it has been much more of a "broadening" of my thoughts on purpose, which already before 2022 was a part of the music I played and produced.

Art for me is particularly wonderful since it allows for expression without clear restrictions, in the sense of the amount of "stuff you can express". The only real restrictions would come from you. As I have so far expressed with mathematics, it is personal because it is a tool to express your own mathematical ideas, whether it's about formulating a question about something based in the world, or whether it's about understanding something purely mathematical, without any regard to if it is applicable in the real world or if it is meaningful to others. In this sense, science, mathematics, entrepreneurship, and art, are different forms of expression, but with different types of restrictions. Scientific fields are restricted by what we actually experience in nature. Entrepreneurship is restricted by other people's judgment and needs. Mathematics is constrained by logical truth, and to me it has fewer restrictions than for example physics; in mathematics we are not constrained by empirical reality. Art, however, is arguably the weakest forms of expression, and so it becomes very complementary to more "technical interests", and not so distant to them as it might seem. These activities aren't separated by whether they are creative. Instead, they differ partly by what is allowed to tell you that you are wrong.

Coming back to art in the recent year has in turn broadened my view on art, from consisting solely of music, to now being more about this weaker form of expression. I would encourage others to care less about a particular art form. Coming from years of music production, mixing, mastering, arranging, writing, a lot of what art was to me was defined by music, where "doing art" became "producing music". Exploring completely new forms—texts, visual notebooks, typography, websites, personas, photography—has been both amazing and incredibly fun! The point is not that you need to have background in all forms, or even in *any* form. The point is to allow yourself to just do stuff anyways, based solely on your feelings and ideas. It's more about allowing yourself to commit to an idea, rather than practicing a particular form. So take photos even if you've never have taken a photography course, draw even if you haven't drawn a picture since you were a child, and allow yourself to explore what feels natural in the moment.

Earlier creation was for me partly oriented around becoming good, impressive, succeeding, or producing something that met an imagined standard. I am happy I am at a place where I'm able to make something because the act itself produces satisfaction and where I am genuinely enjoying my own works as. Listening to music I played, reading texts I wrote, seeing visual layouts I created, and looking at images I composed. I love it! This perspective might seem egocentric, but because you are not disregarding others' perspectives, I don't believe it is. Instead, I strongly believe one of the primary goals of doing art is understanding yourself. It's interesting to keep coming back to your own works as it forces you to retrospectively understand yourself, who you were, how you changed, and who you now are becoming. It puts *you* first, rather than any form of restriction like usefulness, axiomatic structure, money, or recognition. Embracing this, art has for me become more expressive than I ever have felt before, feeling both free and natural. But do not misinterpret me as saying art shouldn't be shared, or that artworks shouldn't be able to become profitable. I am simply saying that I believe you should put yourself first and create whatever comes to mind, in a way that feels natural to you.

If there is a pattern in the art I am currently engaged in, it is a willingness to commit to seemingly "stupid ideas", ideas that you used to simply keep or let vanish. Personally, what feels natural is to just allow yourself to commit to these types of ideas, spend hours thinking about and building upon them. No matter what they are, they are yours, and giving yourself time to express them can be incredibly rewarding. Creation is sometimes an experiment that later allows you to evaluate the idea. A question can sound silly until you formalize it. A startup idea can sound brilliant until you talk to customers. An artistic idea can sound stupid until you make it. In fact, this reflection itself has been one of those experiments. I began writing it without knowing exactly what it should become, and only through writing it did I begin to see connections between parts of my life that I had previously treated separately.

References

Christian Blohmann. Elastic diffeological spaces, 2023. URL https://arxiv.org/abs/2301.02583.

Robin Cockett, Geoffrey Cruttwell, Jonathan Gallagher, Jean-Simon Pacaud Lemay, Benjamin MacAdam, Gordon Plotkin, and Dorette Pronk. Reverse derivative categories, 2019. URL https://arxiv.org/abs/1910.07065.

Geoffrey Cruttwell and Jean-Simon Pacaud Lemay. Reverse tangent categories, 2023. URL https://arxiv.org/abs/2308.01131.

Bruno Gavranović. Fundamental components of deep learning: A category-theoretic approach, 2024. URL https://arxiv.org/abs/2403.13001.

Bruno Gavranović, Paul Lessard, Andrew Dudzik, Tamara von Glehn, João G. M. Araújo, and Petar Veličković. Position: Categorical deep learning is an algebraic theory of all architectures, 2024. URL https://arxiv.org/abs/2402.15332.

Giovanni Luca Marchetti, Vahid Shahverdi, Stefano Mereta, Matthew Trager, and Kathlén Kohn. Algebra unveils deep learning – an invitation to neuroalgebraic geometry, 2025. URL https://arxiv.org/abs/2501.18915.

Nelson Niu and David I. Spivak. Polynomial functors: A mathematical theory of interaction, 2024. URL https://arxiv.org/abs/2312.00990.