NLP in Financial Services

Summary of Research Findings

LSEG LABS

Introduction

Geoff Horrell

Global Head of LSEG Labs

Unlike advances in time series or quantitative analysis, leaps forward in understanding human language have far wider societal and commercial impacts. Consider the use of translation tools, or chatbots or voice assistants in your phone.

In the last few years major advances in natural language processing have been achieved due to increases in processing power, data availability, open source and new techniques. These are already being used across the financial world for sentiment analysis, recommendation systems and many other use cases.

Within the Labs we were curious. Having worked with NLP for many years, we wanted to take stock and see how our customers and the market had evolved since the ‘big bang’ emergence of advanced language models a few years ago.

This report shows that tools are maturing, technology has evolved and data science skills are more widely available. The limit is now the vision, creativity and ability to execute in the new age of machine learning.

> Natural language processing turns the complexities of human language into simple, systematic, powerful patterns that power workflows and analytics.
>
> **– Geoff Horrell, Global Head of LSEG Labs**

A Note from the Authors

Daniel Lewington
Director, Product Design
LSEG Labs

Laura Sartenaer
Strategy and Partnerships Manager
LSEG Labs

This document comprises the output from interviews with LSEG subject matter experts, market research and interviews with our customers – some of the largest financial services companies in the world who are working to advance NLP deployment. LSEG is a business built on open access and so we wanted to share this report with you.

Presented as research findings rather than concrete conclusions, we hope it will provide valuable points of comparison for your own NLP conversations and strategies. For the team in Labs, it has helped us shape our own internal conversation and those we hold with our customers and ultimately made us better equipped to provide them with the data, tools and support that they need.

If you have any thoughts or feedback you would like to share, please do contact us at **refinitivlabs@refinitiv.com**.

Finally, if this report has been of interest, you may also be interested in our 2020 Artificial Intelligence & Machine Learning Survey.

LSEG Labs were previously called Refinitiv Labs. We changed our name after LSEG completed the acquisition of Refinitiv and we began partnering with the Data & Analytics, Capital Markets and Post Trade divisions to help inform and accelerate the creation of new customer experiences and products.

Contents

Methodology and Concepts

LSEG
LABS

Research Methodology

This document contains the key findings
from each of these research workstreams

01

LSEG NLP experts

02

Market research

03

Customer interviews

Core NLP Concepts

Definition and background

NLP is a branch of artificial intelligence that is used to help machines understand the structure and meaning of human language by analysing various aspects like syntax, morphology, semantics and pragmatics.

Dedicated to the harnessing of human language in programmatic ways, using linguistics, computer science and machine learning, it converts unstructured data, the written or spoken word, into structured data. This information can then be interpreted and acted upon by machines.

There are an enormous number of NLP use cases in financial services given its ability to analyse vast amounts of unstructured data available, unlocking new sources of actionable insights and driving operational efficiencies.

Models

Language models power NLP applications. They learn to predict the probability of a sequence of words and are a crucial first step for most NLP tasks. Language models are rapidly evolving and improving the capabilities of NLP.

Deep Recurrent Neural Network (RNN) Language Models

- Gated recurrent units (GRUs) and long short-term memory networks (LSTMs)
- Bidirectional RNNs
- Attention mechanism and memory-based networks
- ELMo

Transformer-based Generative Language Models

- Google BERT
- Open AI GPT 2&3
- XLNet
- Baidu ERNIE

Huge advances over the past 2 to 3 years

Tasks, techniques and processes

NLP can be applied to many different tasks. Most advanced applications require these tasks to be combined to generate desired outcomes. The relative importance of each depends on the use case or application.

Examples include:

- Named entity recognition
- Syntax and morphological analysis
- Word disambiguation
- Sentiment analysis
- Information extraction
- Word embeddings
- Machine translation
- Intelligent tagging
- Entity resolution
- Topic modelling
- Clustering
- Text-to-speech
- Speech-to-text
- Conference resolution
- Semantic analysis
- Relation extraction

The NLP Market Landscape

LSEG
LABS

NLP Ingredients

"NLP matters so much. We don't communicate in numbers."

– Global Head of Trading Analytics at LSEG

Unstructured data

- Information is encoded in language and ~80-90% of all data is unstructured, by most estimates.
- Financial services have a vast amount of sources to comb through, such as news, research reports, company filings, transcripts of quarterly earnings calls, social media and Internet sites.
- Most financial analysis over the last several decades has been on structured, numerical data and it is increasingly hard to generate differentiated value.

