



**Young Potential ALIS Managers**

Source: <http://www.texasenterprise.utexas.edu/2015/05/13/workplace/how-strangers-unlock-our-creativity>

## Data De Groove

**Michael Weinberg**  
MOV37

There is a new generation of investment managers which will disrupt asset management. These managers – which engage in Autonomous Learning Investment Strategies (ALIS) – use unstructured non-financial data, machine learning, and record-low computer processing and storage costs to run innovative investment strategies at lower costs vs. traditional fundamental and quantitative managers. ALIS managers typically don't originate from the traditional finance world of Wall Street and MBAs, but rather they are run by PhDs and have their roots in the counter-culture arena which include gamers and hackers (though benign ones).

It is therefore fitting that we've drawn the title of this paper from a Falco album, *Data De Groove*. The album was released nearly three decades ago, in 1990, and was dedicated to the computer era. At that time society was still in early innings in the 'computer era', and therefore we would state that Falco was decades ahead of its time, at least with *Data De Groove*. (Though the author notes that he personally was an early adopter of technology, having had an original IBM 8086 PC dual floppy disk drive PC with a monochromatic monitor, nearly a decade earlier.)

We have met more than 200 ALIS managers around the world over the past few years as well as other thought leading asset owners and investors, and one of the main things we speak to them about is data. In our view it's one of the most misunderstood ALIS topics. We will attempt to answer some important questions about data usage by ALIS managers, and hopefully clarify a few common misperceptions.



### **How do ALIS Managers Differ From Established Quantitative Managers?**

We believe ALIS represents the proverbial “third wave” of investment management. The first wave was fundamental discretionary managers, which was disrupted by the second wave – quantitative managers. Autonomous Learning Investment Strategies – the third wave – are new managers forming due to the confluence of five unprecedented events.

First, data blew up. There are massive and exponentially increasing amounts of new data being created – much of which is untapped by investors. Second, data science caught up. New analysis and structuring platforms are arising to make all this new data usable. Third, machine learning is surpassing humans. As opposed to the old model of humans programming computers, computers are now learning from experience and “teaching” themselves at a lightening fast pace. Fourth, the cost of computer processing and storage has collapsed. Prohibitively expensive server rooms are now replaced with cheap cloud computing. And fifth, discretionary management is being left behind, as their information edge has been eroded by regulatory enforcement.

As a result of these five events, small teams that combine the power of human + machine can now effectively compete with discretionary managers as well as the large established quants. Quant strategies that used to take an army of PhDs and massive investments in servers can now be run by a couple of PhDs and the cloud, at a much lower cost.

### **Data is Expensive - Don't Managers Need Large AUM to be Able to Afford Expensive Data Sets?**

We couldn't agree more. Some data is expensive. Credit card data is an example that falls in this category. We believe that almost all large hedge funds subscribe to this data. Similarly, the providers of this data typically are quite prevalent due to large marketing budgets to tout their offerings, which managers pay for through substantial subscription fees, resulting in a virtuous cycle. For precisely these reasons, we believe the value of some of this ‘off-the-shelf’ data has likely diminished to the point that it is commoditized and the residual alpha from this (and similar) sources has in many cases decayed to the point of being immaterial.

### **With These Expensive Data Sets – How Much Value is in Them?**

Our sense is there may be some value in them, but not in and of itself. Marcos Lopez del Prado, a world leading machine learning expert, analogizes that a good way to find alpha sources is akin to using a sieve to strain gold out of substrate rather than searching for nuggets. In other words, the aggregation of many smaller particles will be greater than that found in nuggets. We believe this analog may hold with some of these more expensive and potentially commoditized data sets – that in conjunction with other data sets, there may be some residual value to them. Another analog that we would use is that it is akin to combining orthogonal low(er) Sharpe strategies that result in a higher Sharpe strategy.

	9	CUST_ID	CUST_CAT	MERCH_ID	MERCH_CAT	MERCH_SUBCAT	FRAUD	GROUP
1	■	Customer001	65andOver	Merchant0001	Retail	DrugStores	0	1
2	■	Customer001	65andOver	Merchant0002	Retail	FoodStores	0	1
3	■	Customer001	65andOver	Merchant0003	Services	Restaurants	1	1
4	■	Customer001	65andOver	Merchant0004	Services	Restaurants	0	1
5	■	Customer001	65andOver	Merchant0005	Services	OtherServices	0	1
6	■	Customer001	65andOver	Merchant0006	Services	OtherServices	0	1
7	■	Customer001	65andOver	Merchant0007	Retail	General	1	1
8	■	Customer002	35to44	Merchant0008	Services	OtherServices	0	1
9	■	Customer002	35to44	Merchant0009	Retail	GasStation	1	1
10	■	Customer002	35to44	Merchant0010	Retail	OtherRetail	0	1
11	■	Customer002	35to44	Merchant0011	Services	OtherServices	0	1
12	■	Customer002	35to44	Merchant0012	Retail	NonStore	1	1
13	■	Customer002	35to44	Merchant0013	Services	OtherServices	0	1
14	■	Customer002	35to44	Merchant0014	Retail	FoodStores	0	1

**Credit Card Data:** Credit card data is an example of an expensive data set that is marketed to investment firms, so much so that alpha from the 'off-the-shelf data' may have diminished since the data is becoming commoditized.

