# Research Plan

Joseph Tooby-Smith

My research sits at the intersection of computer science, mathematics and physics. I am intrested in the building of a bridge between these areas using interactive theorem proving. I have PhD in theoretical physics from the University of Cambridge, have completed a postdoc at Cornell University in which I focused on the application of theorem proving software in Physics. For the academic year 2024-2025 I am undertaking a postdoc in computer science at the University of Reykjavik.

## Past research

The underlying theme of my research has been the application of techniques in pure mathematics and computer science to problems in the physicical sciences. This has lead to an expertise in two areas related to computer science: interactive theorem proving and category theory. Let me dicuss these in turn.

**Interactive theorem proving:** The main paper which demonstrates my skills in this area is .... This presents a program to digitalise results (meaning definitons, theorems and calculations) from high energy physics into the interactive theorem prover Lean 4. This is the first anything like this has being attempted in high energy physics. There a four important motivations to of project:

1.
2.
3.
4.

Despite the application of this work been physics, the main challange of this project is use the correct tools in from computer science, and in particular functional programming. To make the digitilisation as easy as possible. One such tool is the use of monads and operads from category theory. This brings me onto my next area of expertise.

**Category theory:** I have a strong background in the application of category theory outside the ivory towers of the pure mathematicians. Historically, my main use of category theory is as a language to recast problems from the physical sciences and to use this language to derive new previously unkown results. As a specific example, in high energy physics there is a relatively new notion called a "generalised symmetry", in ..., we used special types of categories called higher topoi to define and derive new results about these symmetries.

Higher topos theory itself is related to homotopy type theory, which is actually the path that lead me to interactive theorem provers.

Outside of interactive theorem provers and category theory, I also expertise in the theory of Lie groups and their algebras. This is demonstrated by a number of papers e.g. .... Where the this theory was used to computationally search, with the help of graph theory, a discrete space of physics theories for those satisfying certaint conditions.

## Main future project: Theorem proving and AI in the physical sciences

Going forward my main research goal is help progress interactive theorem provers, specfically Lean, so that they can be used more easily in the physical sciences. In addition, I wish to work to further convince academics in the physical sciences that interactive theorem provers are a way forward in academic reasearch, and help build the bridge between the physical science, computer scientists working on interactive theorem provers, and those working on the use of AI in mathematics.

To achieve this goal I plan to undertake the following steps:

1. In Lean 4 there is notion of blueprint for a theory. This is a English-written document containg all of the steps that must be taken to turn the prove of an English-written prove into a Lean written prove. This can be thought of as pseudo-code for Lean. To help build the above bridge I would produce such a pseudo-code for an theory in physics.
2. Most work on AI in mathematics has looked at e.g. math Olympiad problems in Lean (e.g. Google Deepminds work). I would like to see the use of AI to solve problems from the physical science in Lean. To do this I plan to create a data set of Lean 4 written theorems from physics that can be used for AI testing and training.
3. Overlapping a bit with AI, high energy physics use heavily tensors. As part of Lean 4 I would like to develop tactics that help formally verify results related to tensors.

## Plan for undergraduate student involvement

Part of the paper ..., discusses how HepLean can be used as a pedagogical tool, and give students the ability to get involved in research. My plan is to develop three list of undergraduate-level projects around HepLean.

The first list will be concerned with functional programming type projects. As an example: the handling of lists in Lean to efficiently undertake computations needed for index notation (a notation used by physicists to deal with tensors). Other examples will involve meta-programming in Lean to make the user-experience easier.

The second list of projects will be concerned with the use of AI for physics and mathematics. Simple example involve auto-formalisation of theorems in physics (turning a human written result into a result written in Lean), as will as the inverse problem, 'auto-informalisation'. These have being heavily explored in mathematics, but not in the physical sciences.

The third list will be at the boundary of physics, computer-science and mathematics. These projects will involve a proving theorems from physics using Lean. There are many such problems that involve very little prerequiests in physics, once the theorem has being written down. Part of my plan above, with the blueprint, will be a first step in this direction. An example of such a project will be to formalisation of properties of the two-Higgs doublet model potential. This is a potential, and physicists are interested in its properties, such as its minima, whether it is bounded or not etc.

Each of these lists of projects will involve 'home-work style projects' which will take no more than a couple of hours to complete, and more detailed thesis level projects. The benefit of having a large project like HepLean is that coming up with such a diversity of projects is relatively easy.

## Other future project: Higher category theory in computer science

Category theory plays an important role in functional programming. A key example of this is the notation of a 'monad'. A 'monad' is a special case of a more general object in the theory of Higher algebra. This is an area I have expertise, since it overlaps with my use of category theory in physics, as demonstrated in *e.g.* .

The aim of this project would be to explore the possible application of these more general versions of monad in computer science. Part of this process may involve developing or using a langauge which can interactive with the structures of higher categories in some useful way.