# The AI Alpha Lab AI-model

## 1. BACKGROUND

From a given universe of assets AI Alpha Lab uses Artificial Intelligence (hereinafter **AI**) to forecast which assets to invest in. AI Alpha Lab creates various products around this and sells them to clients. Together, that is the purpose of the use of AI in AI Alpha Lab.

## 2. WHAT IS AI AND MACHINE LEARNING?

AI is a great buzzword, but today AI as such does not exist anywhere in the world. What we do have is advanced models that can solve complex domain specific tasks much better than humans, but the models do not understand the physical reality in which they live. The models are dependent upon human intelligence in order to specify the objective and the framework for solving the problem at hand.

At AI Alpha Lab we use machine learning. While AI encompasses the idea of a machine that can mimic human intelligence, machine learning does not. Machine learning aims to teach a machine how to perform a specific task and provide accurate results by identifying patterns. To put it simple, machine learning is a very complex analysis driven by computer power.

## 3. WHY USE MACHINE LEARNING ON INVESTING?

AI Alpha Lab strive to optimize the intersection between financial domain knowledge and machine learning with the purpose of making better forecasting models. *The investment decision in its very essence is a decision under uncertainty,* and we want to help investors with their two main problems:

1) *Data*: Investors are forced to navigate in an abundance of unstructured data and the variables investors try to forecast are non-stationary.

2) *Too simple models*: The data problem is being exaggerated by the models that investors are applying to the data, since these were never meant to provide optimal decision making under uncertainty. They are simply not fit for the problem.

The problem investors face tomorrow is not the same problem they faced yesterday and as such the variables of interest to investors (future returns, future volatility, and future correlations) are highly time-varying.

## 4. OUR AI-MODEL

Our objective at AI Alpha Lab is to create models that are fitted to the causal structures available in data and not to spurious correlations and noise. Machine learning and the increase in available computer power has facilitated a new way of fitting models which rely on a large number of generated models that are all capable of explaining data sufficiently well.

In AI Alpha Lab we can - in simple terms - create a large number of models explaining the data at hand and by using all of them (basically taking the average value) we can create a model which is superior at forecasting which assets to invest in. That is what we refer to as our AI-model.

The uniqueness of our AI-model comes from the team of quantum physicists sitting behind the model - and not from machine learning. Quantum physics is tailormade for solving problems associated with significant uncertainty and uses probability to describe unobservable quantities like future returns on financial assets.

Machine learning is what facilitates the use of concepts from scientific fields such as quantum physics, enabling us to build models that, contrary to the models used today, are suited for solving complex problems like the investment problem.

At AI Alpha Lab we employ a Bayesian multi-layer perceptron (neural network) to do our inference (a conclusion that is drawn from evidence and reasoning) on the assets in question. By doing Bayesian inference we move from a paradigm postulating that the model that best fit the data equals the true model and instead postulating that the simplest model that generalizes the best is the most likely to be the true model.

This approach not only provides us with a more robust forecast of the future, but we are also provided with the uncertainty surrounding the forecast. In other words, our AI-model estimates both the future return of the asset in question and the probability (or uncertainty) of the return estimate.

### *4.1. Uncertainty*

Optimal decision making under uncertainty requires a set of possible scenarios and the associated probabilities with which these scenarios are realized. The scenarios or outcome space is in itself useless if we cannot quantify the uncertainty surrounding each future state, known as predictive uncertainty.

Predictive uncertainty is composed of two types of uncertainty:

1) *Data uncertainty*: Unstructured noise in data that has no causal explanatory power.

2) *Model uncertainty*: A measure of how uncertain the model of choice is with regards to its ability to explain structure in data.

Broadly speaking, investors today thinks that data uncertainty equals predictive uncertainty. As a result, model uncertainty is a large uncompensated risk inherent in most investment portfolios today. Using a Bayesian neural network, we can quantify model uncertainty and incorporate it into the investment process.

## 5. DEVELOPMENT

### *5.1. The base AI-model*

The base AI-model upon which the AI Alpha Lab AI-model is built is developed and maintained by Desupervised ApS. Desupervised ApS continuously develop the base AI-model, and all development is supervised by AI Alpha Lab co-founder and CTO Michael Green, who is the founder and CEO of

Desupervised ApS. With Michael Green being strongly associated with both companies, there is a strong control with the development of the base AI-model seen from an AI Alpha Lab perspective.

A new base version of the AI-model is very rarely applied in AI Alpha Lab. Meaning the base AI-model stays the same for years. A new base version is only applied if it generates better results (returns, robustness, or less required computer power) and after it has been thoroughly tested on all products in AI Alpha Lab with satisfactory results. This is a huge task involving a lot of testing on historical data which in itself limits how often a new base AI-model can be applied. Last time a new base AI-model was applied was in December 2019.

### *5.2. The AI Alpha Lab AI-model*

The AI Alpha Lab AI-model is developed with a quantitative approach and taught to take into account a wide range of investment methodologies and investment concepts such as momentum, portfolio tranching, ensembles and process- and implementation diversification. The current version of the AI-model has been in service since December 2019 and has not been changed in the meantime.

With regards to development of the AI-model the same principles apply as for the base AI-model in section 5.1. The AI Alpha Lab AI-model is only changed if the new version produces better results and has undergone the same testing as described in section 5.1. If the changes were more frequent and did not undergo rigorous testing it could very easily end up just being a data mining exercise.

