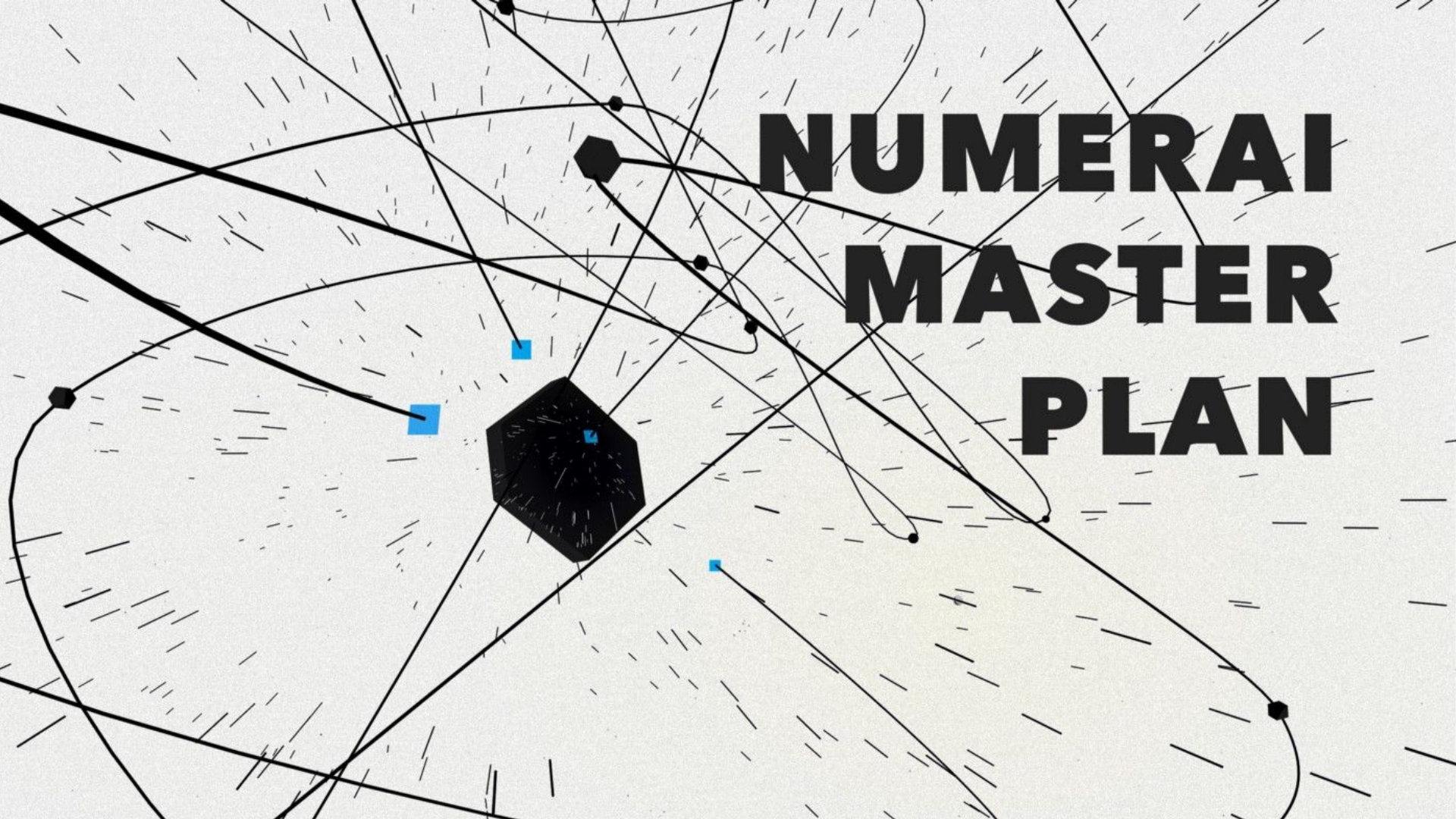
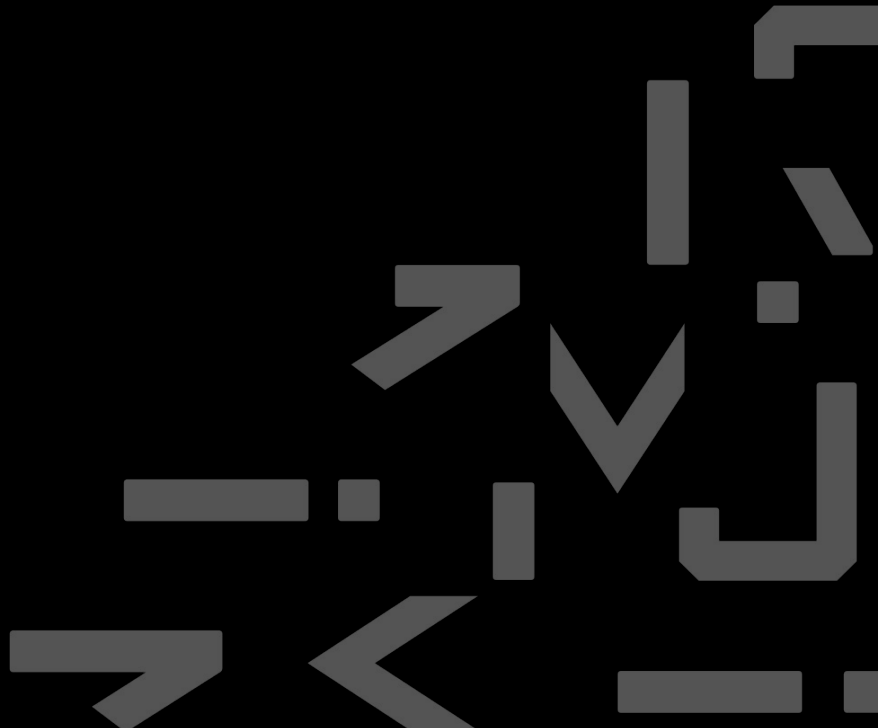
NUMERCON 2022


NUMERAI
MASTER
PLAN

THE NUMERAI MASTER PLAN

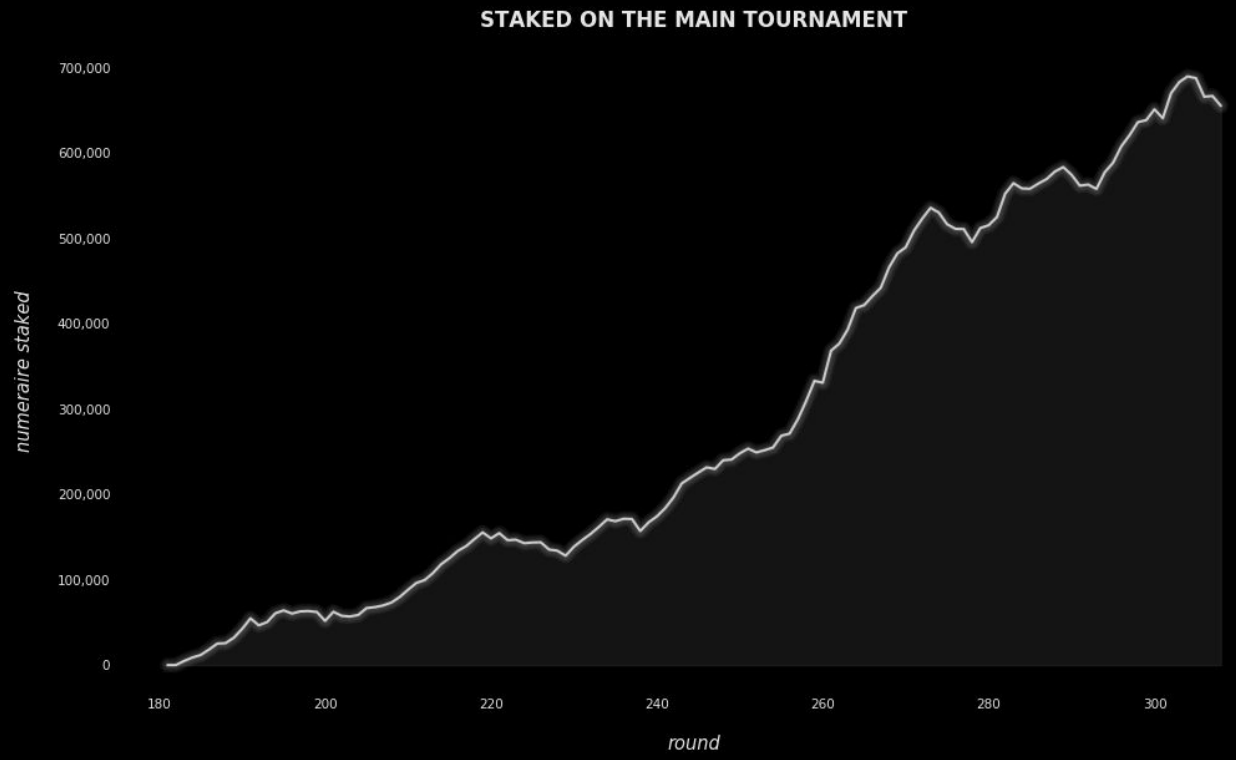
1. Monopolize intelligence
2. Monopolize data
3. Monopolize money
4. Decentralize the monopoly

1+2=3 and 4 would be awesome



i. INTELLIGENCE

Progress in Intelligence: 15x models, 100x USD staked



growth in staked models
=
growth in intelligence?

growth in models contributing intelligence

=

growth in intelligence

INTELLIGENCE CONTRIBUTING MODELS?

AUC against target?

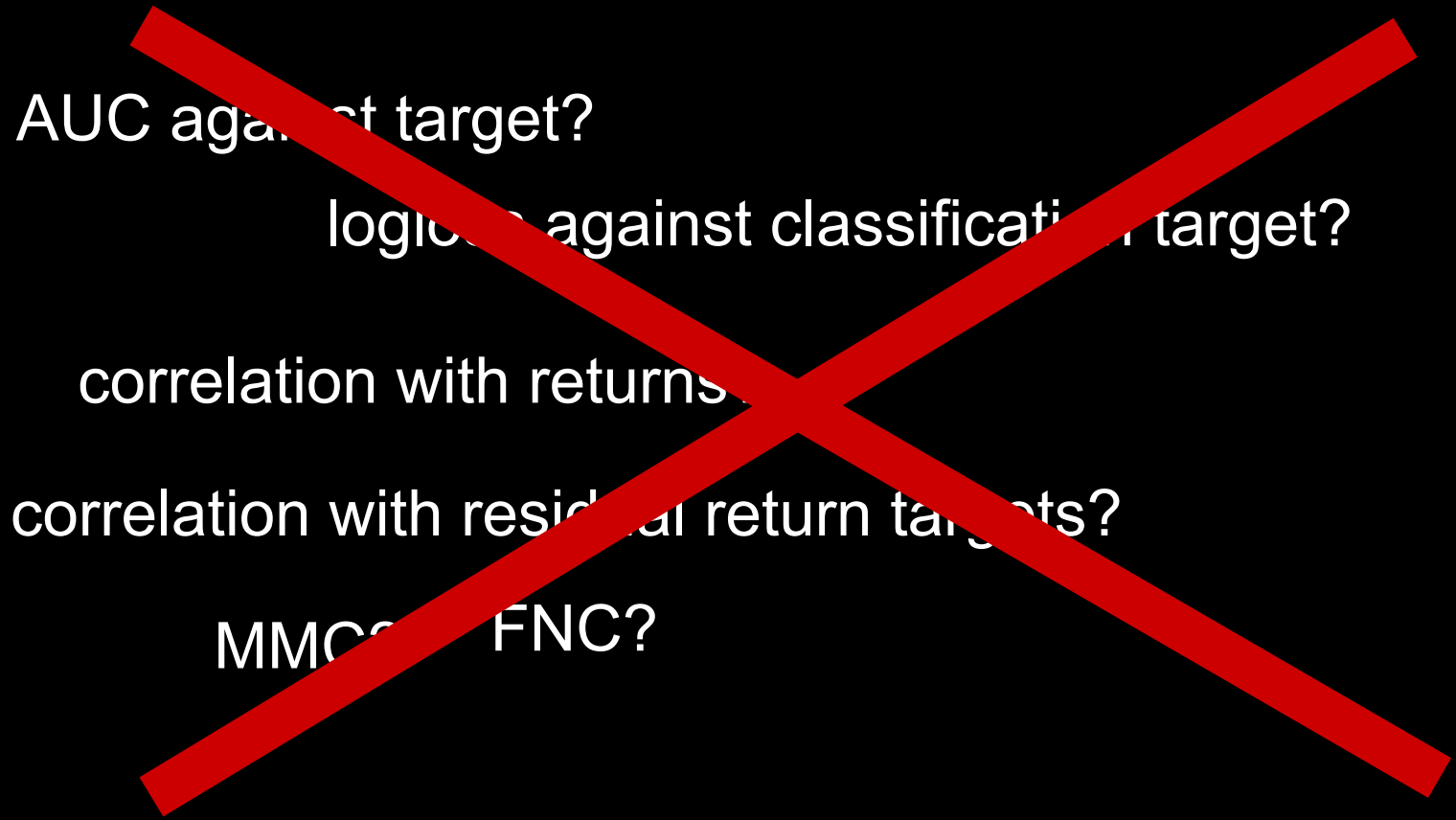
logloss against classification target?

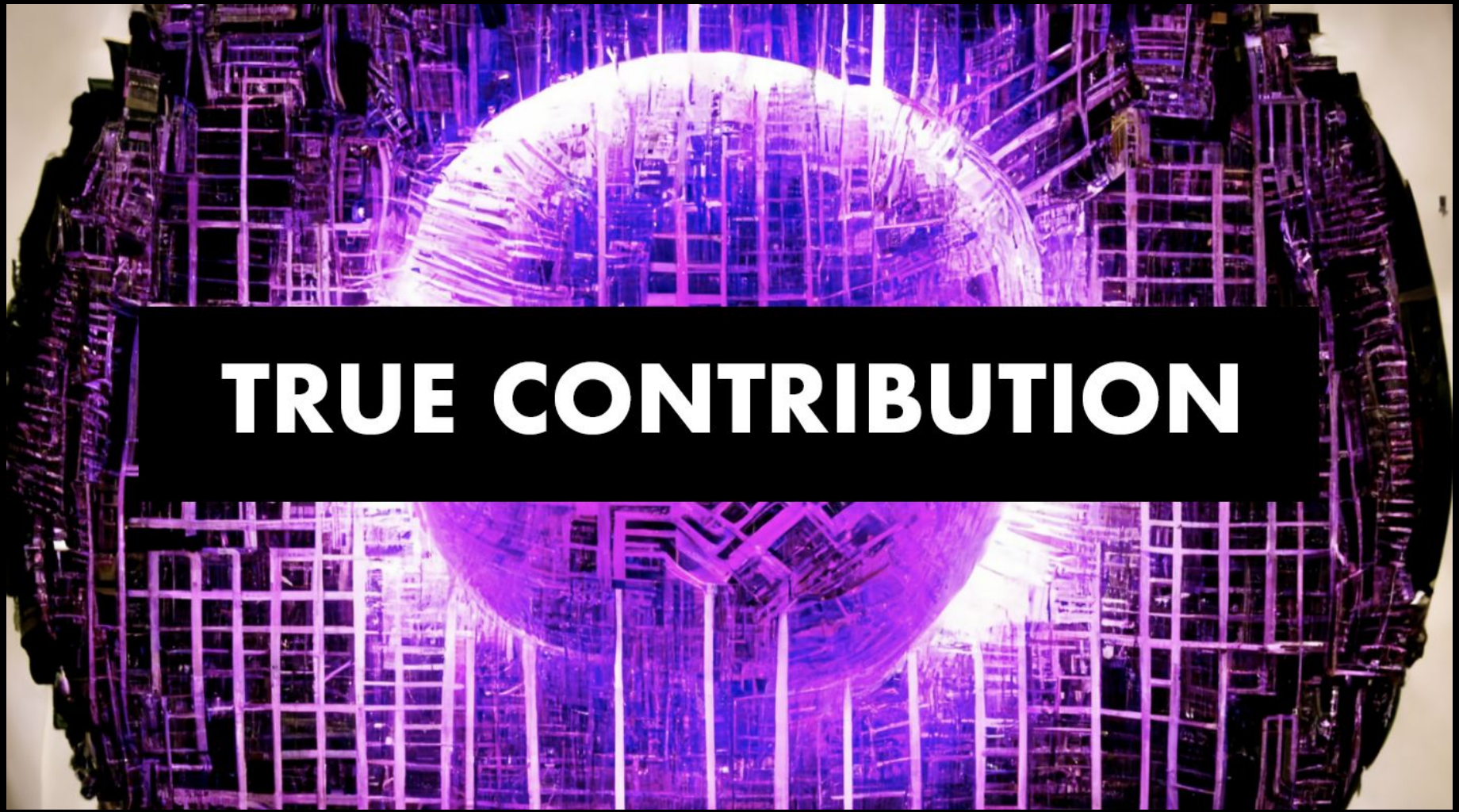
correlation with returns?

correlation with residual return targets?