Source: <http://morgan.dartmouth.edu/Docs/sas92/support.sas.com/documentation/cdl/en/grnvwug/61307/HT ML/default/n0n6fnvllk0wwun1t38tds6cpl2y.htm>

	<u>Trading Days</u>	<u>NYSE Group Block Share Volume (Millions)</u>	<u>NYSE Group Block Trades (Thousands)</u>	<u>NYSE Group Block Dollar Volume (Billions)</u>
January	20	5954	103.2	\$224.80
February	19	5566.9	102.1	\$212.60
March	23	7799	126.1	\$304.90
April	19	5599.3	105.1	\$202.50
May	22	6843.2	131.4	\$244.10
June	22	9212.7	132.7	\$326.20
July	20	6058.9	100.9	\$233.70
August	23	6164.8	118.3	\$237.50
September	20	7418	109.5	\$296.60
October	22	6322.4	116.5	\$249.90
November	21	6493.5	115.1	\$260.00
December	20	7618	101.9	\$333.50

**NYSE Group Block Volume in NYSE Listed, 2017:** Structured financial data often can be obtained for free online, like data available on the NYSE's website.

Source: <https://www.nyse.com/data/transactions-statistics-data-library>



## How do ALIS Managers Procure Data Differently Than Large, Established Managers?

The short answer, in one word - Creatively!

The long answer, in more than one word....

As we previously stated, ALIS managers often originate from the counter-culture, including gamers and hackers. There are successful managers who see no need to subscribe to Bloomberg, Reuters, FactSet or Capital IQ. They procure their technical structured financial data gratis from open sources and avoid paying for the aforementioned data services.

These managers subscribe to the philosophy that data is like air. It is no less available than air is when one exists and should be the same price, free. They employ complex web scraping techniques to systematically obtain data from the internet to generate alphas. Sources range from search engines, to social media, to news to other third-party websites. These managers then systematically use Convolutional and Recurring Neural Networks, CNN and RNNs, to analyze and process the data, which is in turn used in its models to make investment (or divestment) decisions.

ALIS managers also may train their neural nets to work better than 'off the shelf' products or approaches used by more traditional quantitative or computational finance funds. For example, sentiment analysis of text is quite common, and also possibly commoditized. However, we believe that these managers have a better mouse trap that is effectively more nuanced in ascertaining the meaning of words, phrases and sentences, resulting in material and accretive alpha, where competitors may only find commoditization.

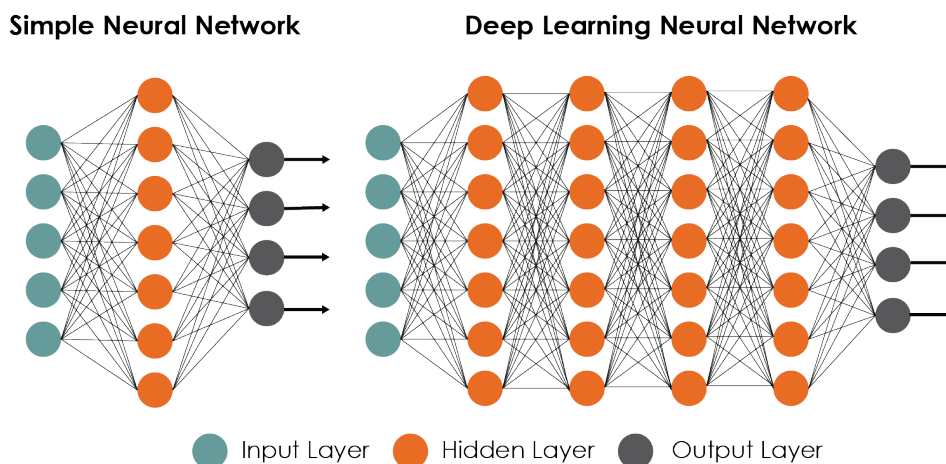
This is not to say that ALIS managers don't pay for data; some better ones certainly do. And what they have found is that the data market, like the security markets, is similarly inefficient. There are the expensive, well marketed data sets, as described above, that typically provide low returns and/or alpha per cost, and there are under or un-marketed data sets that may provide high returns and/or alpha per cost. These valuable, alpha and return rich data sets often are procured in a one-off way as they are inherently under-marketed.

## How do ALIS Managers Approach Charging Data Costs to Their Investors?

We've found that most ALIS managers, having come from outside the Wall Street world of MBAs, view fees differently. ALIS managers believe fees should be reasonable and that high fees should only be earned when there is high performance. This differs from some established managers who earn high fees on large asset bases, even during years of mediocre performance.

