# PhD Research Avenues

Meeting of October 6, 2022

## 1 Big Ideas

Long-term, ambitious ideas that can be the aim of the PhD as a whole.

1. *New properties*: so far, one of our contributions was to formulate new structural properties on Neural networks. An interesting result would be to verify a larger system whose correctness depends on a new NN property (that is not reachability). The challenges would be first to identify such a system and the desirable property, then to model and verify it.

2. *Structural properties on a network's output*: instead at looking at the network's own structure, we might want to enforce some structural property on the network's output. For instance, when using NNs to generate trees, the problem of generating outputs with a valid structure has occured.

3. *Properties formulated by ML practitioners*: ML practitioners (e.g. Johnaton Hare) have identified desirable properties for ML algorithms that are not currently being investigated by the NN verification community. We could work towards verifying those; this would probably involve using mathematical proof libraries (e.g. Coq's mathcomp).

4. *Contructive proofs*: provide constructive proofs/certificates for NN verification + a formally verified checker to check these proofs. As of now, SMT solvers may give wrong results due to implementation errors and in a competition, if there is no consensus on a verification task among the different contestants, it is hard to tell who is right. Compared with "usual" problems submitted to SMT solver, where the user has an idea of what the result should be, it is hard to know if the result for a NNV task is correct.

   Providing constructive proofs along with a checker would allow us to verify their correctness. This is the aim we chose to follow for now for several reasons: it arose from a "small idea" (interfacing Vehicle and Imandra) so we have a clear idea of where to start; and it would bring a significant contribution to the field (make sure that the verification proofs are indeed correct).

## 2 Small Ideas

More concrete ideas that could be implemented in the short/medium term.

1. *Imandra-vehicle connection*: as discussed in big idea 4. above. Two possible interfacing: call Imandra from Vehicle or call Vehicle from Imandra; in the first case, it would allow us to use Imandra's constructive proof in Vehicle; in the second case, it would allow to use specialised NN verification tools like Marabou from Imandra.

2. *Unified implementation of CheckINN*: at the moment, we have multiple parts of the library that each respond to one verification goal. We want to have a nice unified library, where one implementation would work for all (or most of) the verification techniques and properties (e.g. CNNs can be implemented with Matrix as lists)

3. *Equivalence checking*: as discussed with Grant, using some sort of equivalence checking between the original model and a compressed (i.e. quantised or binarised) one, then we can transfer the more easily verifiable properties on the compressed version to the original version. Work on approximte equivalence in the field of differential privacy could be of interest.

4. *Handle non-linear approximation functions*: as mentioned in the reviews, we could use recursive Cauchy sequences/recursive functions to approximate non-linear activation functions (e.g. final softmax layer in a classifier).