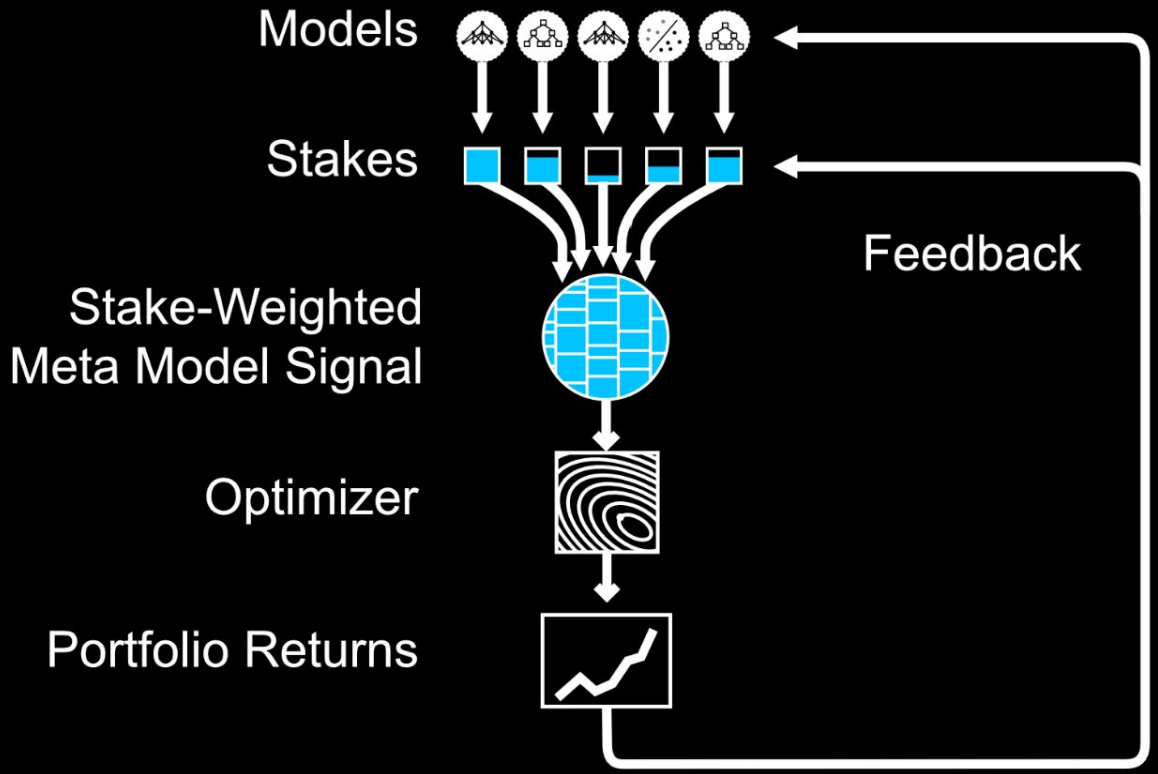
MMC? FNC?

INTELLIGENCE CONTRIBUTING MODELS?




TRUE CONTRIBUTION

TRUE CONTRIBUTION



Leaderboard HB_		▼ CORR	MMC	FNC	TC	STAKE
2	HB_APOLLO 2XMMC	0.0502	0.0208	0.0296	0.0284	8.53 NMR
15	HB_SCOUT 2XMMC	0.0475	0.0143	0.0196	0.0124	765.19 NMR
87	HB_EUROPA	0.0414	0.0168	0.0283	0.0325	—

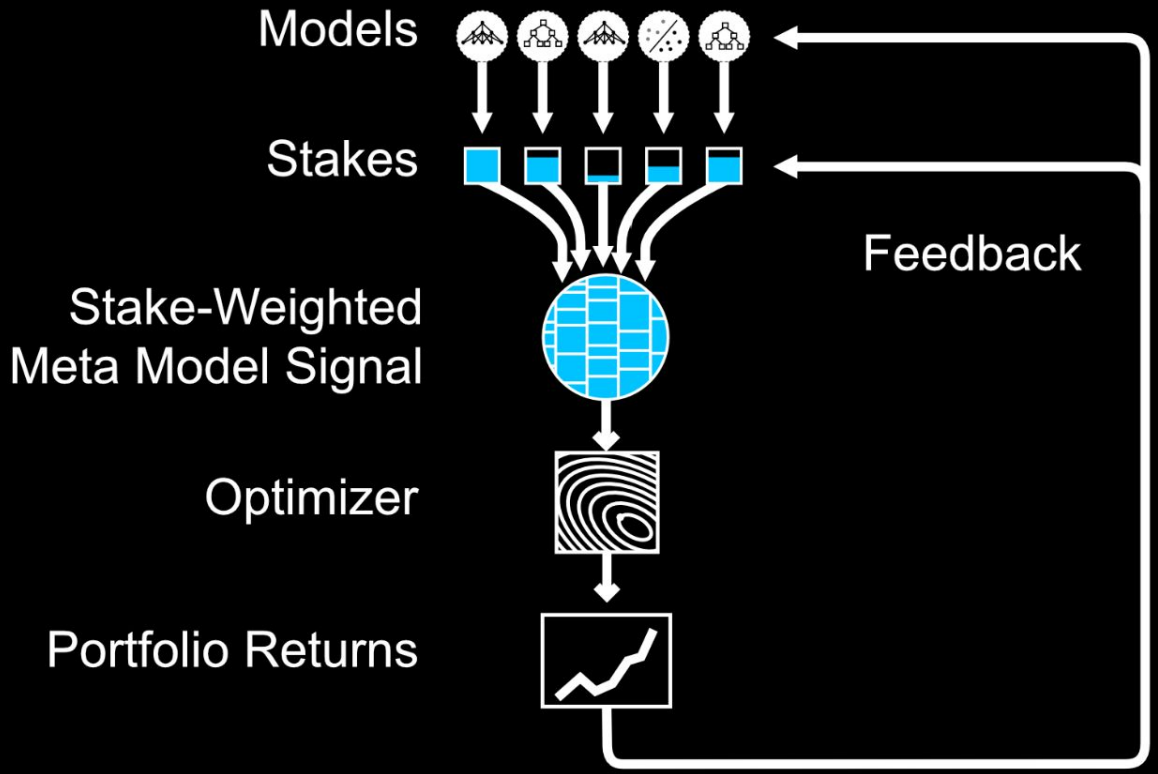
HB's highest staked model has low True Contribution and his model with high True Contribution

STAKING ON TC BEGINS APRIL 9TH

What is True Contribution and how do I get it?

- TC rewards improvements to the portfolio
- Depends on signal, optimizer, and everyone else
- Aligns incentives

TRUE CONTRIBUTION



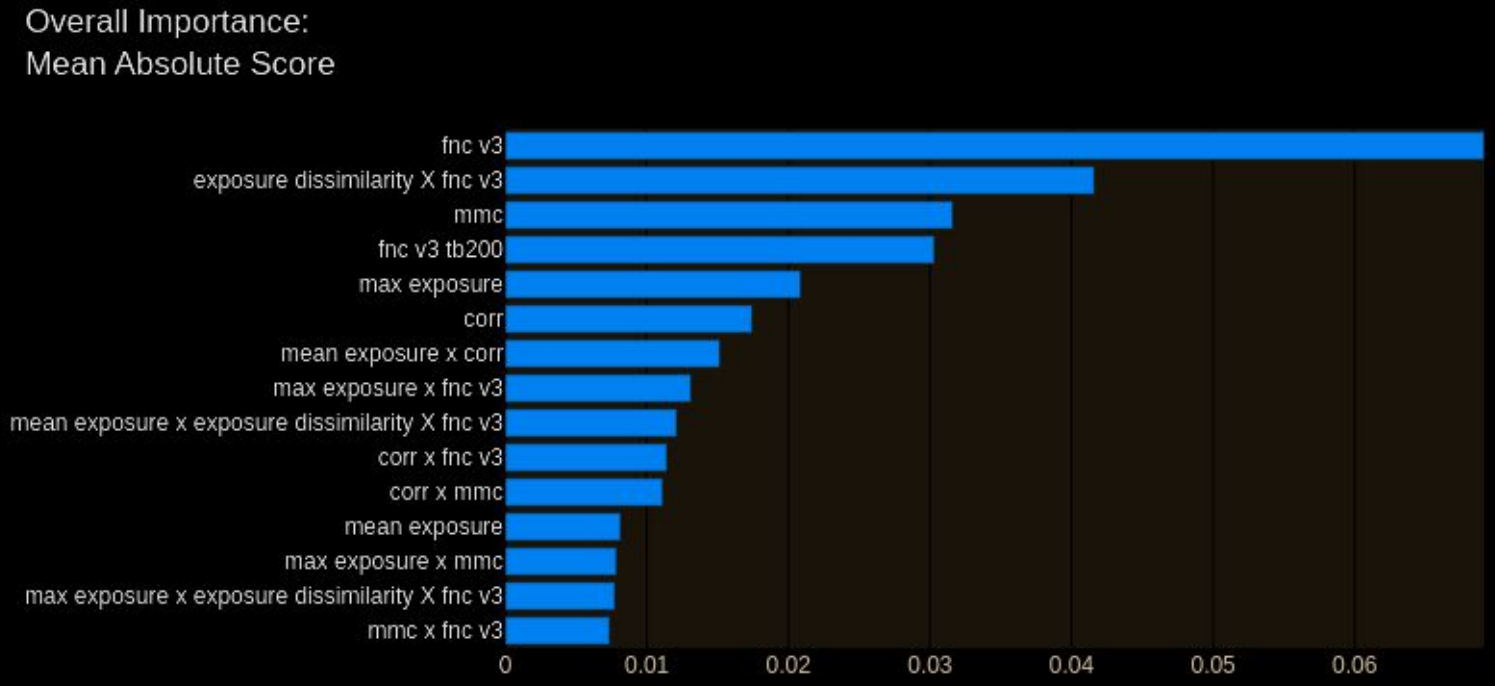
TC implemented in PyTorch w/ cvxpylayers

```
swmm = SWMModel(len(stakes), context=context, optimizer=n1_optimizer)

# set weights of linear layer to be user stakes
swmm.lin1.weight.data=stakes.T

swmm.zero_grad()

# get optimized portfolio and swmm signal
swmm_port, swmm_signal = swmm(user_preds)

# calculate portfolio returns and then stake gradient wrt returns
portfolio_returns = swmm_port.T @ stock_returns

# calculate gradient
portfolio_returns.backward()

# extract gradients from Linear stake weighting layer
stake_grads = swmm.lin1.weight.grad.numpy().copy()
```

Feedback drives evolution of Meta Model

- Gradients/TC scores move stakes in the right direction to improve the Meta Model
- People respond to incentives and improve their model to improve TC scores which improves the Meta Model
- Human in the loop stochastic gradient descent

What metrics are related to True Contribution?

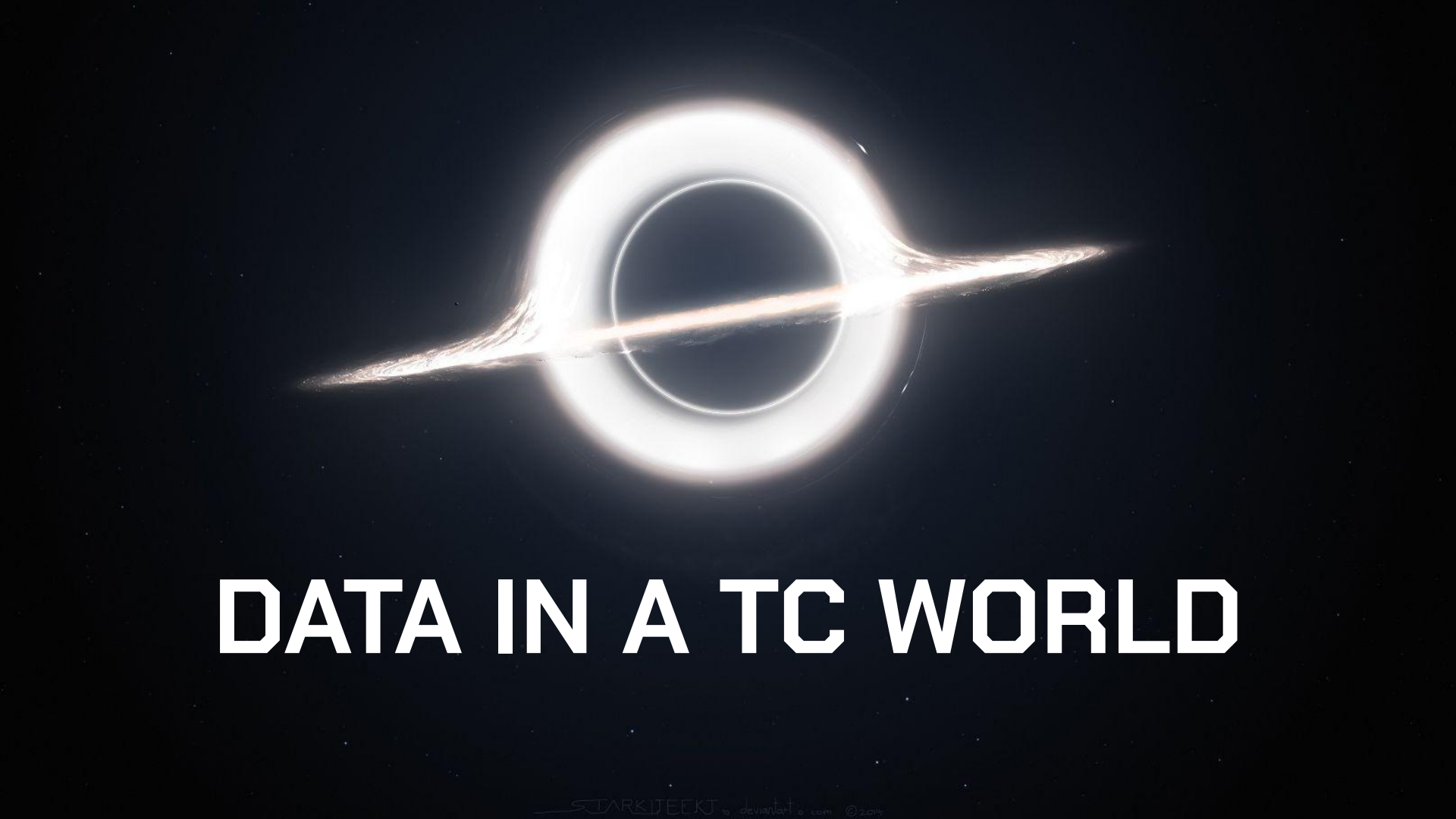


How to get TC?