### *5.3. Product development*

When developing new products, for instance a new portfolio with a universe or strategy not used before in AI Alpha Lab, the process consists of the following parts:

1) Always begin with first principles: What is our hypothesis? We want to test a hypothesis not generate a hypothesis which is a very common problem in finance/investing.
2) What data do we need to test the hypothesis i.e., what data do we believe contain information about our hypothesis?
3) How should we train the AI-model? Training period, prediction period etc.
4) Test the AI-model through rolling out of sample (walk forward test).
5) Test the robustness of all parameter choices that are put on top of model output i.e., cut-off thresholds, smoothing, rebalancing etc.
6) Documentation/logging to secure the AI-model history.

### *5.4. Development criteria*

Under normal circumstances the main criteria for product development are:

1) The AI-model's performance, i.e., the generated return.
2) The robustness of the performance, i.e., the likelihood of continuously generating the performance.
3) The explainability of the performance. Not towards clients but internally in AI Alpha Lab. In other words, does the results make sense given the investment universe, data, training etc.
4) What are the limitations and how does the AI-model react to different market conditions.
5) That the factors influencing the AI-model are highly time-varying which must be taken into account.
6) Data selectivity. Which data types improve the results - and which does not.

## 6. DATA

Through years of development and testing we know that the AI-model is highly dependent on data quality. Data quality and data cleaning is always an integrated part of any development or training of the AI-model.

AI Alpha Lab acquires all data from a third party. Before the AI-model is trained on any data, the data goes through the following process:

1) First, we prefer to work with point in time data meaning data which is not altered at a later point in time like for instance inflation figures.
2) We compare data from several independent data providers on a sample basis.
3) Data quality checks are automatically run by an in-house application to detect errors and outliers.
4) A manual visualisation of data (graphs or charts) is being performed to visually check for data hiccups.

When the AI-model is live (as opposed to being tested) it is normally trained on rolling one-year data. That is after one month, one month's new data is taken in and the oldest month's data is deleted. Based on our testing, the conclusion is that there are simply not good enough patterns in data from financial assets allowing the use of for example what happened 10 years ago for something significant.

It is therefore better to train the AI-model on "fresh" data, as this is more predictive of what will happen in the near future (1-3 months). In that area, AI in the financial area differs greatly from other areas such as image recognition, where it can be advantageous to give a model substantial amounts of data to train on, because the problem you are trying to solve is more "stationary".

However, when developing and testing it is crucial to have access to older data to see how the AI-model reacts to different market conditions. As a result, when putting together the investment universe for a specific product, a requirement for each asset in the universe is access to a certain period of historical data.

In general, it is our experience that when it comes to data types, we must be selective. Giving the AI-model access to further data types does not automatically yield better results. In addition, with an increased amount of data, the need for computing power also increases, i.e., it takes longer to train the model.

## 7. TRAINING

The overarching purpose of the AI-model is to generate stable/robust returns on financial assets. That is the main purpose of any training of the AI-model. Therefore, most choices when training the AI-model is driven by this factor.

Naturally, with better potential investment returns the investment risks also increase. Mitigation of investment risks are to a large extend baked into the AI-model, as it will estimate both the return of the asset in question and the probability (or uncertainty) of the return estimate. Meaning it is possible to adjust the risk of for instance a portfolio generated by the AI-model to a client’s liking.

With regards to data quality please see section 7 above. Again, it is worth highlighting that when it comes to data types, careful selection is the key word. Based on our experience we know that giving the AI-model access to further data types does not automatically yield a better result.

With regards to over- or underfitting our sampling techniques not only ensures that we fit causality in data, not correlations, but it also provides a natural haircut through a Bayesian Occam's razor procedure to ensure that we do not overfit. Like Einstein said: "*Everything should be made as simple as possible, but not simpler*. Meaning that if two models provide the same explanatory power, but the models defer in the number of parameters used, we automatically select the simplest model.

Also, by not forcing our AI-model to provide an answer, we greatly minimize the risk of overfitting. We see the transition from an optimization paradigm to a sampling paradigm as essential to safe, robust AI. Optimization is about finding the best model, whereas sampling is about finding the true model. The best model rarely exists, and we have to take that into account when using AI.

## 8. ROBUSTNESS AND EXPLAINABILITY

The robustness of the AI-model has been touched upon in several other sections of this memo, but it makes sense to summarize the thoughts behind the AI-model from a robustness perspective. We have ensured the robustness of the AI-model the following way:

1) By having a well-defined hypothesis. We think there is information to extract from prices on assets (momentum).
2) By using the best suited AI-model. We believe that is a probabilistic model, because it estimates both the future return of the asset in question and the probability (or uncertainty) of the return estimate.
3) By testing on several assets. We can see the AI-model works on assets where momentum is present.
4) By testing over time: We can see the AI-model works over large periods of time, not only a couple of years.
5) As a main rule, we do not change the AI-model.
6) By ensembling. When testing back in time we are always willing to give up performance for robustness.

This is also how we secure that the AI-model is not biased.

The explainability of the AI-model to some extend goes hand in hand with robustness. The AI-model must produce robust results, otherwise it will be hard to explain what the AI-model is doing.

However, explainability is also a two-edged sword. On one side we must be able to explain what the AI-model is doing - to some extent. On the other side giving a full explanation for all the AI-model's choices is simply not possible in practice because of the millions and millions of calculations being made.