Compute

- Exponential increase in computational power with the ability to handle an unprecedented volume of data has led to development of highly sophisticated deep learning neural networks.
- High-performance compute hardware is enabling large-scale deep learning. Recent M&A developments and the strong focus on AI chips from big tech companies is further accelerating this trend.

Open-source

- Open source technology has created a collaborative model that has contributed to the high growth of NLP usage.
- Open source model and data set sharing is driving NLP's Cambrian explosion.

Technology advancements

- It is widely accepted that NLP performance is far better than it was 5-10 years ago, given advances in language models and computational power.

High investment

- In the 'race for AI', tech giants and VC firms have poured vast amounts of money into NLP companies, which has led to rapid technology development.
- It is estimated that >$50b was invested in AI start-ups between 2011 and mid-2018[1].

Customer experience

- The last two decades have seen widespread recognition that a company cannot be successful without having a clear understanding of its customer's wants and/or needs.
- NLP enables organisations to better understand their customers and tailor the customer experience.

Democratisation

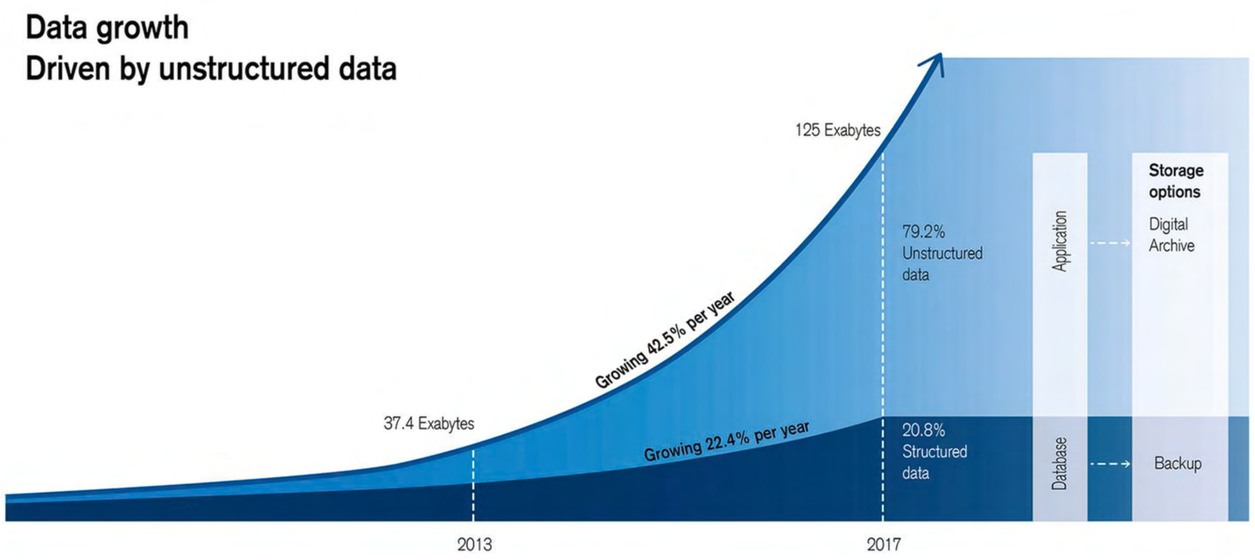
- NLP has become an integral part of our day-to-day lives and we no longer need to be convinced of its potential.
- Predictive typing, spell checkers and virtual assistants (e.g., Alexa, Siri) are all examples of how NLP surrounds us.

1 OECD – https://www.oecd.org/going-digital/ai/private-equity-investment-in-artificial-intelligence.pdf

Data

According to multiple estimates, 80-90% of the world's data is unstructured.

Differential value in structured data is being squeezed out. The use of unstructured data is growing rapidly.



125 Exabytes of enterprise data was stored in 2017; 80% was unstructured data.
SOURCE: Oracle and IDC

Investors must extend beyond the evaluation of traditional data sets to systematically track segments of the market by including social media, news and blogs in their market monitoring strategy.