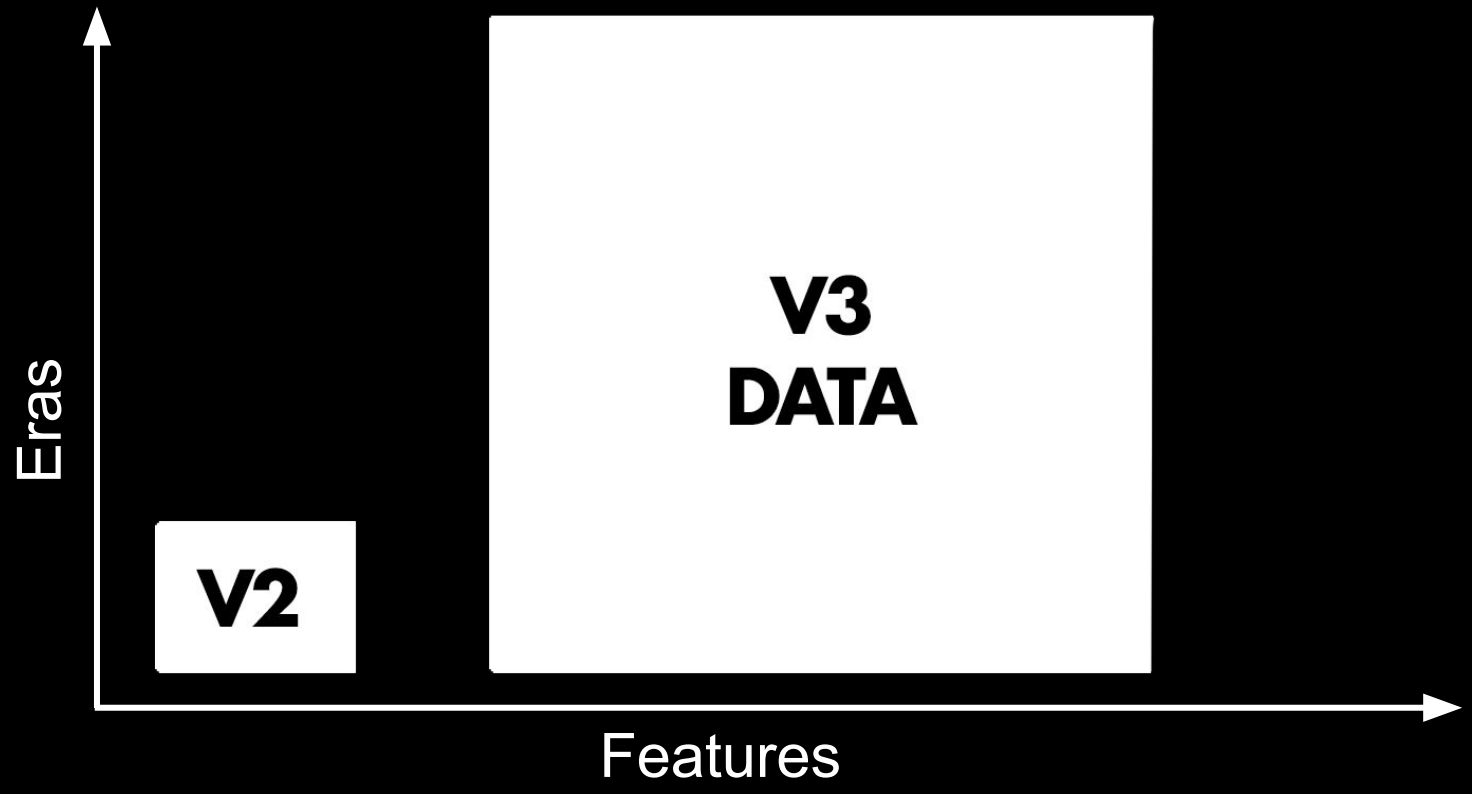
Make your signal …

1. have predictive power that isn't *linearly* explainable by the features
2. have predictive power in the tails of your predictions
3. have only low to moderate correlation with single features
4. unique

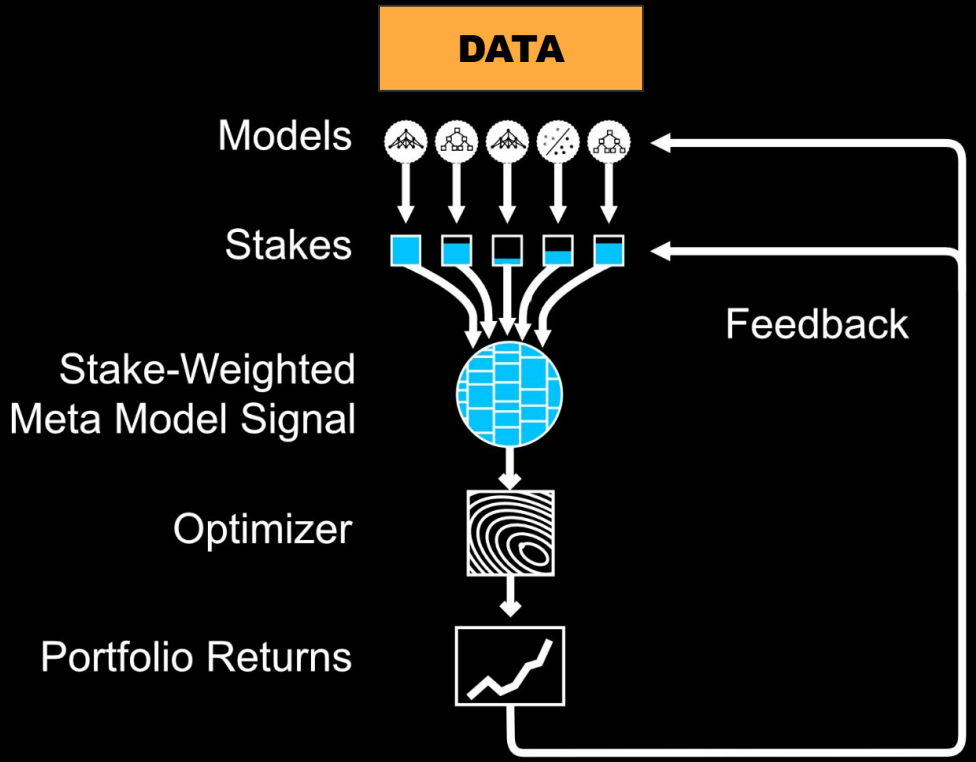
ii. DATA


DATA IN A TC WORLD

PROGRESS IN DATA



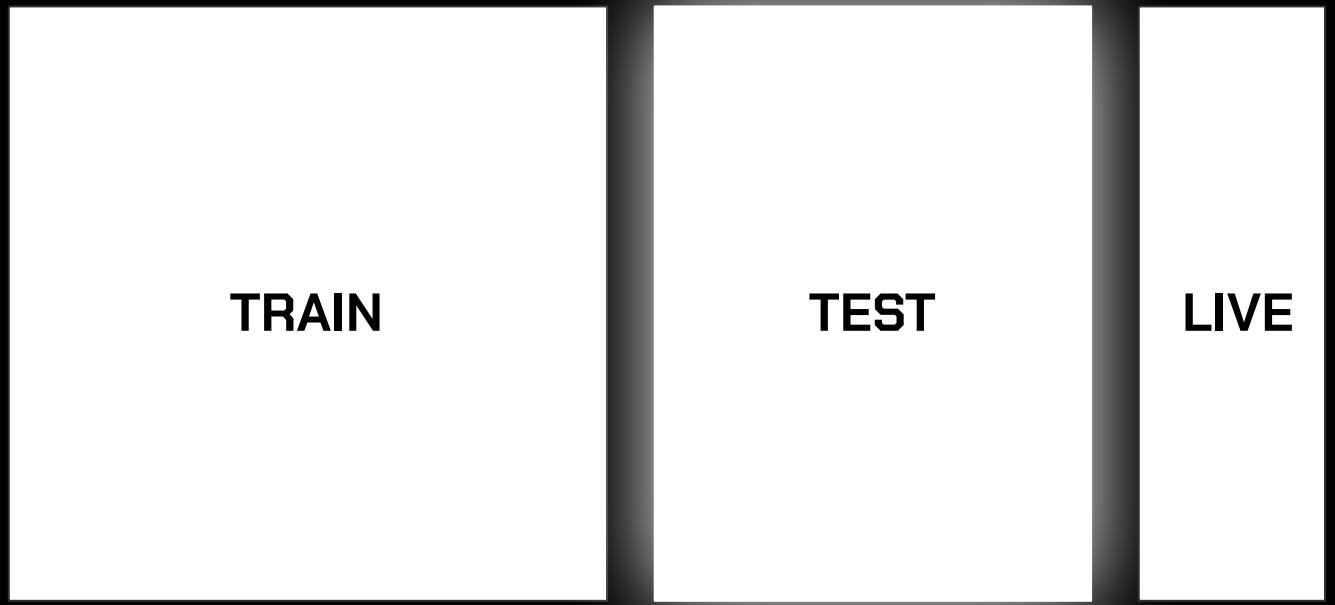
Data in the TC World



TEST DATA

TEST DATA

TEST DATA



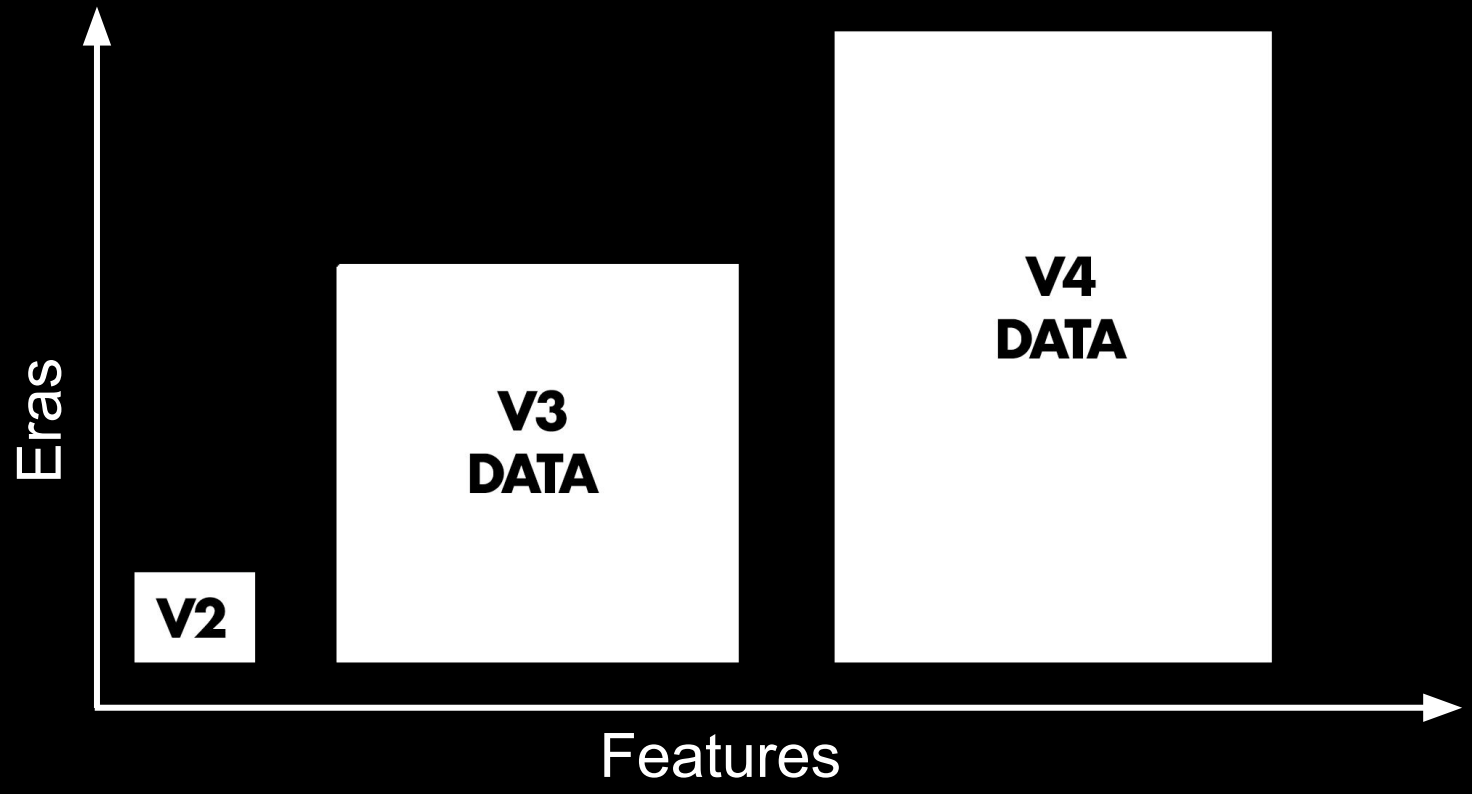
V4 DATASET

Over 100 new features

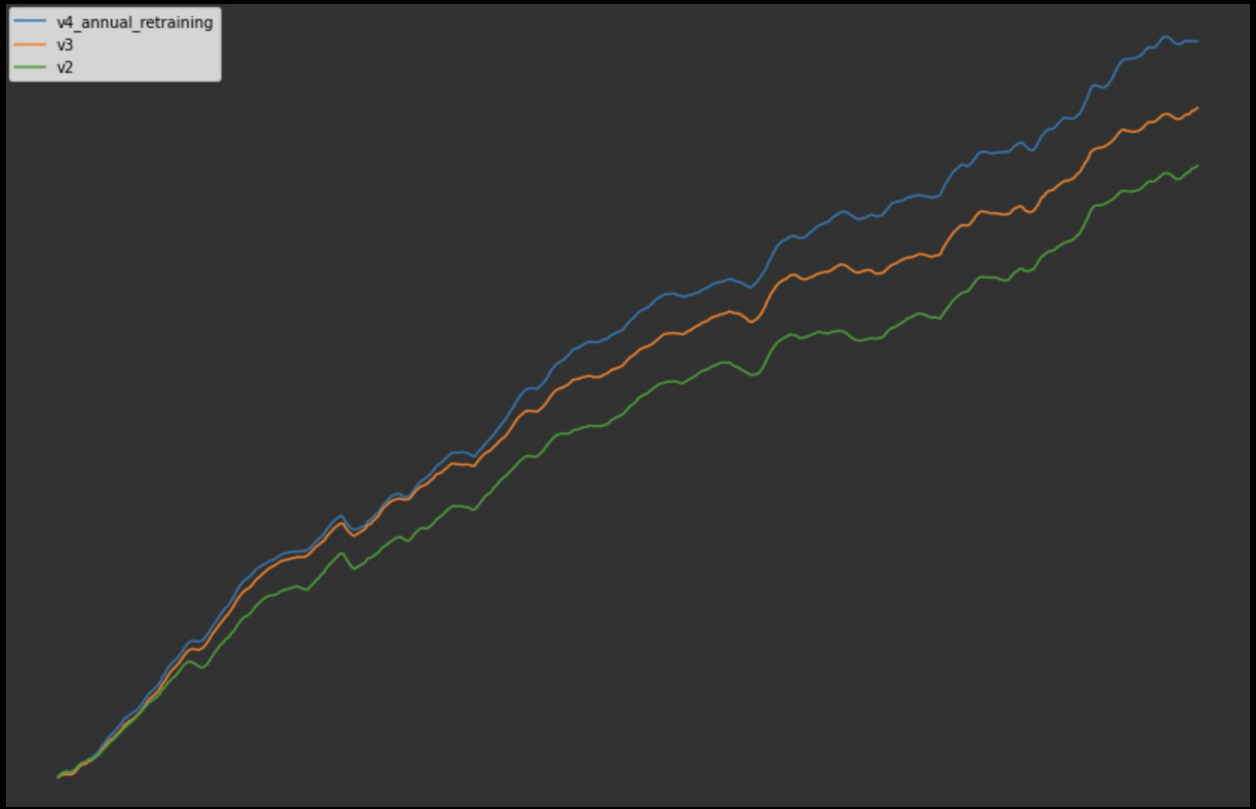
50% more training data

Newest data every week

V4 DATASET



V4 DATA PERFORMANCE



NEW MODELING POSSIBILITIES

How else can we help users get TC?

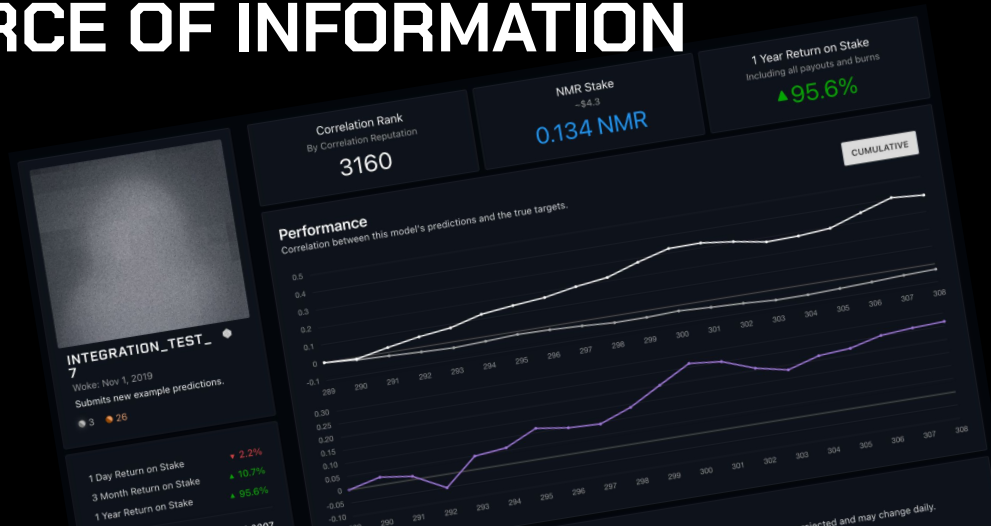
- Sharing modeling ideas
- Improving the research experience
- Giving more feedback on models

