**Leash Bio Competition**

**Link to Kaggle:**

[**https://www.kaggle.com/competitions/leash-BELKA**](https://www.kaggle.com/competitions/leash-BELKA)

**Notes:**

* Evaluation: Mean Average Precision (micro) between predicted probability and observed target
* **Chemical Representations**:
  + **Strings**: SMILES/SMARTS, International Chemical Identifier, FASTA
  + **Images**: Molecular Graphs(2D, 3D), 3D Coordinates,
  + **Numeric**: Molecular Finger Prints, Molecular Descriptors, Pysiochemical descriptors, Atomic Environment Matrices, Voxel Graphs, Density Maps, SASA, Molecular Electromagnetic Potential

**String:**

* Look into different variants of SMILES
* NLP Models
  + Finetune
  + Prompting?
  + Feature extraction?
  + Multitask?
  + MoE?
  + Masking?

**Graphs/Images:**

* CNNs/ViT/other machine vision models
* Graph Neural Networks

**Numeric:**

* ANN
* Bayesian Neural Network
* VAE
* Gaussian Process Regression/Deep GPR

**Ensemble Models:**

* Numeric Model -> String Model (use numeric outputs as input to the Gate in MoE)
* Numeric Model -> Graph Model (MoE)
* Multiple Graph Models -> Numeric
* Multiple NLP Models -> Numeric
* Numeric -> Graph + NLP -> Numeric

**Current Problems:**

* Dataset is too damn big