IDEAS MADE TO MATTER | ANALYTICS

Tapping the power of unstructured data

by Tom Herbert | Feb 1, 2021

https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data

GameStop: Why Informed Investment Decisions In A Digital Era Matter

11 February 2021 Marina Goche

https://www.alpha-week.com/gamestop-why-informed-investment-decisions-digital-era-matter

OPINION

Hidden in Plain Sight – The key to finding insights in unstructured data

https://www.cio.com/article/3604475/hidden-in-plain-sight-the-key-to-finding-insights-in-unstructured-data.html

Artificial intelligence ✓ Added

Asset management's fight for 'alternative data' analysts heats up

https://www.ft.com/content/2f454550-02c8-11e8-9650-9c0ad2d7c5b5

BUSINESS SCHOOL RESEARCH

How Twitter can help institutional investors make better trading decisions

https://www.theglobeandmail.com/business/careers/business-education/article-how-twitter-can-help-institutional-investors-make-better-trading/

Robo-surveillance shifts tone of CEO earnings calls

Trading algorithms leave a mark with deeper focus on the spoken word

https://www.ft.com/content/ca086139-8a0f-4d36-a39d-409339227832

As payments giant Wirecard enters the pantheon of shocking corporate failures, here are 12 great scandals that rocked the business world...

By THIS IS MONEY
PUBLISHED: 10:51, 26 June 2020 | UPDATED: 11:32, 26 June 2020

https://www.thisismoney.co.uk/money/markets/article-8460299/Here-12-fraud-scandals-rocked-business-world.html

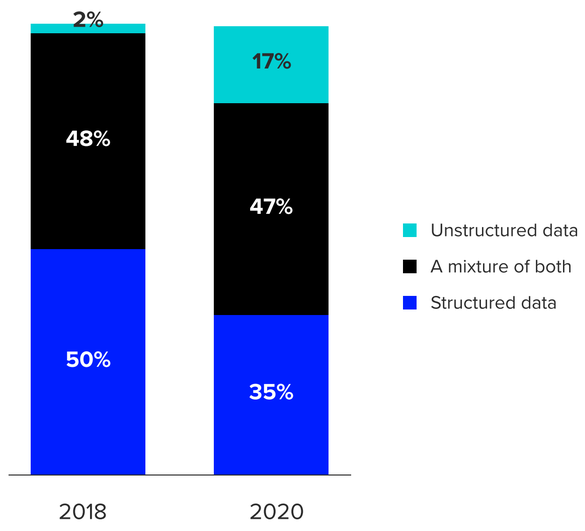
Investment

The NLP market was valued at ~£8b in 2019 and is expected to grow to ~£27b by 2025, with a CAGR of 21.5%[1].

The use of unstructured data is increasing year-on-year.

- Number of firms using only **unstructured data** with their ML models has grown from **2% to 17%** in the last year.
- **62%** of respondents use **News** in their ML models making it our most-used data set.

Structured vs. unstructured usage

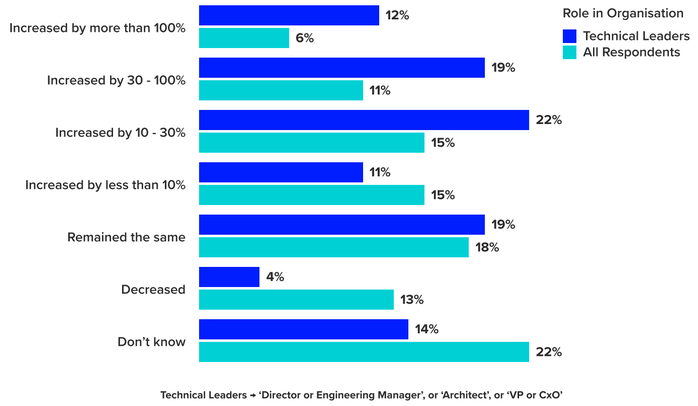


SOURCE: Refinitiv AI/ML Survey, December 2018; June 2020

Budgets increased specifically for NLP across industries in 2020.

- **64%** of technical leaders had **larger NLP budget**.
- **47%** of all respondents **had larger NLP budget**.

Compared to 2019, the budget allocated to NLP projects in your organisation has:



2020 Budget for NLP projects. Each bar reflects the share of respondents within the given group.

SOURCE: Gradientflow.com

NLP practitioners are a growing and increasingly important customer segment.

Artificial intelligence and machine learning skills are in constant demand. LinkedIn reported a growth of **40%** in global hires in 2020.

Within AI, natural language processing stands out as an area of talent growth.