Diagnostics

Forum

NO CENTRAL SOURCE OF INFORMATION

Example Scripts

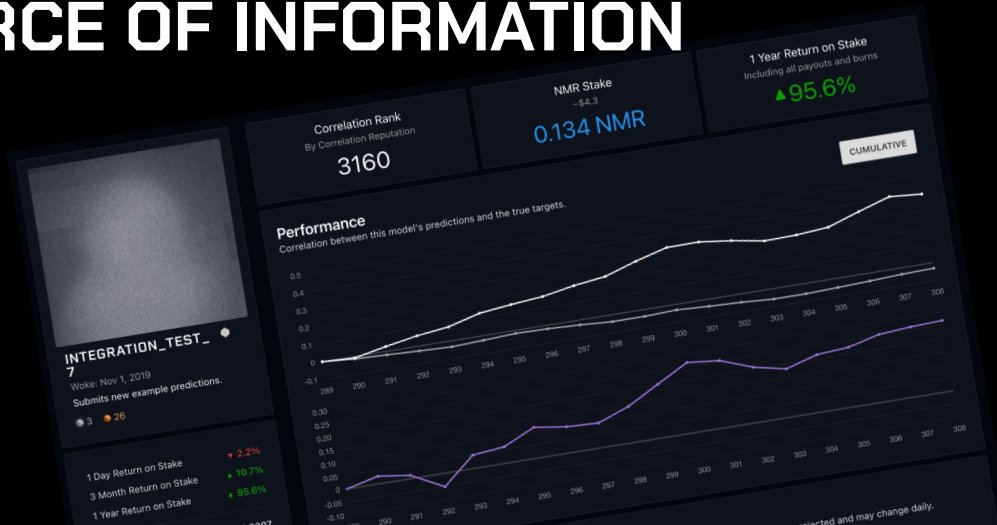


Diagnostics

Forum

NO CENTRAL SOURCE OF INFORMATION

Example Scripts



BENCHMARK MODELS

BENCHMARK MODELS

- Open-Source Team and Community built models
- Scores, including TC, backfilled for 2+ years
- Predictions available each week through the API

BENCHMARK MODELS

Model Name	1 Year TC	1 Year Corr	Corr w Meta Model
V2 Era-Boosting	-0.004	0.022	0.81
V4 Continuous Retraining	0.013	0.024	0.78
V3 UMAP Feature Augmentation	0.015	0.013	0.24

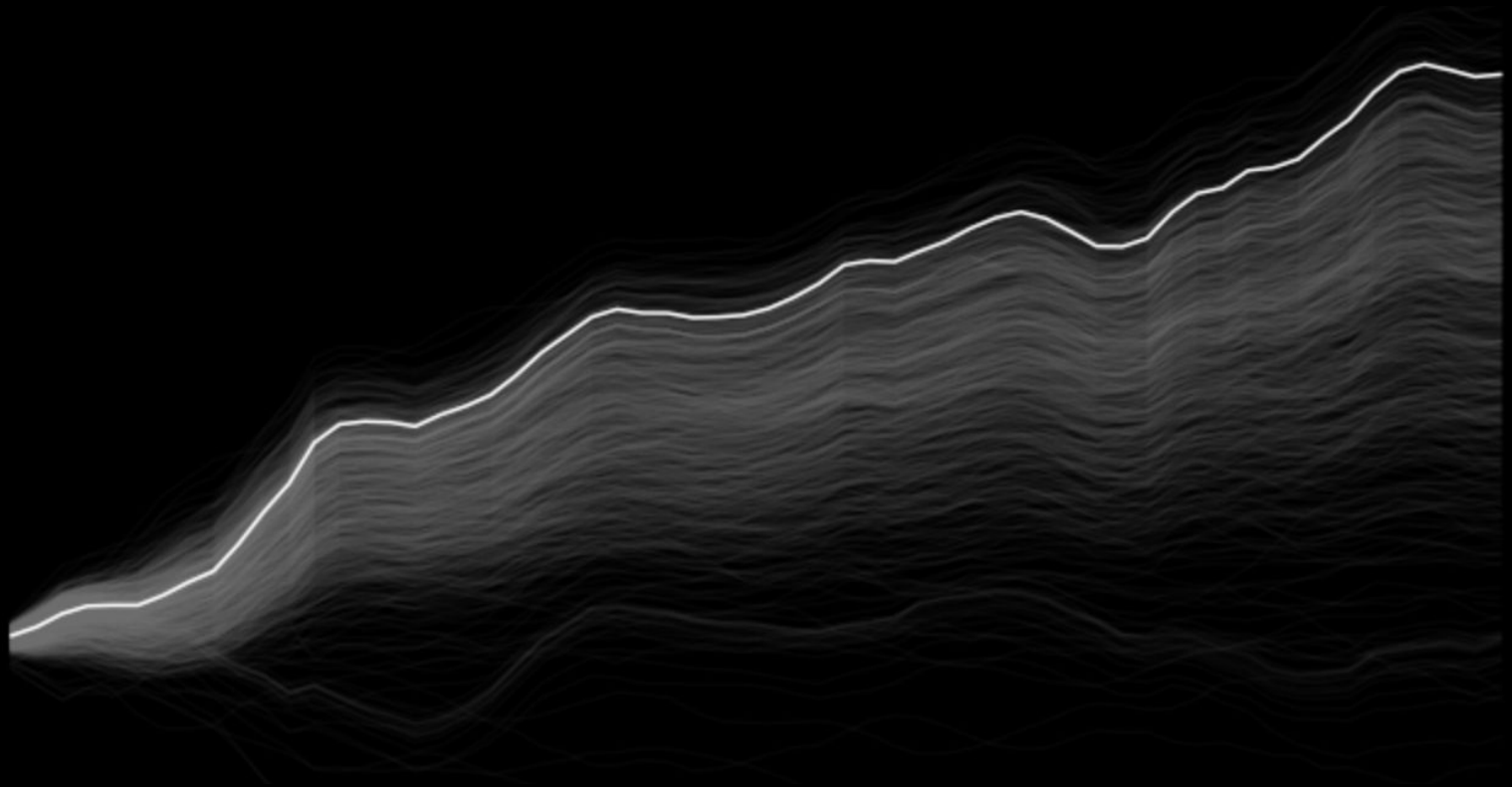
BENCHMARK MODELS

Model Name	1 Year TC	1 Year Corr	Corr w Meta Model
V2 Era-Boosting	-0.004	0.022	0.81
V4 Continuous Retraining	0.013	0.024	0.78
V3 UMAP Feature Augmentation	0.015	0.013	0.24

TYPE IV HEDGE FUND

- A signal which no one can make profitable alterations to
- True Contribution builds towards this directly
- The Data Team's responsibility is to give users the resources needed to approach this signal

THE ULTIMATE BENCHMARK



THE ULTIMATE BENCHMARK

- Sharing Stake Weighted Meta Model
- Allows comparison of performance
- Allows comparison of specific predictions

UNBLOCKING NUMERAI USERS

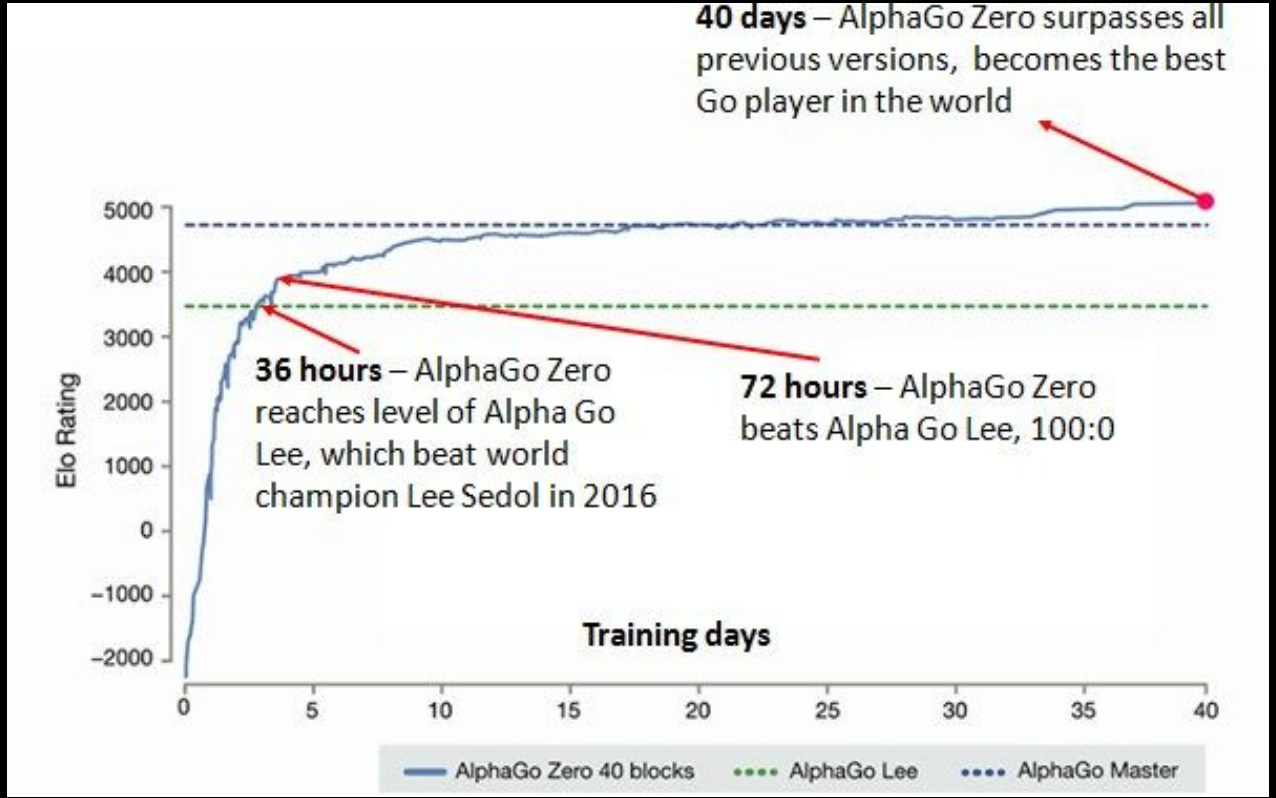
- Give all of the data in the world
- Share all of the experiments using the best models in the world
- Allow users to directly compare themselves with the Meta Model
- Type IV Hedge Fund


THE FUTURE OF DATA

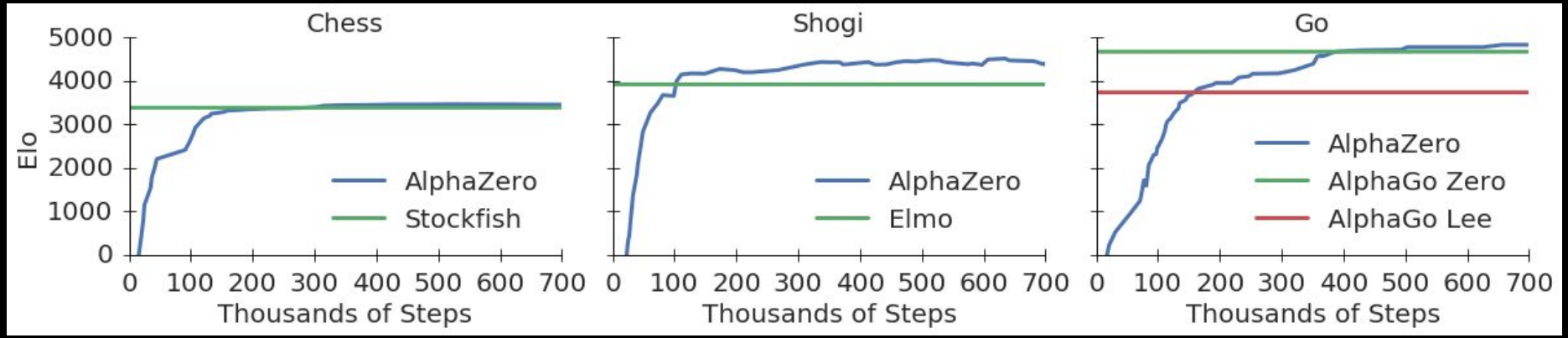
STOCK MARKET DATA IS LIMITED

How well could we perform if it wasn’t?

AlphaGo Zero became super-intelligent through self play



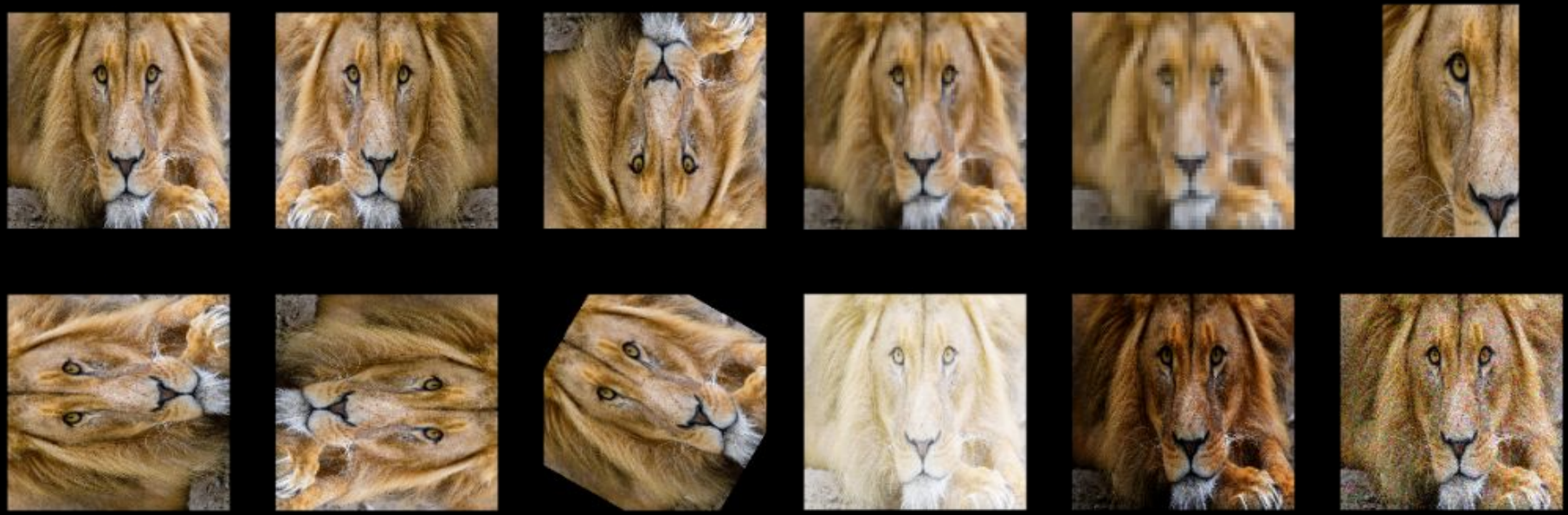
LEARNING VIA SELF PLAY



SYNTHETIC DATA

- P(features) - sample to generate realistic features
- P(returns | features) - sample to generate realistic return relationships
- P(returns, features) - sample to generate realistic features and returns

Sample P(features) with simple transforms



www.mygreatlearning.com/blog/understanding-data-augmentation

Sample P(features) with Deep Dream

ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

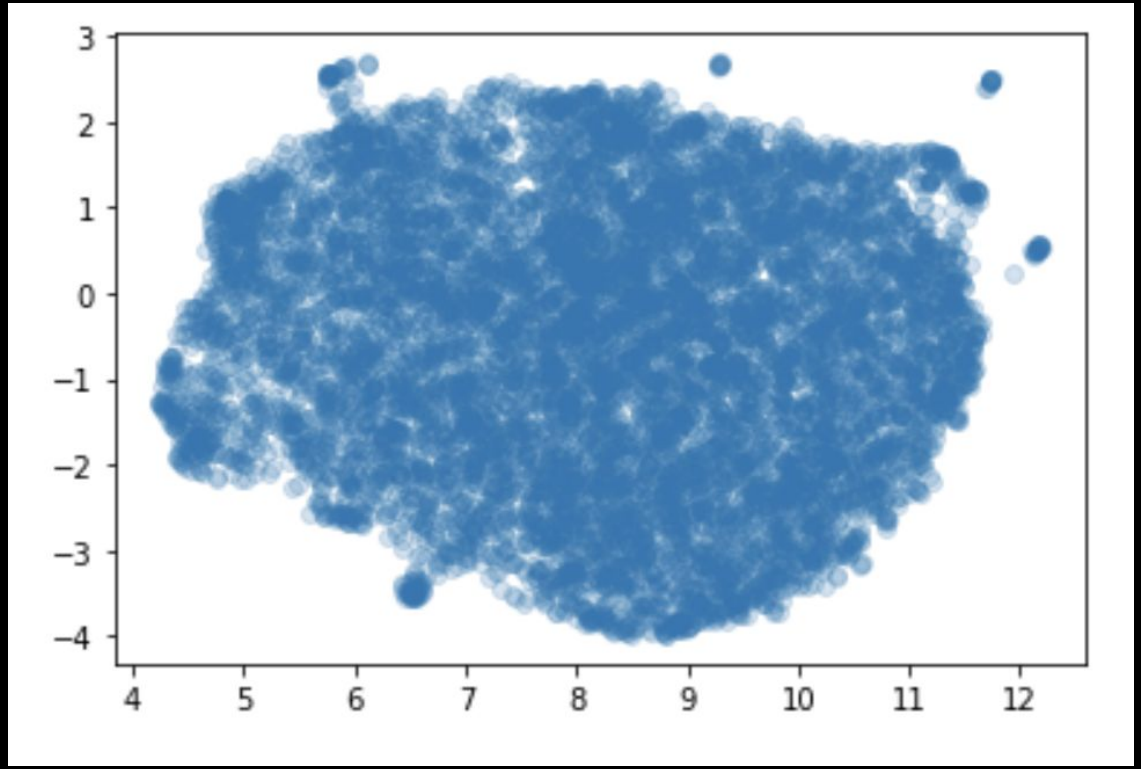
Can be used for any type of data!

credit: nyuton

Latent Space of P(returns | features)

credit:
jefferythewind

P(returns, features)



credit:
jefferythewind

Lots of ways to approach this problem!