One place we see a difference is in the much lower prevalence of "pass-through" expense structures among ALIS managers. These pass-through expenses for data and related costs are relatively common among larger quantitative managers and can range from 0.5% to multiple percentage points and are on top of fees of 2 and 20 or 3 and 30 or even higher.

Another difference between ALIS managers and established quants is ALIS managers are more open to fee structures that incentivize performance over asset gathering. We published an article in *Pensions & Investments* called "Hedge fund fees – a perfect solution", describing a 1/10/20 fee structure that we created and some ALIS managers have adopted. Investors pay a fixed 1% management fee, a 10% incentive fee for net returns below 10%, and a 20% incentive fee for net returns above 10%. Investor interests are better aligned



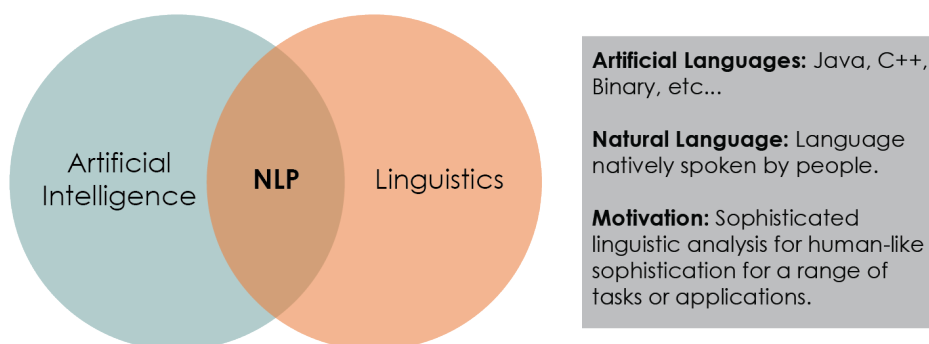
Neural networks, used to analyze unstructured data such as text scraped from websites, can vary in effectiveness, meaning that data which may be commoditized to a manager using a simple off-the-shelf Neural Network may be valuable to a manager using a more complicated Neural Network that may uncover more relationships.

because there are more incentives to produce high returns (and earn a 20% incentive fee), and less of an incentive to stick investors with large pass-through expenses and management fees. ALIS managers can offer these investor friendly fee structures because they are not tied to expensive legacy systems and processes and can tap cheap computing power and machine learning to run their strategies with fewer staff.

In summary, ALIS managers who typically are small, emerging and often from the counter- culture, can be more creative in data acquisition and spending for multiple reasons. These managers don't feel compelled to use traditional Wall Street vendors, and may philosophically believe that content should be gratis. They also may have fee structures that incentivize them to minimize data and other expenditures.

Concomitant with what we believe is a vastly inefficient data market, where expensive data can be commoditized, while off-the-run, free, web-scraped or inexpensive data may be very valuable. We believe Falco nailed it nearly three decades ago, with the prescience of naming his album *Data de Groove*.

## What is Natural Language Processing?



**Goal:** Have computers understand natural language in order to perform useful tasks.

### Disclaimer

The views expressed are those of the author at the time of writing and are subject to change. The author is a member of MOV37, a registered investment adviser. The information in this document has been obtained or derived from sources believed by the author to be reliable, but neither the author nor MOV37 represents that this information is accurate or complete. This material has been distributed for educational/informational purposes only, and should not be considered as investment advice or a recommendation for any particular security, strategy or investment product. Past performance is not a guarantee of future returns. As with any investment vehicle, there is a potential for profit as well as the possibility of loss.

### Author Bio



**Michael Weinberg**  
MOV37

For 25 years Michael has invested directly at the security level and indirectly as an asset allocator in traditional and alternative asset classes. He is the Chief Investment Officer at MOV37 and Protege Partners, and on the investment, management and risk committees. MOV37 invests in ALIS, systematic strategies that deploy machine learning/artificial intelligence and data science. Michael is also an Adjunct Professor of Economics and Finance at Columbia Business School, where he teaches Institutional Investing, an advanced MBA course that he created.

He also testifies as an expert witness in financial and technology litigation. He was a portfolio manager and global head of equities at FRM, a multi-strategy investment solutions provider. Prior to that, Michael was a portfolio manager at Soros, the macro fund and family office, and at Credit Suisse. Before that he was a real estate analyst at Dean Witter.

Michael is a board member of AIMA and on its Research Committee. He is on the management advisory council for the Michael Price Student Investment Fund and an advisory board member for the NYU Stern Investment Management and Research Society. Michael is a founder and advisory board member of YJP CIO, a young professional organization. He is a member of The Economic Club of New York. Michael is a former Chair at CFANY, where he has received multiple awards, including Volunteer of the Year. He is a research contributor to The World Economic Forum on the impact of AI on Finance.

Michael is a published author, having written for *The New York Times*, *Institutional Investor*, and investment books. He has been interviewed by the *Wall Street Journal*, *Financial Times*, CNBC, Bloomberg and Reuters. Michael is a frequent panelist, moderator and lecturer for investment banks, institutional and family office organizations and business schools, including SALT and Harvard. Michael has a BS from New York University and an MBA from Columbia Business School.