According to a UK-based recruitment survey, NLP talent mainly sits within London, with 6,606 professionals in this space of AI. However, there are 638 professionals in Edinburgh and 626 in Manchester, both growing tech hubs for machine learning, and particularly NLP.

Amazon, Facebook and Google are the top three companies employing this talent.

The University of Edinburgh, University College London and the University of Cambridge scored as the top universities producing this talent in the UK.

~14,000

NLP practitioners in financial services

SOURCE: Understanding Recruitment AI/ML Talent survey; LinkedIn

1 ReportLinker – https://www.reportlinker.com/p05877815/?utm_source=PRN

Vendor Landscape

Segment	Insight	Detail	Example Provider			
Use Case Providers	A crowded fintech segment with providers focussing on one use case, reflecting the proliferation of NLP use cases and the fact that one NLP model does not fit all. Models must be built, governed, and maintained for specific use cases.	NLP applications in: **Corporate operations** • Chatbots/customer service • Data reconciliation • Document processing • Financial reporting • Fraud and risk **Investing and trading** • Data reconciliation • Opportunity scanning • Market forecasting • Sentiment analysis • Trend spotting	**Accern**	**Agolo**	**AlphaSense**	**Arkera**
			Beautiful Soup	**Cleo**	**Clinc**	**Codeq**
			Dataminr	**DataVisor**	**Event Registry**	**FeedStock**
			hyScore	**Kensho**	**Reorg**	**SAS**
			Sentieo	**Signal**	**Sysomos**	**Text2Data**
			Uniphore			
Open Source	Open source technology now powers much of the digital economy and has created a collaborative model that has contributed to the high growth of NLP usage.	• The open source community has developed a vast array of tools that can be used to implement NLP more effectively. • It is so robust you can easily build 'on the shoulders of giants' using just a small, highly focussed team and a platform approach.	**Allen NLP**	**Google BERT**	**Hugging Face**	**John Snow Labs**
			OpenNLP	**Python NLTK**	**PyTorch NLP**	**spaCy**
			Stanford NLP	**TensorFlow**		
Big Tech & Compute	Big Tech companies have been investing heavily in their NLP capabilities. They both open source their capabilities and leverage open source technology. Much is built for a horizontal use case.	Big Tech companies focus on NLP both as a capability and baked into solutions: • Google: Natural Language – mix • AWS: Amazon Comprehend – baked-in solution • IBM: Watson Natural Language Understanding – mix • Microsoft: Azure AI – do-it-yourself	**Amazon Comprehend**	**Google Cloud**	**IBM**	**Microsoft**
Data Providers	There is still a huge volume of untapped unstructured data, such as news, documents, earnings calls, analytics, etc. Data providers are a vital part of the market as the NLP story is a data story.	NLP is being applied to improve data acquisition and transformation work leveraging typical NLP tasks: theme extraction/topic modelling, named entity recognition, tracking changes over time, identifying shifts in language, trends and tone. Differentiation lines can be drawn around maturity of NLP in these businesses around data breadth, depth and quality.	**Bloomberg**	**FactSet**	**IHS Markit**	**Morningstar**
			Refinitiv	**S&P Global**		

Source: BusinessWire, IDC, The AI and Data Science in Trading Digital Day, Refinitiv Market Intelligence

How Financial Services are Using NLP

LSEG
LABS

Workflow

Most of the workflow is enabled by open source technology

Workflow	Text data acquisition and data engineering	Text pre-processing	Data labelling	Text representation & language modelling	Task-specific modelling and evaluations	Deployment
Tasks and techniques	ETL (Extract, Transfrom and Load), Text wrangling and formatting	Text segmentation, Tokenization, lemmatisation	Annotation	Word embedding, state-of-art language models (ELMo, BERT, XLNet, OpenAI GPT etc.)	**Example tasks:** sentiment analysis, name entity recogition, relation extraction, question answering, auto-summarisation **Processes:** model fine-tuning, model selection, model evaluation	Model production
Example vendors and tools	**Data Provider:** Refinitiv, Bloomberg, S&P Global, FactSet Twitter, Financial Times, online news sources **Database:** Elasticsearch, AWS Athena, Google BigQuery	Python NLTK, spaCy, Stanford NLP	AWS Groundtruth, Figure8, Prodigy, Snorkel, Tag Tog	**Framework:** Google BERT, OpenAI, Hugging Face **Computing:** Nvidia, Google Cloud Platform (GCP), AWS, Microsoft Azure	TensorFlow, PyTorch	AWS SageMaker, Kubeflow, TensorFlow Serving, Cortex
Output	Digestible and searchable text data or database	Clean and normalised text ready for the task	Labelled training data	Embeddings, pre-trained language models	Predictions from fine-tuned models, interpretable evaluation metrics	Model deployed for future usage
Trends	• News is a major source of textual data • SEC filings, earnings calls and social media are all frequently leveraged	• Mostly open source NLP capabilities/libraries, which has largely destroyed the value of software packages • Technological edge comes from fine-tuning parameters for specific tasks and agile management of pipelines • Big Tech firms have pivoted from 'just' providing compute to open sourcing NLP capabilities as well				• Must be transparent. Cannot be black-box solutions or output only