- Can decompose joint distribution:
  P(returns, features) = P(returns | features) * P(features)

- Transformer architectures are a great fit for this kind of problem

- We would love to hear your ideas and collaborate!

∞ DATA

OVERVIEW

This is 1000 years of synthetic data created by a generative model designed to produce realistic features and targets for each era.

If you're not sure where to start, take a look at the example-scripts repository.

Otherwise you can download data directly from our API using numerapi:

```
> pip install numerapi
```

```
from numerapi import NumerAPI
napi = NumerAPI()
napi.download_synthetic(years=range(1000))
```

COMING SOON!

SYNTHETIC.PARQUET

1000 years of synthetic data and targets used to train your model

ORGANIC.PARQUET

Organic data from this reality used to validate or train your model. This expands every week.

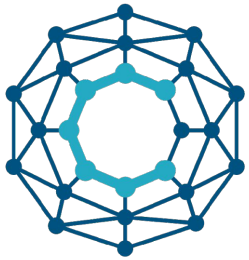
LIVE.PARQUET

The live data that your model predicts on. This changes every week.

COMMUNITY TALKS

NumerBlox

Tools for Solid Numerai Pipelines



CROWDCENT

Carlo Lepelaars

@carlolepelaars

perfect_fit

Jason Rosenfeld

@jrosenfeld13

jrai

CROWDCENT

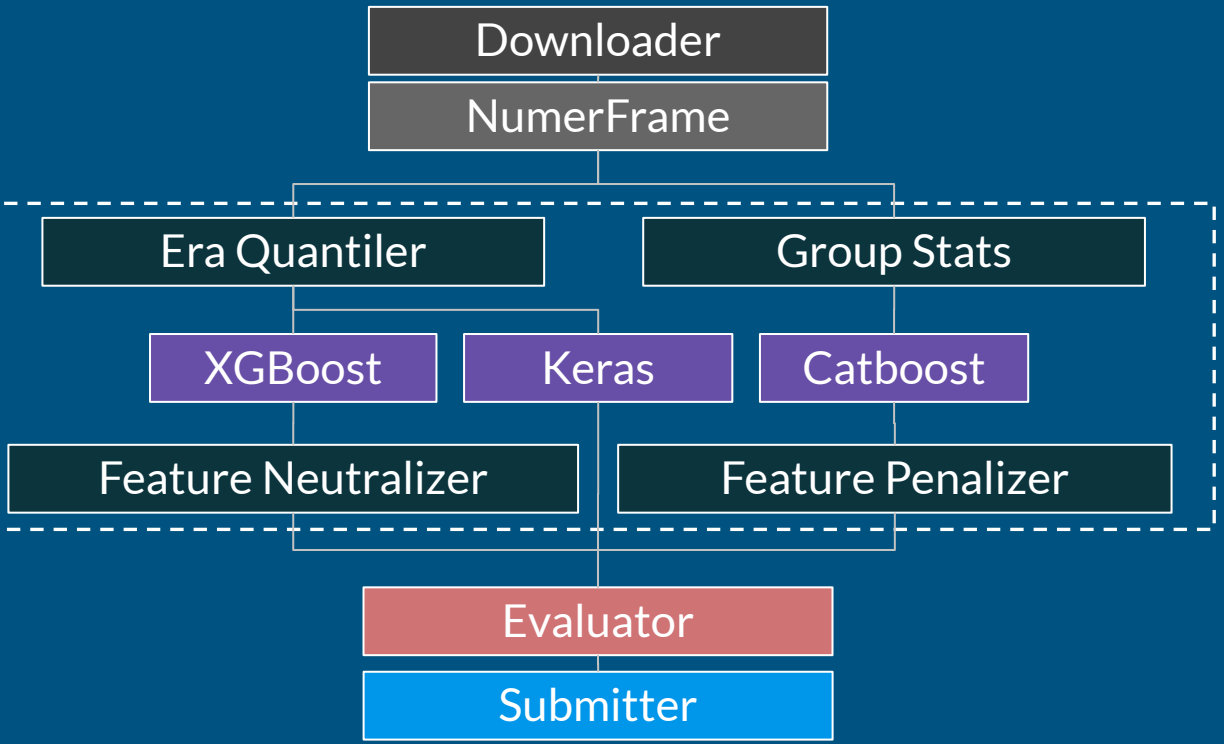
- Goal: Decentralize Investment Management
- Hedge Fund Legos
  - NMR denominated fund
  - 20+ Classic & Signals staked models
- Financial Network Effects
  - Code contributions
  - CoE donations
  - Shared learning
- Numerai benefits from teams, DAOs, and institutions

NumerBlox Context

- We needed:
  - Production framework for Numerai inference pipelines
  - Rapid iteration building and consistent model evaluation
  - Combinations of multiple models and processors
- Abstract the Numerai process into building blocks
  - e.g. Downloader, PreProcessor, Model, PostProcessor, Submitter
- Data Science first
  - Notebook development (nbdev)
  - Data integrity checks
  - Tested components

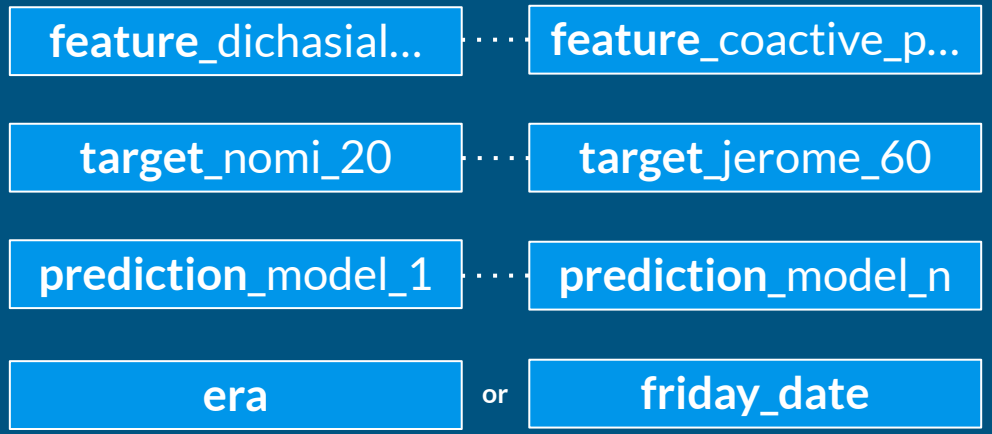
The “Blox”

1. Download/Load
2. Preprocess
3. Predict
4. Postprocess
5. Evaluate
6. Submit



NumerFrame

- Custom Numerai DataFrame
  - Extends Pandas DataFrame
- Standardized naming
  - feature, target, prediction, era
- Metadata



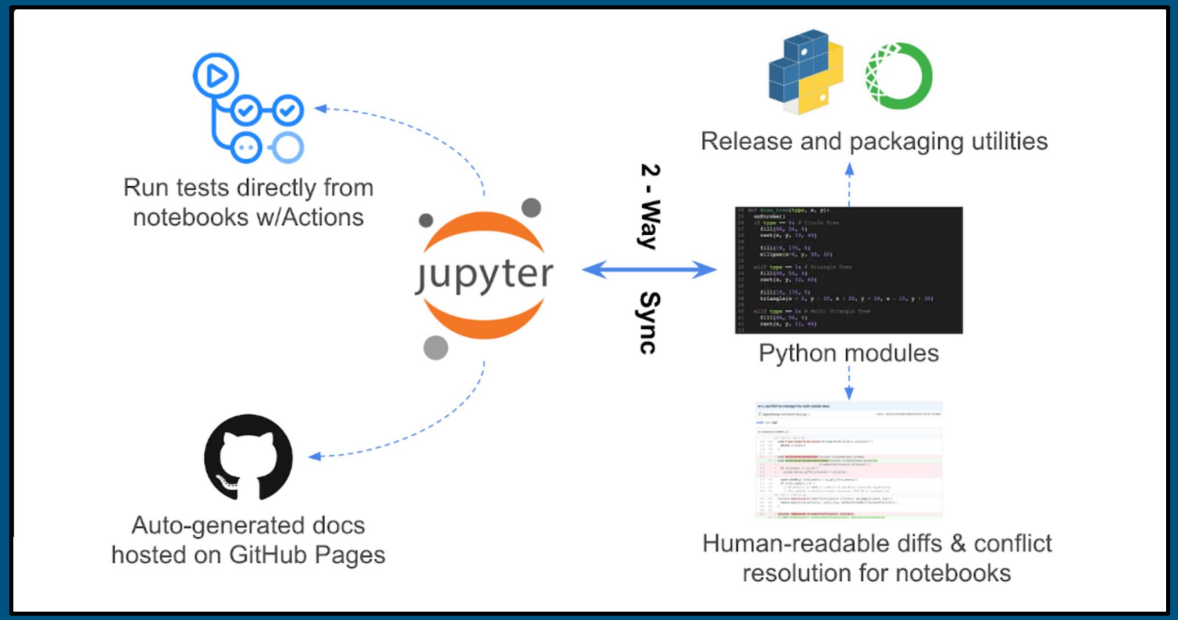

```
dataf = create_numerframe("numerai_validation_data.parquet",
                          metadata={"version": 2, "type": "validation"})
```

```
dataf.meta
```

```
{'era_col': 'era',
 'era_col_verified': True,
 'version': 2,
 'type': 'validation'}
```

fastai/**nbdev**

- Literate programming (Knuth, 1981)



Source: *Hamel Husain, 2020 (github.blog)*

NumerBlox Plans

- FNCv3 and Exposure Dissimilarity for Evaluators (TC proxies)
- NumerBay integration
- NLP (downloaders, preprocessors and models)
- Generative models
- (unsupervised) Autoencoders

pip install numerblox

Github

RocketChat

Carlo Lepelaars

@carlolepelaars

perfect_fit

CROWDCENT

Jason Rosenfeld

@jrosenfeld13

jrai

DISCLAIMERS

The factual information set forth herein has been obtained or derived from sources believed by the author and CrowdCent, LLC ("CrowdCent") to be reliable but it is not necessarily all-inclusive and is not guaranteed as to its accuracy and is not to be regarded as a representation or warranty, express or implied, as to the information's accuracy or completeness, nor should the attached information serve as the basis of any investment decision.

Past performance is not a guarantee of future performance. This presentation is not research and should not be treated as research. This presentation does not represent valuation judgments with respect to any financial instrument, issuer, security or sector that may be described or referenced herein and does not represent a formal or official view of CrowdCent. The views expressed reflect the current views as of the date hereof and neither the author nor CrowdCent undertakes to advise you of any changes in the views expressed herein. It should not be assumed that the author or CrowdCent will make investment recommendations in the future that are consistent with the views expressed herein or use any or all of the techniques or methods of analysis described herein in managing client accounts. CrowdCent and its affiliates may have positions (long or short) or engage in securities transactions that are not consistent with the information and views expressed in this presentation.

This document has been provided to you solely for information purposes and does not constitute an offer or solicitation of an offer or any advice or recommendation to purchase any securities or other financial instruments and may not be construed as such. Any such offering of the Fund will be made only in accordance with the terms and conditions set forth in the Fund's Private Placement Memorandum. No part of this material may be (i) copied, photocopied or duplicated in any form, by any means, or (ii) redistributed without CrowdCent's prior written consent.

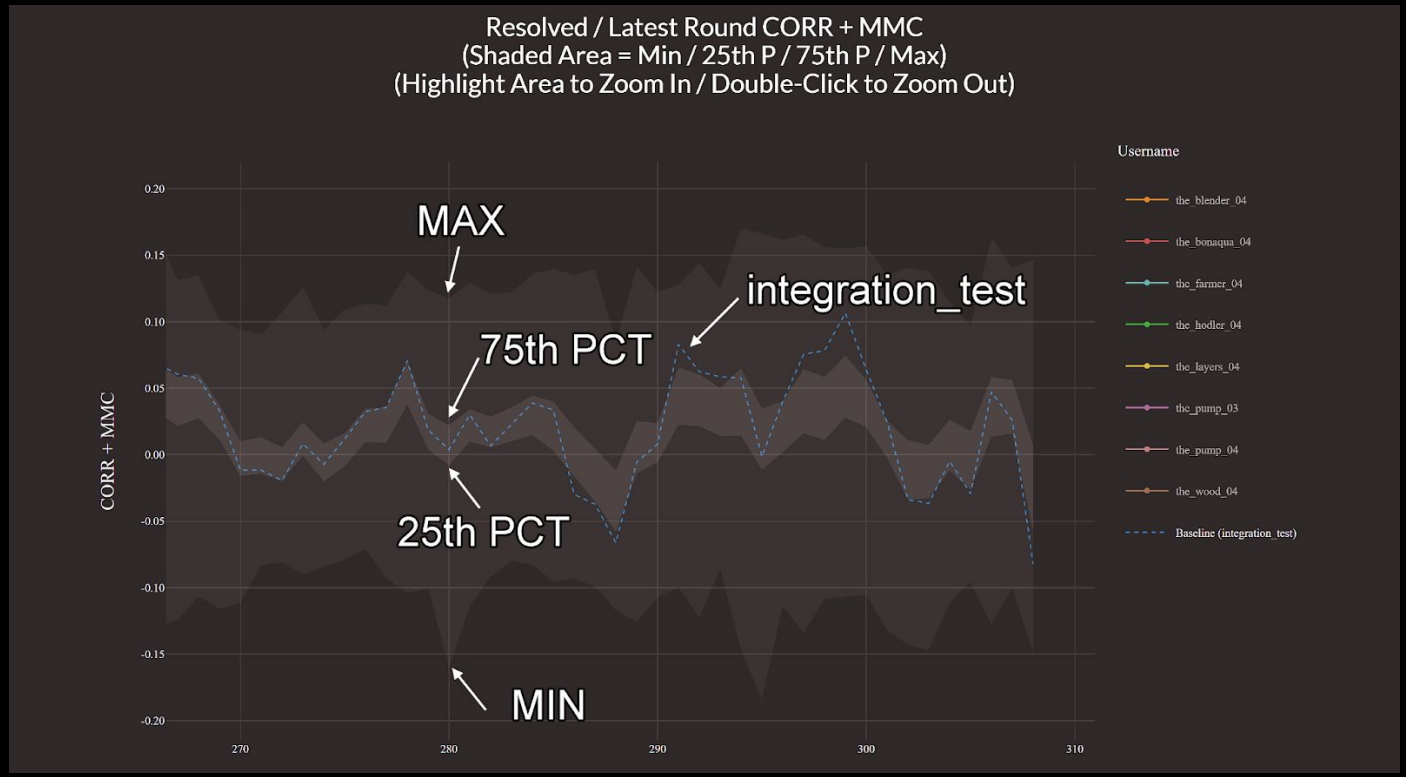
ABOUT JOE

Numerai: ia_ai
Real Name: Jo-fai Chow
Twitter: @matlabulous

Work:
Civil Engineer →
Data Scientist →
360 Photographer @ H2O.ai

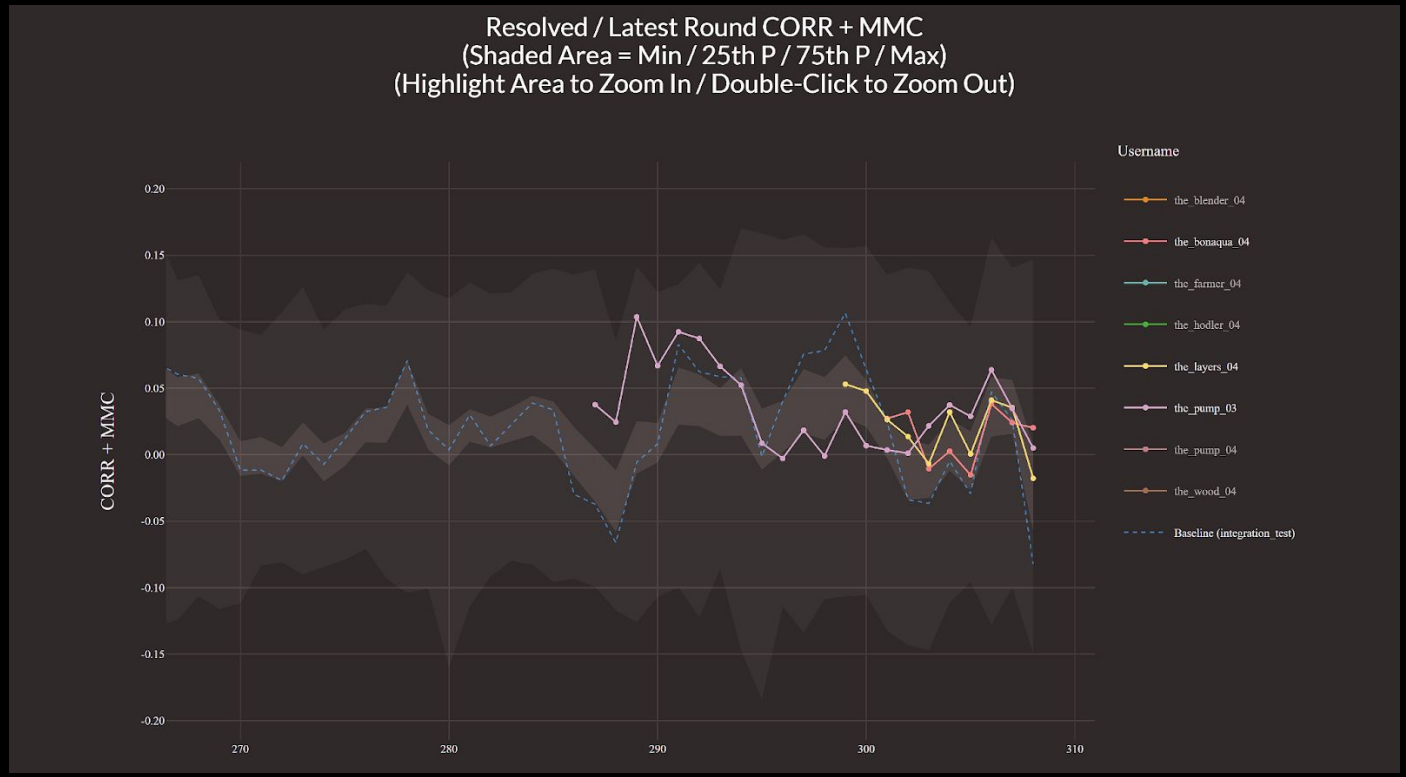
NUMERATI DASHBOARD

WHY?



