# 1. What are the core defining features of the delusion?

## 1.1 What makes the Delusion interesting?

### 1.1.1 What aspects about the delusion make me exited?

Working with machine learning and python

Finally finishing my batchelors

It’s a real challenge

### 1.1.2 What is the most interesting way that I could realize this delusion?

A novel state-of-the-art glycan prediction model applied to an interesting problem as a test

### 1.1.3 What of my <core interests> do the delusion tie into?

* Complex systems
* Programming
* AI

#### 1.1.3.1 What aspects of the <thematic core>?

* Runaway systems
* Oracles

### 1.1.4 What parts of the delusion are not interesting?

Applying an already known method mechanistically to a known system

Writing the paper

### 1.1.5 Why do you want to do this delusion? What is its purpose?

To learn and to get my batchelors done and perhaps even to open up new possibilities

## 1.2 What is the gist of the delusion?

Taking the embeddings from a Glycan language model trained on SMILES data and piping it in as features in the SweetNet Graph neural network to improve its glycan property prediction

### 1.2.1 What is the central concept of the delusion?

Using the embeddings from a glycan language model as starting featurization in a glycan GNN to make it perform better than when using randomized starting values

### 1.2.2 What is the central idea of the delusion?

The glycan language model has learnt something about the different sugars and bond types that will give the gnn a heads-start when the training starts, leading to better fitting in the end and higher performance

### 1.2.3 What is the central theory of the delusion?

GNNs benefit from non-random featurization

### 1.2.4 What is the central vision of the Delusion?

A emergent result of a new gnn that performs better than the constituent parts

### 1.2.5 What are the top-three objectives of the delusion?

1. I do Something that I can write a thesis about
2. A working Sweetnet with the glm embeddings, doesn’t matter how it actually performs
3. An improved version of sweetnet

#### 1.2.5.1 Specific, measurable, ambitious, relevant, and testable (S.M.A.R.T)?

1. Using Sweetnet to predict a specific property (to test the base system, and all iterations)
2. A Sweetnet with glm embeddings that I can benchmark and compare to the base Sweetnet
3. A Hyperparameter optimized Sweetnet that outperforms the original (with or without embeddings)

### 1.2.6 Can the delusion be restated as a Hypothesis?

#### 1.2.6.1 As a verifiable hypothesis?

Does a GNN featurized with embeddings from a GLM perform better than the base system?

##### 1.2.6.1.1 What would be an experiment to test this hypothesis?

Run a evaluation test using a specific task (or even better several different tasks https://glycanml.github.io/project/ ) On both models

#### 1.2.6.2 As a Popperian (bold & falsifiable) hypothesis?

Will my system perform better at [specific task] than the base system?

##### 1.2.6.2.1 What would be an experiment to test this hypothesis?

Use both models to do the task and compare the results to reality

### 1.2.7 Can the delusion be rewritten as a problem?

How could (I use embeddings from a GLM to) improve a Glycan GNN?

#### 1.2.7.1 What are some possible solutions?

* + Use embeddings as featurizations
  + Improve hyperparameter optimization
  + Rebuild architecture

### 1.2.8 Can the delusion be rewritten as a thought experiment?

If… No I’m just getting stuck here. NaN

### 1.2.9 Can you think of any metaphors that mirror aspects of the delusion?

* To learn one programming language and then use that base knowledge to learn another
* To have a runner run-in a new pair of shoes before handing them over to a hiker

### 1.2.10 What do I want to prove?

That the Glycan language model learns something useful about the glycans, and that this is transferable

### 1.2.11 What is the Goal of the delusion?

To Improve SweetNet using GLM embeddings, or prove that this technique isn’t useful

## 1.3 Where are the boundaries & constrains of the delusion?

### 1.3.1 What assumptions does the delusion rest on?

* That the embeddings taking from the pre-trained GLM has captured something about the structure and chemistry of the glycan
* That starting with non-random features improves the model after training
* That this method doesn’t lead to over-fitting

### 1.3.2 What are the known unknowns of the delusion?

Will the embeddings improve the model?

How will I pipe in the embeddings?

### 1.3.3 What is the set of qualities that makes this delusion what it is and none other?

* No one has tried using GLM embeddings in a GNN before
* Super-focused

### 1.3.4 What are the riskiest aspects of the delusion?

I lack the expertise to actually do what I want to do in python, making a fool of myself

The embeddings don’t improve the model

The embeddings don’t map well to Sweetnet

### 1.3.5 Are there any contradictions?

The GLMs don’t perform that well in their base configuration, how do I know that they actually learn?

#### 1.3.5.1 Can they be resolved?

Trying other GLMs

### 1.3.6 What might be impossible?

Making a new state-of-the-art

### 1.3.7 What is possible?

Finding out if the method works

Slight improvement through hyperparameter tuning

### 

## 1.4 What is the theoretical basis of the delusion?

### 1.4.1 What field(s) does this delusion belong to?

* Machine learning
* Glycobiology
* Programming
* Graph neural networks
* Hyperparameter optimization
* Language models
* Hybrid approaches to machine learning
* Feature engineering

#### 1.4.1.1 Select three that capture the delusion

1. Graph neural networks
2. Hybrid approaches to machine learning
3. Feature engineering

### 1.4.2 What knowledge could I seek in inchoate delving?

* I’ve done quite a bit of pre-delving
* Perhaps enough

### 1.4.3 What do I need to find the state-of-the-art of?

* Glycan language models
* Graph neural models
* Feature engineering

### 1.4.4 What theories might be useful?

* Linear algebra
* Message passing

### 1.4.5 What would be good to know?

* Object oriented programming

### 1.4.6 What have you done that could be relevant to this delusion?

The explore part of the EDGE of chaos

VEX programming

#### 1.4.6.1 Do you have good things for the research pile hidden somewhere?

### 1.4.7 What should I find a wide review article of to get good starting points?

Feature engineering

Embeddings in Neural models

## 1.5 What is the delusion?

I am going to try to see if I can improve the performance of the SweetNet Graph Neural network by using the pre-trained values of embeddings of a Glycan Language Model as features at the start of training to see if the relationships they have captured through training improves the learning of the new hybrid model.

#### +be as specific as possible, keep it to one sentence. Take your time

## X.X develop the catech template, save with new version

### X.X.1 Remember to open up the original template

### X.X.2 add ideas and experiments to the bottom of the original.

# Iterative development.

## Things to add in future iterations (ideas):

* More storytelling tie-in.
* A more scientific waypoint.
* In some delusions certain aspects can be captured in their own stand-alone prototypes, prototyping one part at a time and putting them together in the final deutrotype

## Things to vary/experiment with:

* I should try to find a nice middle state between deterministic and non-deterministic chaoskampf
* I wonder if I should remove the one-sentence limitation of the prototype description?
* Perhaps Each iteration could be a feature?
* After the first round of inchoate catech I should make a through definition of the delusion itself

## Hypotheses:

## Points of friction (not necessarily bad):

* 1.2.5.1 X
* 1.2.6.2 XX
* 1.2.8 X