SOURCE: Towards Data Science

Use Cases

01

Internal efficiency

Sifting through messy textual data to extract relevance from forms/text, triaging data to route actions efficiently and enhancing customer experience through chatbots.

02

Accelerating workflows

Search and the linking, clustering and personalisation of information delivery are all based on applied NLP capabilities. This has revolutionised the discovery and access to information.

03

Creating signals

Investment and risk signals can be created with NLP techniques and backtesting them by combining with structured data.

Use Cases

Categories of Problems	Buy-side		Sell-side	Risk & Compliance
	Asset Management	Wealth Management		
Increase Efficiency • Maximise value while minimising effort • Extract and classify topics/themes • Reduce noise – identify what matters in a document • Achieve scale	• Earning season preparation • Search and discoverability • Automated market and financial analysis, assembly of reports and recommendations • Signal to noise	• Robo-advisory to increase efficiency: minimal human intervention required • Use of NLP to generate sentiment score on companies/ entities	• Automation of research, M&A analysis and strategy development • Predictive analytics for core business performance management (e.g., early warning asset churn, etc.) • Improving customer service while cutting cost • Understand market news and trends	• Processing and understanding communication between traders/investment managers • KYC checks – sentiment analytics in negative news media • Market surveillance • Management and regulatory reporting • Transaction monitoring • Fraud detection and identity verification (e.g., image recognition/voice biometrics for customer authentication)
Generate Revenue Growth • Find alpha • Derive value • Categorise data to feed into a model • Understand sentiment • Analytics – identify trends, patterns, anomalies, priorities	• Idea generation • Event-based detection and prediction • Sentiment-derived signals • Theme extraction – ESG, Supply chains, Commodities • Algo trading/liquidity discovery • NLP to derive new data • Portfolio optimisation	• Idea generation • Robo-advisory to drive sales: augmentation and automation of investment decisioning • Use of AI to derive new data (e.g., metadata & alternative data)	• Reason to call • Deal origination • Personalised insights to drive sales outreach • Pre-market price setting • Predict market moves • Flagging investment opportunities based on timing – e.g., financial distress, pre-IPO, capital raising	

Key Customer Research Themes

LSEG
LABS

About Customer Research

Inspires and informs intuition through a variety of methods with related intents: to expose patterns underlying the rich reality of people's behaviours and experiences, to explore reactions to probes and prototypes, and to shed light on the unknown through iterative hypothesis and experiment.

SOURCE: Informing Our Intuition: Design Research for Radical Innovation, Jane Fulton Suri, 2008

Method applied and details

- **Goal:** To understand the problem/opportunity space and broad thematic insights
- **Format:** One-hour guided interviews
- **Interviews:** 1 x facilitator, 2 x subject matter experts (strategy and NLP engineering)
- **Analysis:** Team debrief, transcription analysis, affinity mapping for theme detection

Our Customer Research in Numbers

Customer interviews	Minutes of customer insight	Internal interviews
20	1200	30

Customer interview distribution

Buy side	Sell side	Analytics providers	Consultancies
11	05	02	02

- Analysts and portfolio managers
- NLP experts and leaders
- Engineers working in NLP
- Quants and NLP specialists

NLP Maturity

Needs vary according to customer segment, but adoption increases across the board.