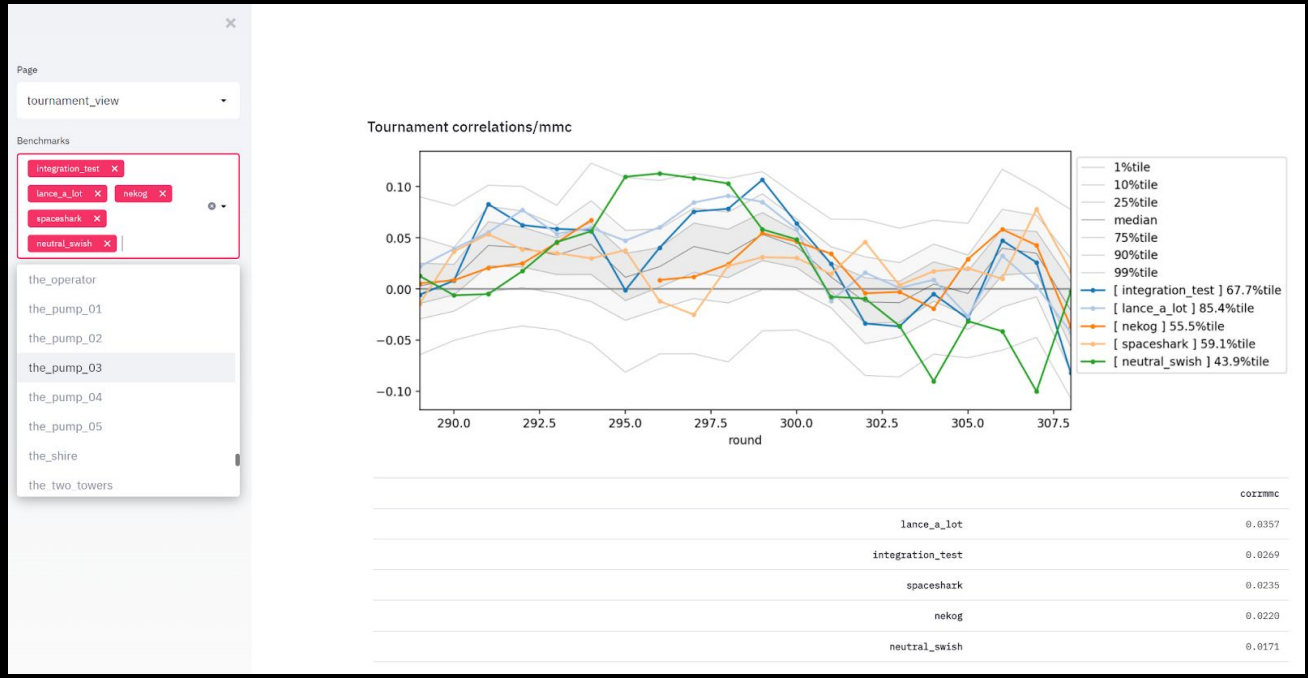
NUMERATI DASHBOARD

WHY?



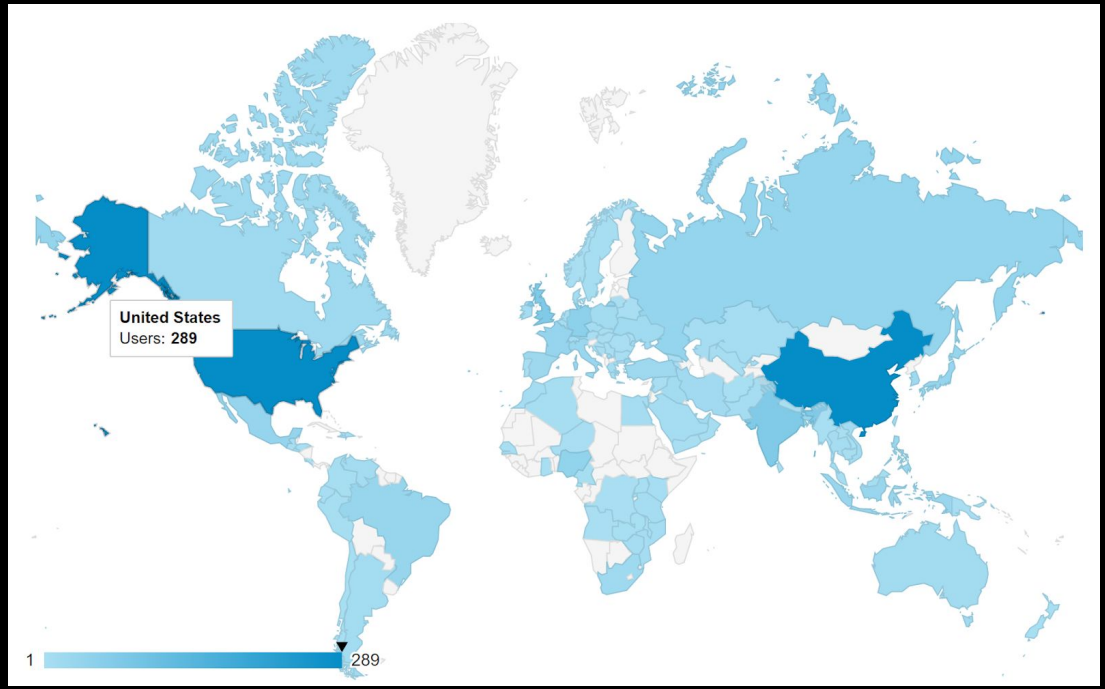
Joe’s CSV + Ceunen’s Streamlit

Joint Community Effort.



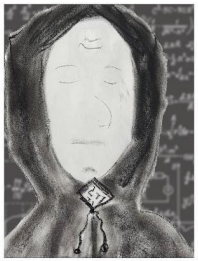
Numeratis
around the World.

1684 Users
106 Countries
since June 2020



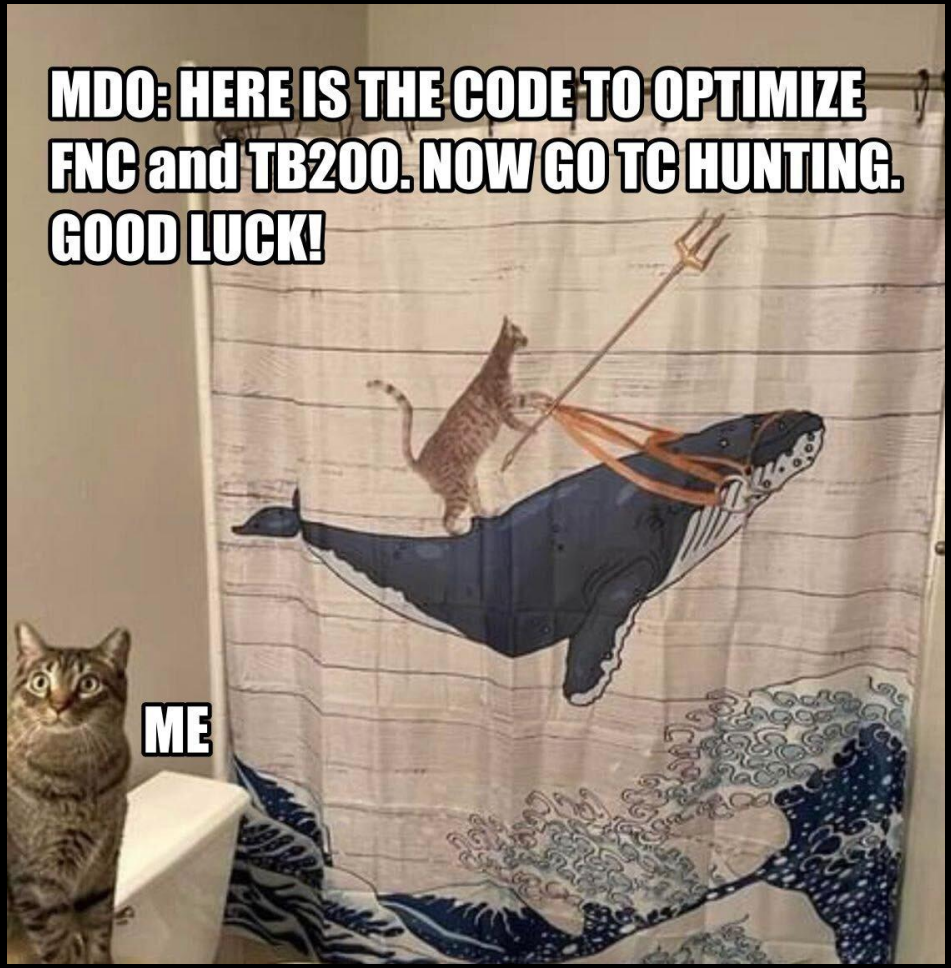
Life is Short.

Talk Memes to Me.

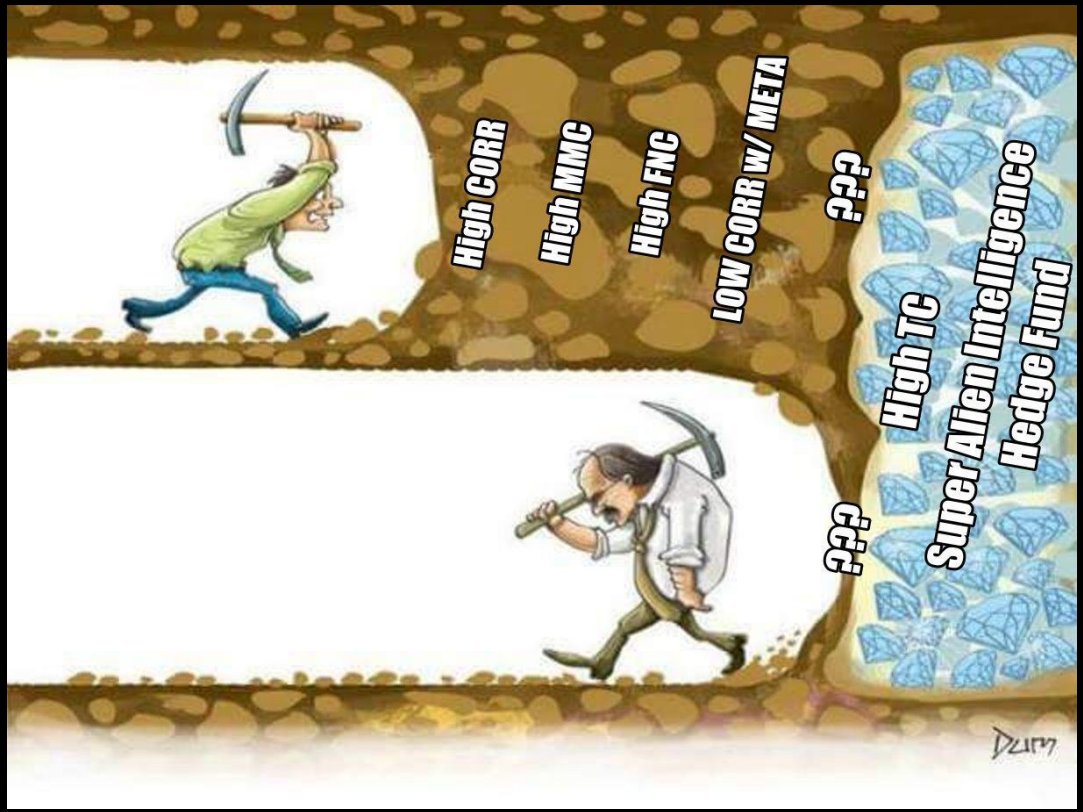


What is the Meta Model?

TC IS HERE.
LET’S GO!



WHERE ARE WE NOW?



iii. MONEY

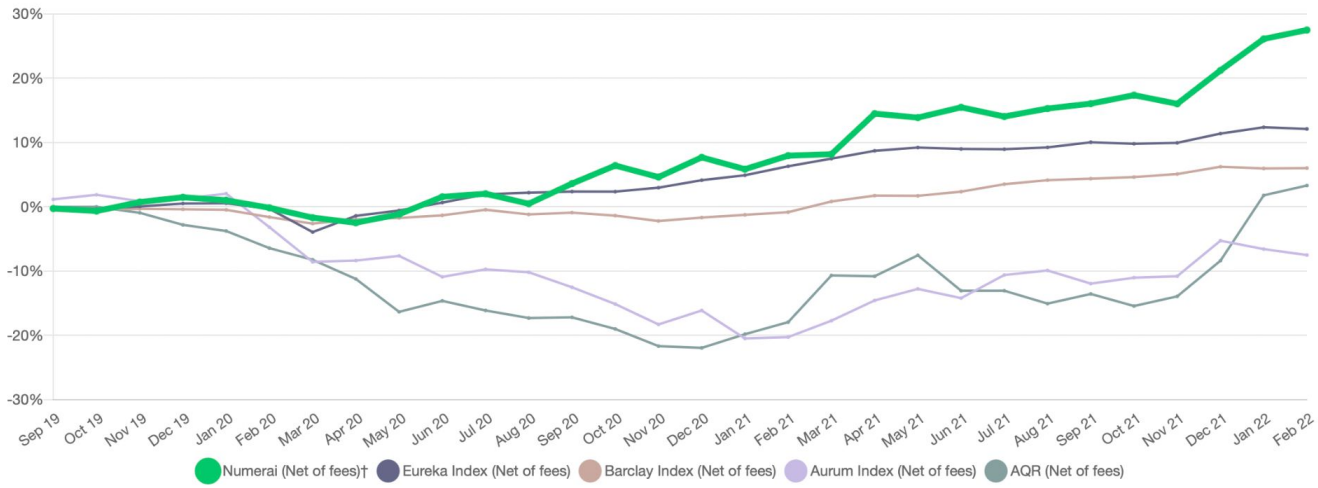
Performance results of Numerai's hedge fund is presented for information purposes only. Numerai's fund performance result was calculated net of management fee and incentive allocation, assuming a management fee at the rate of 1/12 of 1% (1% annually) of beginning net assets each month paid in advance and an annual incentive allocation calculated at a rate of 20% of trading profits. Rate of return is calculated by dividing net performance by beginning net assets. Performance of individual investors may vary based upon differing management fee and incentive allocation arrangements, and the timing of contributions and withdrawals. Returns are inclusive of the reinvestment of dividends and other earnings, including income from new issues. Returns may vary for investors who are restricted from participating in new issues. Performance results have been reviewed and audited by an independent accountant. ***The information provided is historical and is not a guide to future performance. Investors should be aware that a loss of investment is possible. No representation is being made that any investor will or is likely to achieve profits or losses similar to those shown.*** *Performance figures are net of costs and fees.*

UPDATE ON MONEY

- Numerai One, first institutional high Sharpe product
- Numerai One has 10x assets ($7m -> $70m)
- First institutional anchor investor, Canadian pension fund

...and over $200m in capacity taken

PERFORMANCE



+27.51%	+12.12%	+6.03%	-7.50%	+3.32%
Numerai	Eureka Equity Market Neutral Index	Barclay Equity Market Neutral Index	Aurum Quant Equity Market Neutral Index	AQR Equity Market Neutral

The information provided is historical and is not a guide to future performance. Investors should be aware that a loss of investment is possible. No representation is being made that any investor will or is likely to achieve profits or losses similar to those shown.