Increasing strategic relevance of NLP

Testing the waters

Common attributes

- Small-to mid-market firms, not being applied in a meaningful way
- Piloting solutions – single business unit or across the institution. One or two solutions in production (e.g., chatbots)
- Command-driven as opposed to NLP-driven workflows
- Open to pre-packaged analytic platforms; buy over build approach to NLP

~25%[1]

Percentage of firms in this category

Convinced and investing

Common attributes

- Senior sponsorship and established NLP teams
- Systematism of use case identification, piloting and deployment in BUs and, often, corporate functions
- Considerable work on enablement (data environment, consistent ML pipeline approach, etc.) and adoption

~20%

Percentage of firms in this category

NLP is fundamental

Common attributes

- Most advanced NLP techniques deployed, including 'black-box' deep learning approaches
- Want the raw data – all pre-processing regarded as proprietary and valuable
- Proprietary sensitivity means cautious with respect to cloud
- Buy-side firms that have evolved to become technology firms

~20%

Percentage of firms in this category

1 Assumptions based on desk research and customer interviews. Assumption that ~35% not doing anything in NLP

Common Themes

Community

A broad, sophisticated and active community has evolved around NLP both within companies and within the industry.

Diverse and expanding use cases

Commonality around a core set of techniques but myriad applications, the number of which is growing fast.

Build

Customers are building rather than partnering or buying to gain the knowledge, retain the IP and competitive edge. They build on top of Open Source.

Scale

Major investment in recruitment and embedding of Ph.D. to grad-level expertise. Also investing in centralised NLP pipelines. NLP is no longer a nice-to-have.

Data for models

NLP models are frequently powered by news + transcripts + filings and structured data, as well as internal data.

Model transparency

Models must be ‘explainable’ to fulfil evaluative, auditing, regulative and ethical requirements.

Interview Quotes: Diverse & Expanding Use Cases

Commonality around a core set of techniques but myriad applications, the number of which is growing fast. There is no single opportunity.

So many applications of NLP – risk management, alpha generation, asset allocation, research, market monitoring

HEAD OF DATA AT BUY-SIDE FIRM

We found NLP was additive in our backtesting. Growth and sentiment (fuzzy matching, dialogue, perception) was in our grasp. We were able to take a holistic view and evaluate companies from different angles: customer, CEO, analyst and so on.

EX-QUANT AT BUY-SIDE FIRM

The other opportunity is around thematic investing – a more direct impact of NLP. We looked at whether it could help us launch new products altogether and semi-automate portfolio creation with companies, e.g., biotech, robotics etc.

VP, DATA SCIENCE AT BUY-SIDE FIRM

Use cases include: Risk Management, fundamental stock selection, energy transition, fiduciary reporting.

NLP ENGINEER ASSISTANT MANAGER AT CONSULTING FIRM

There are so many use cases around asking questions, raising queries, obtaining replies, e.g., chatbots, digital banking, IT applications.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

We worked on news analytics – which was all about alpha and volatility prediction, but it evolved to extracting information e.g., for research.

APPLIED AI ML DIRECTOR, ML CENTER OF EXCELLENCE AT SELL-SIDE FIRM

Every two months there are new models trained on new data sets. And you can fine-tune it. It works for simple tasks.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

Corporate restructuring – you could get this out of the box. But what about all-day breakfast at fast food chains. It's not set up out of the box. Not in the ontology. We need to learn evolving topics.

HEAD OF MARKET INTELLIGENCE AT BUY-SIDE FIRM

We're developing M&A and divestment models.

NLP ENGINEER ASSISTANT MANAGER AT CONSULTING FIRM

Interview Quotes: Build

Customers are building rather than partnering or buying to gain the knowledge, retain the IP and competitive edge. They build on top of open source.

No one here is going to accept a third-party NLP solution providing only the output. Ever.

HEAD OF DATA CURATION AT BUY-SIDE FIRM

Do you build, buy or partner?
It's build.

EX-PORTFOLIO MANAGER AT BUY-SIDE FIRM

Vendors are now competing against open source.

HEAD OF SEMANTIC TECHNOLOGY, ANALYTICS AND MACHINE INTELLIGENCE AT SELL-SIDE FIRM

We want to build. We need to keep the IP and knowledge in the team. There is no competitive edge in being vendor-locked.

BUSINESS & TECH ANALYST • DATA SCIENTIST AT SELL-SIDE FIRM

If there is an area where we are one year behind we will then look at vendor products. It's a stop gap. We'll do everything in house eventually.

HEAD OF DATA AT BUY-SIDE FIRM

And as a layman I could write NLP – everything is open source. I didn't have to be a data scientist.

FOUNDER AT CONSULTING FIRM

Do you build, buy or partner?
No. We build our own. More control … we're using pre-processed libraries.