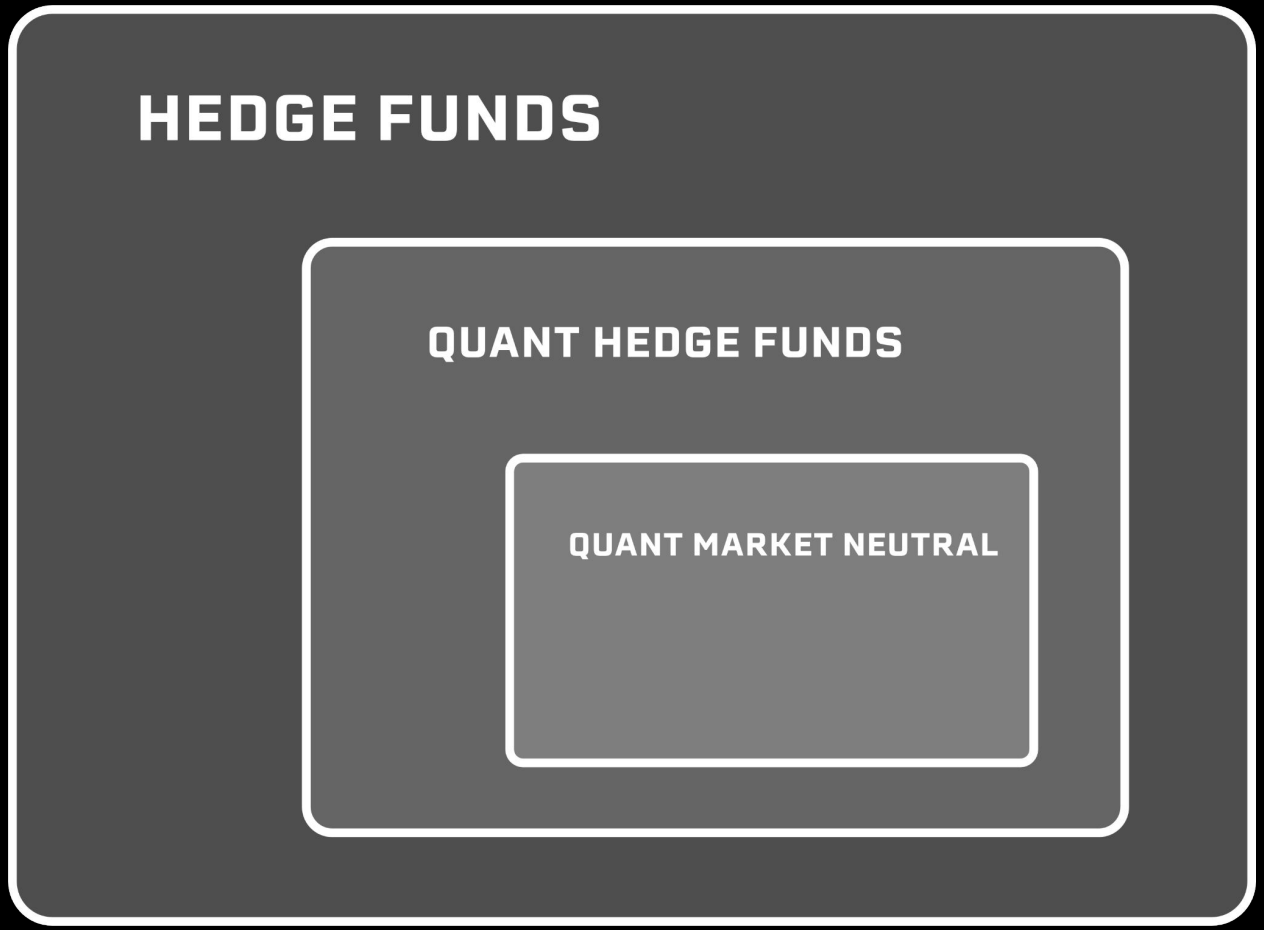
ACCELERATING PERFORMANCE

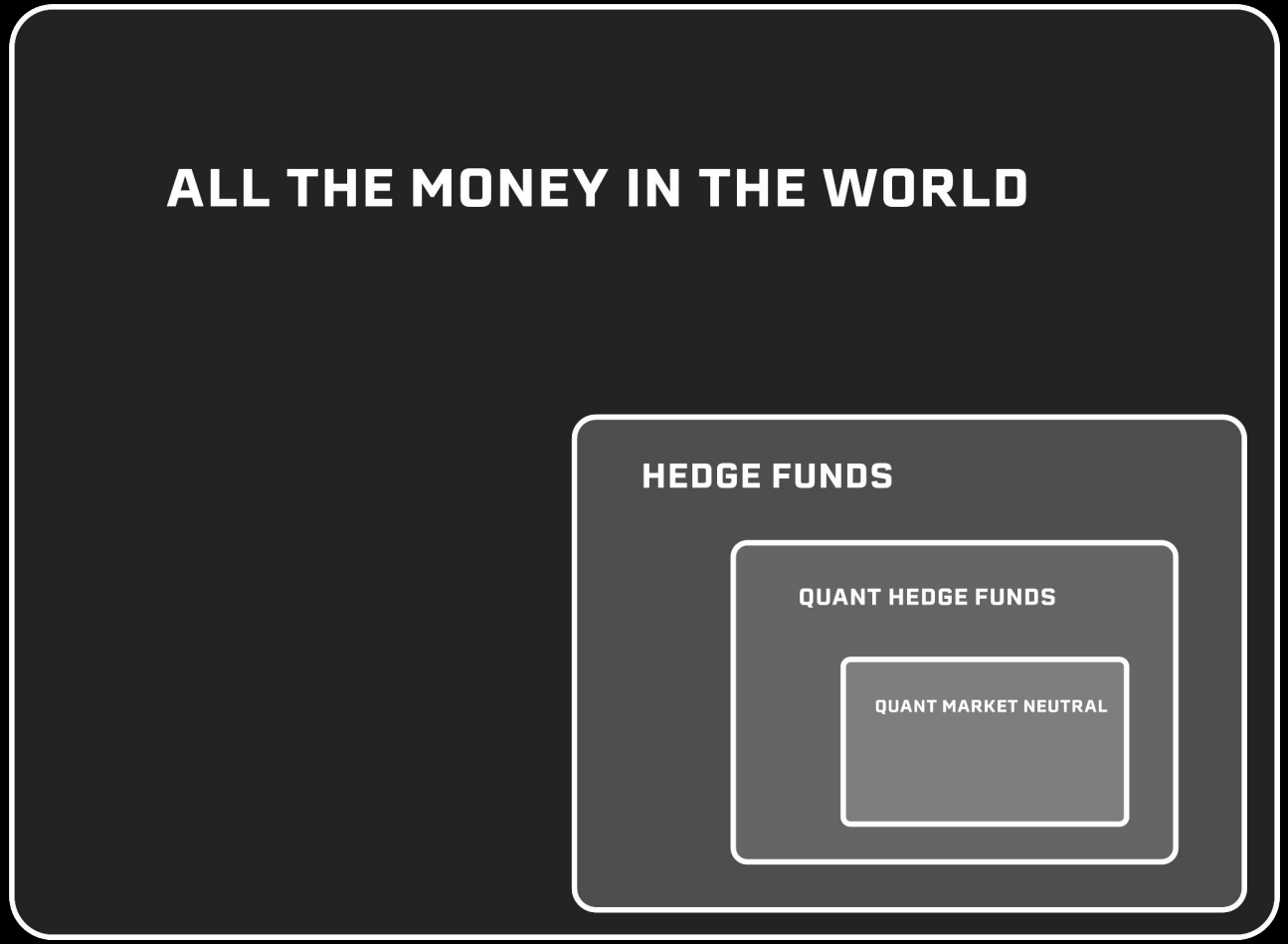
Sharpe ratio calculations are based on returns gross of fees and assume a risk-free rate of 0%

LTM Sharpe of 2.07

Sharpe ratio calculations are based on returns gross of fees and assume a risk-free rate of 0%

Capping Numerai One at $500m


HEDGE FUNDS
QUANT HEDGE FUNDS
QUANT MARKET NEUTRAL


ALL THE MONEY IN THE WORLD
HEDGE FUNDS
QUANT HEDGE FUNDS
QUANT MARKET NEUTRAL

NUMERAI SUPREME

NUMERAI ONE

Risk Management

Maximum position size	< 2% AUM, monitored daily
Market Neutral	Market, country and sector neutral
Factor Exposure	Beta of ~0 with limited exposure to core style factors
Target Volatility	< 10%

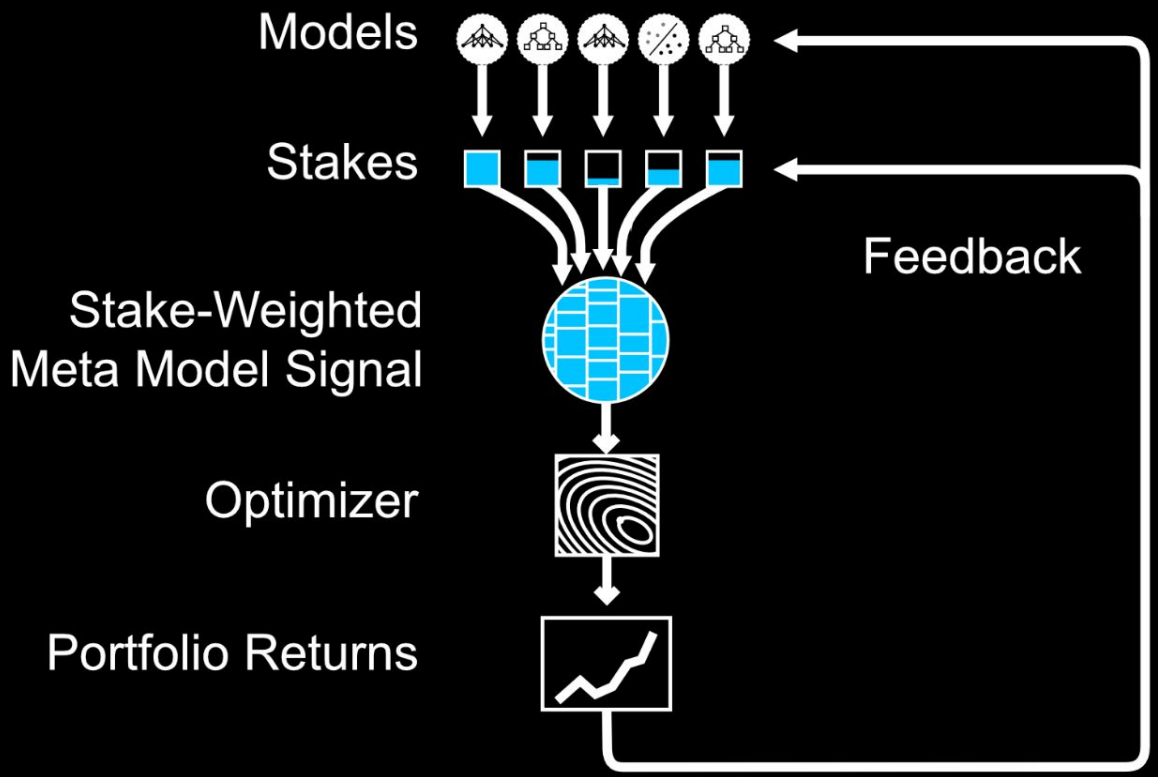
NUMERAI SUPREME

- *Believers only*
- High risk
- High lockup
- High fee
- Limited capacity $100m only


NUMERAI SUPREME

iv. DECENTRALIZATION

TRUE CONTRIBUTION



PROGRESS IN DECENTRALIZATION

- Staking 1.0 (2017)
  - NMR
  - Staking
  - User wallets

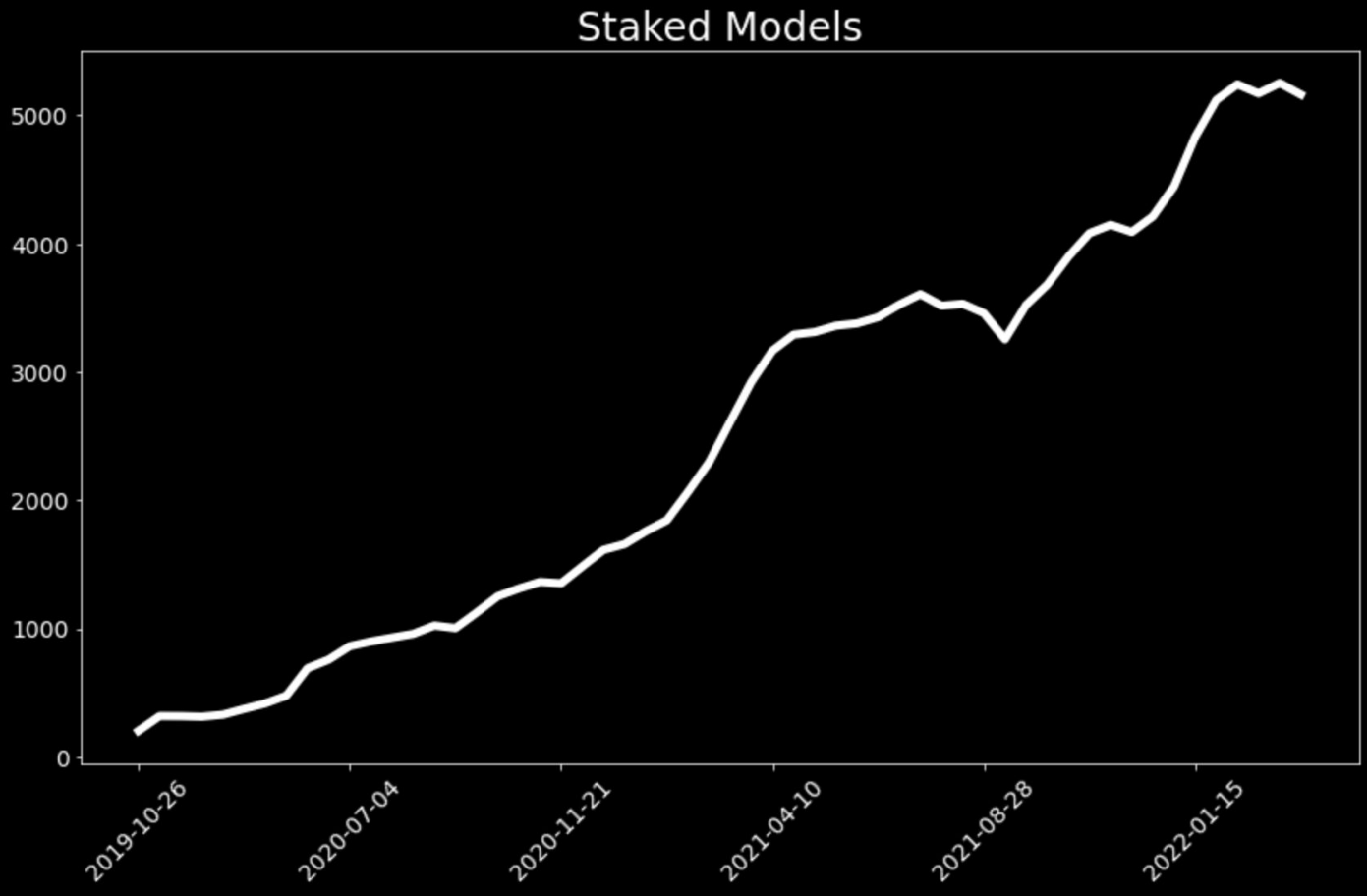
PROGRESS IN DECENTRALIZATION

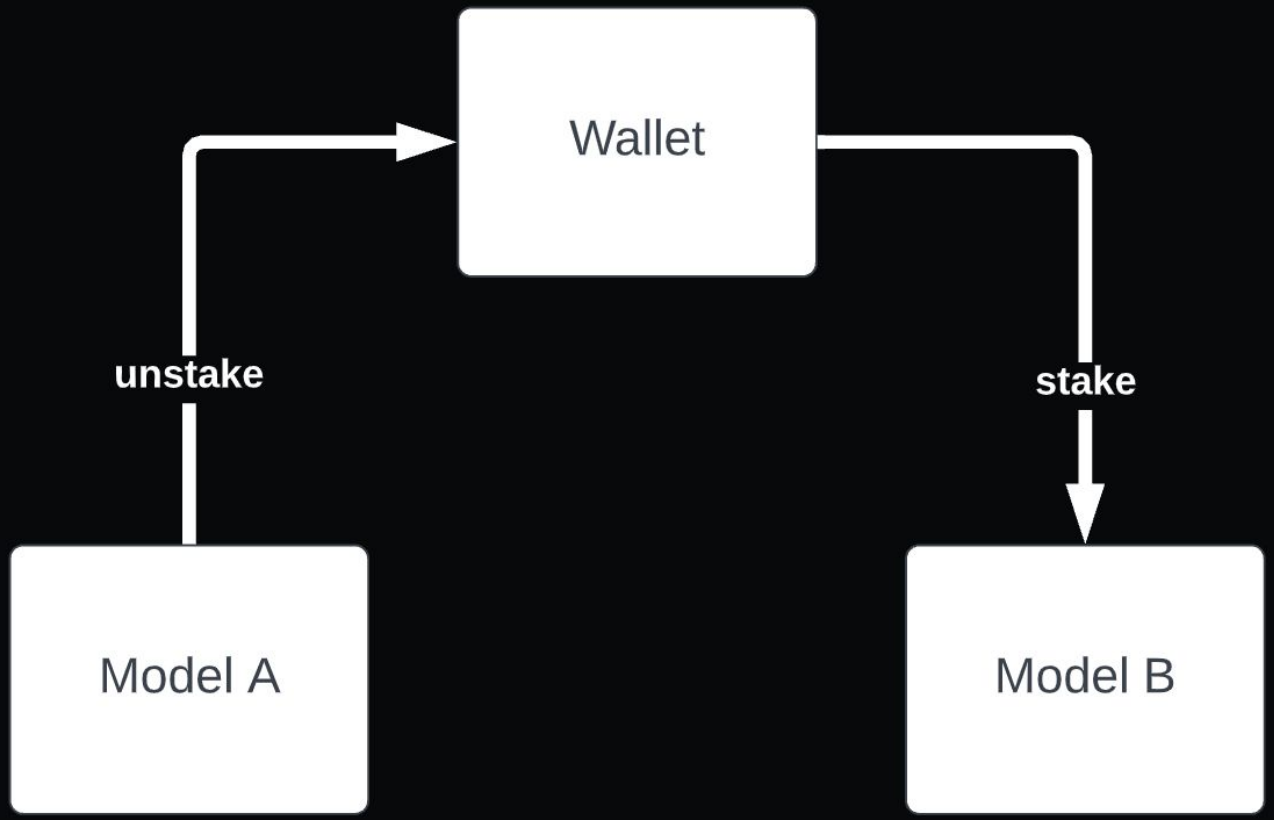
- Staking 1.0 (2017)
  - NMR
  - Staking
  - User wallets

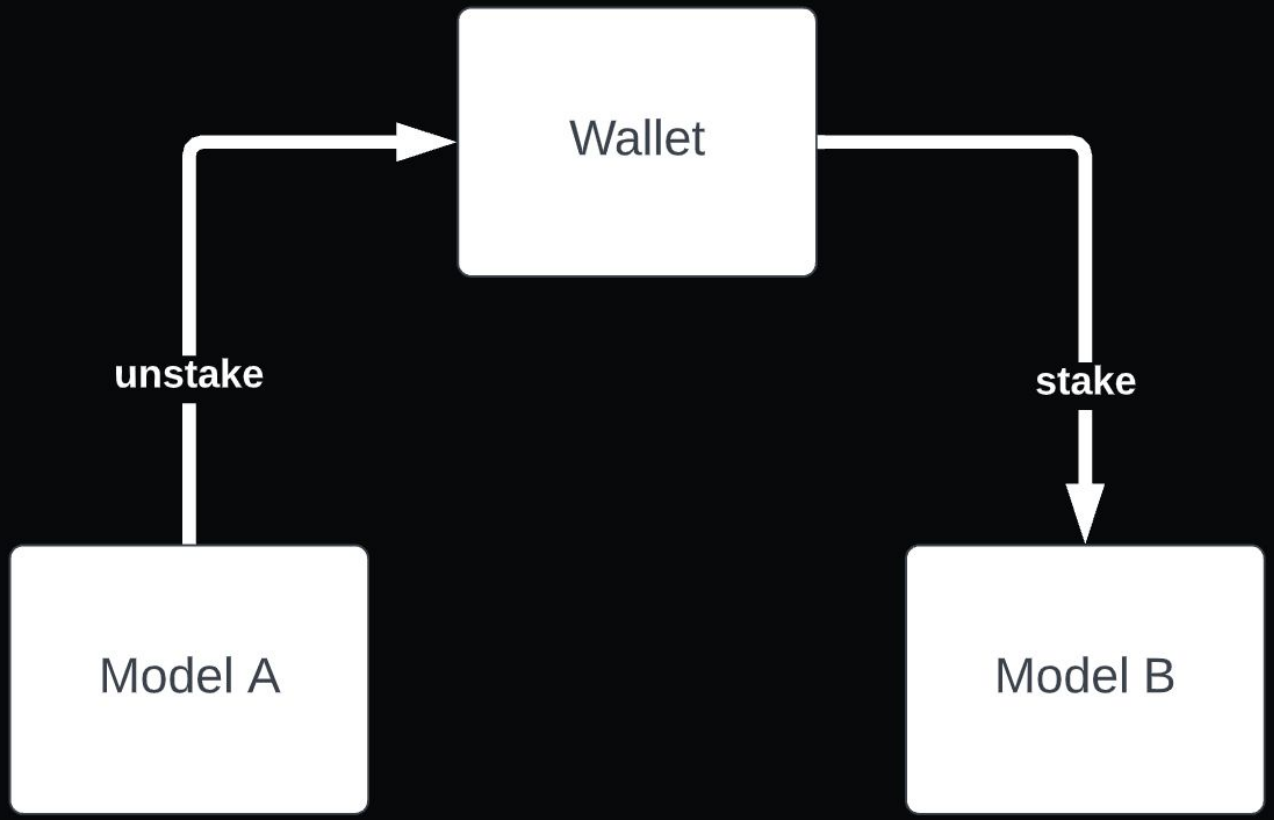
PROGRESS IN DECENTRALIZATION

- Staking 1.0 (2017)
  - NMR
  - Staking
  - User wallets
- Staking 2.0 (2019)
  - Automation
  - Continuous staking

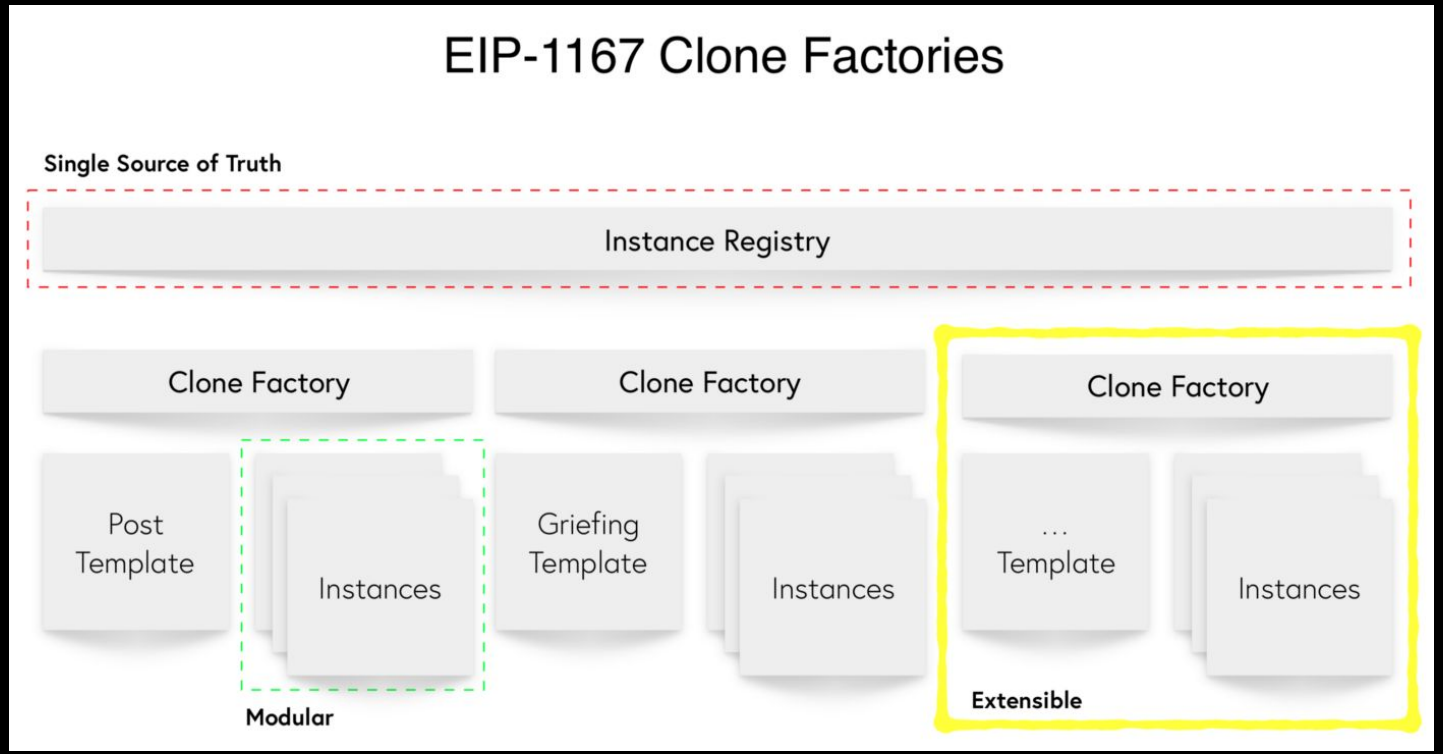
NUMERAI ON THE WEEKEND


Staked Models
5000
4000
3000
2000
1000
0
2019-10-26
2020-07-04
2020-11-21
2021-04-10
2021-08-28
2022-01-15

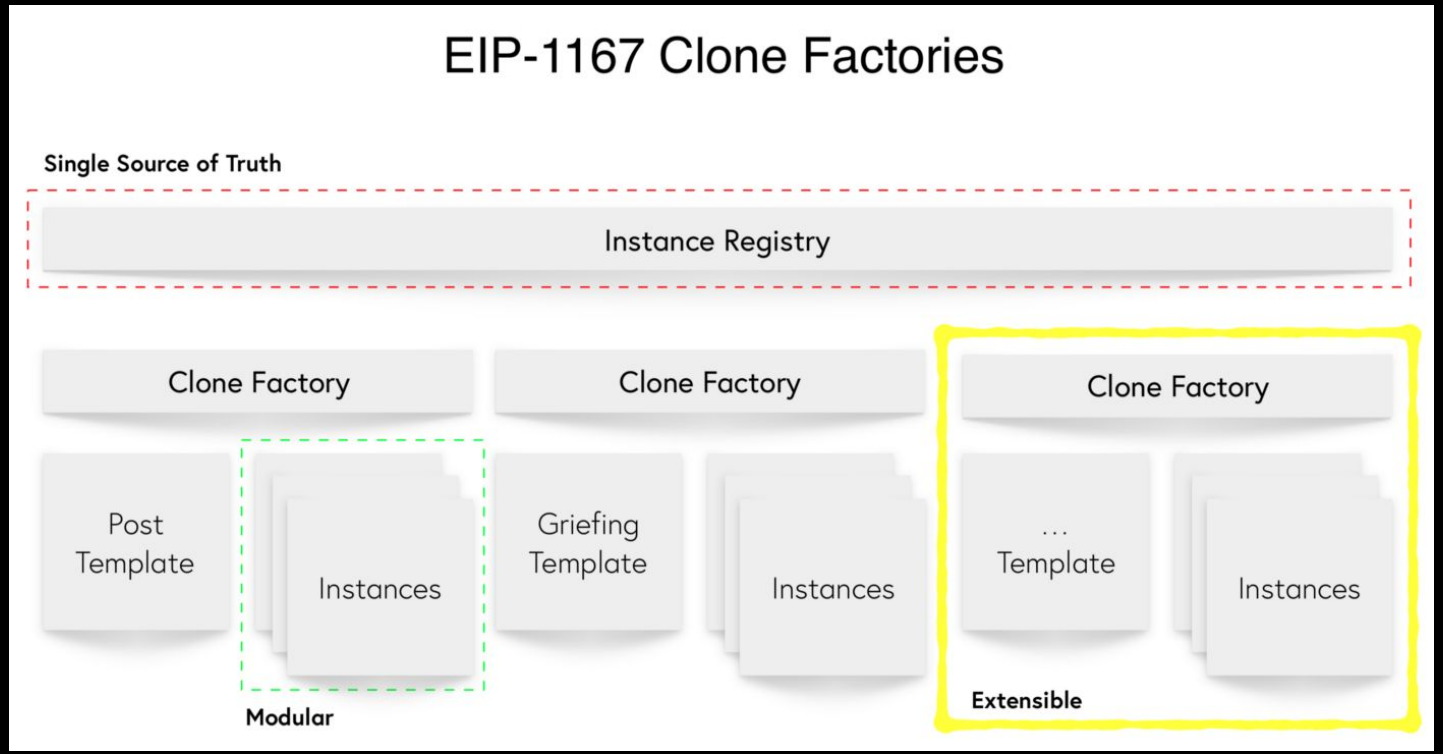

Wallet
unstake
stake
Model A
Model B


Wallet
unstake
stake
Model A
Model B

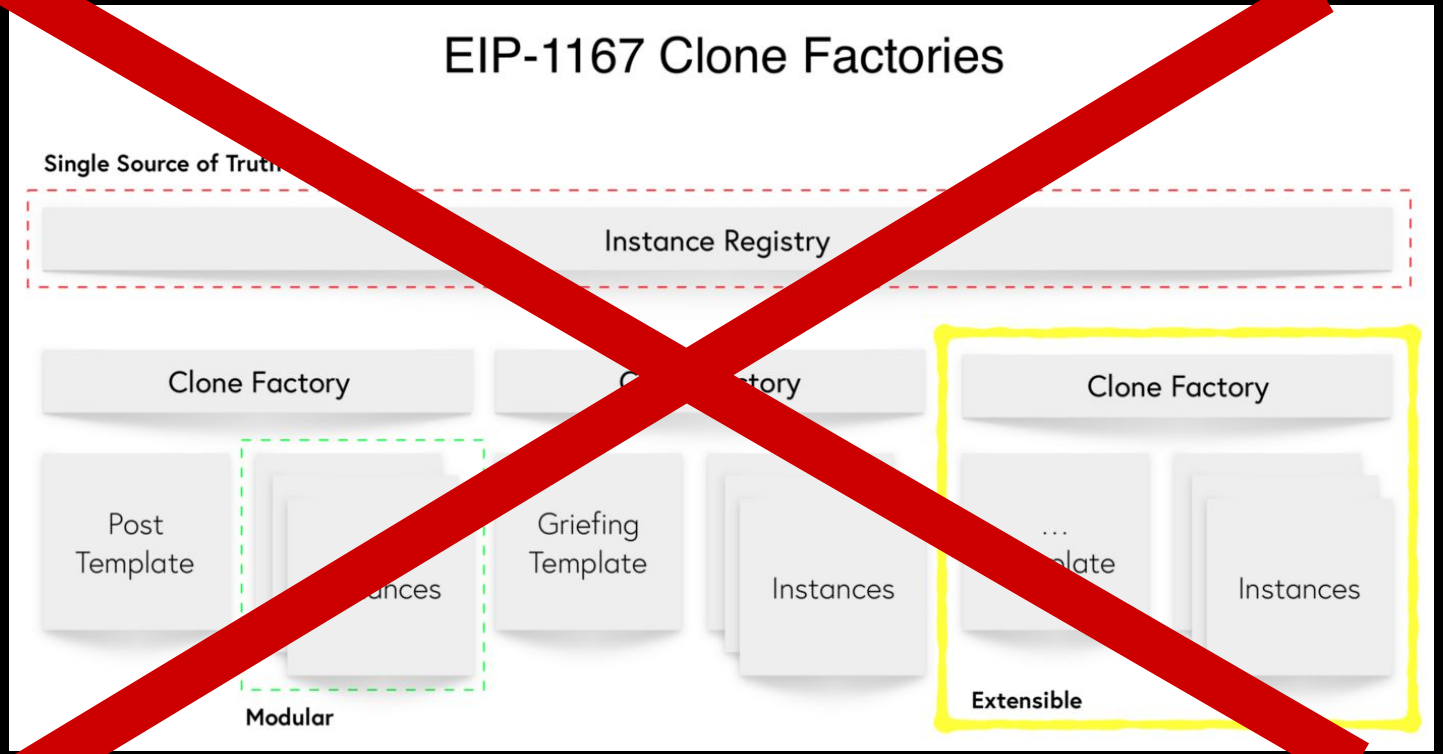
Erasure Protocol



Erasure Protocol



Erasure Protocol



✨ STAKING 3.0 ✨

🚀 Bring your own wallet

🦄 New tournament contract

⚡ Account level staking

NUMERCON 2022