NLP ENGINEER ASSISTANT MANAGER AT CONSULTING FIRM

Data vendors end up building the data set too late. So it's about [us] building the tools to get the answers faster.

EX-QUANT AT BUY-SIDE FIRM

Yes. We're looking at vendor solutions. We have to compare, even if we're building. If it's roughly the same price, then we'll buy because it's a priority.

EXPERT RESEARCHER AT BUY-SIDE FIRM

Investment teams are trying to use NLP to inform investment decisions. Knowledge is kept at team level for trading signals. Lots of IP to it.

EX-QUANT AT BUY-SIDE FIRM

Innovation Labs keep looking at new companies and products. We organise demos, look at their work. Check the use case. If it's good, then we proceed. But it's a mixture.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

Interview Quotes: Scale

Major investment in recruitment and embedding of Ph.D. to grad-level expertise. NLP needs bodies. Centralised pipelines and model re-use. NLP is no longer a nice-to-have.

The goal is to have a single platform for all business lines.

DATA SCIENTIST AT SELL-SIDE FIRM

We need good architecture for ML/NLP core. Common platform for all models. And data centrally.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

We have doctorates in NLP feeding into everything we do.

HEAD OF DATA AT BUY-SIDE FIRM

We have tens to hundreds of people [SMEs] in each business line. More around the hundeds that are not always dedicated but work in the space. Many are at least half capacity on the research and development or operational implementation.

DATA SCIENTIST AT SELL-SIDE FIRM

Every town hall meeting we get to know about status of NLP, how is your deadline going, what are the milestones. We are recruiting so we can keep extracting value from the text.

NLP ENGINEER ASSISTANT MANAGER AT CONSULTING FIRM

They were still hiring when I left. The target number between London and NY was 45 people. These were applied researchers. Few managers but were hands-on except for the head. Around 20+ were focussed on NLP.

EX-APPLIED AI ML DIRECTOR, ML CENTER OF EXCELLENCE AT SELL-SIDE FIRM

We had a research support team trying to understand what is possible with NLP and to build the infrastructure they share with other teams.

EX-QUANT AT BUY-SIDE FIRM

There is a group within the AI Centre of Excellence – Data Governance group who follow up and monitor. It's how the whole group pitches their value. To make sure leadership are aware that the work they are doing is translated to business value.

VP, DATA SCIENCE AT BUY-SIDE FIRM

We have lots of different groups exploring. I'm just in Investment Banking but there are lots of group initiatives. You will see online that Compliance have done a lot of NLP work.

CHIEF DIGITAL INNOVATION OFFICER OF GLOBAL COVERAGE & INVESTMENT AT BUY-SIDE FIRM

Interview Quotes: Data for Models

NLP models are frequently powered by news + transcripts + filings and structured data, as well as internal data.

We need to go past structured data and also get meaning and value from unstructured.

DATA SCIENTIST AT SELL-SIDE FIRM

PDFs are sent to us and they would like to know if something is interesting for them.
Filtering for relevance based on investment mandate.

EXPERT RESEARCHER AT BUY-SIDE FIRM

Is combining structured and unstructured important?
Yes. They don't exist in silos.

VP, DATA SCIENCE AT BUY-SIDE FIRM

We use structured and unstructured text. Internet of things. Uses in article – top-down, macro, fund flows.

HEAD OF DATA CURATION AT BUY-SIDE FIRM

Always asking, what are the humans processing that machines aren't processing yet?

HEAD OF DATA CURATION AT BUY-SIDE FIRM

Lots of unstructured data sources – 10k filings, earnings calls, internal research notes, news.

VP, DATA SCIENCE AT BUY-SIDE FIRM

We're reading any text that might be helpful. Everything is black box. Anything where there may be relationships between entities in text we can act upon. Filings. News. Descriptions. Unique descriptions. Companies. People in companies.

HEAD OF DATA AT BUY-SIDE FIRM

Earnings call transcripts. Annual reports. News article classification. Major ones. Some in pipeline analysing text from LinkedIn, Glassdoor – for the talent team to mine data. … Annual reports. White papers. … Yes. Twitter – we're using external API for that.

NLP ENGINEER ASSISTANT MANAGER AT CONSULTING FIRM

On the investment side, sentiment is a big part. Topic detection. We're seeing it in ESG. Combining sentiment with topic – news articles or other forms of text – all about packaging it up.

EX-QUANT AT BUY-SIDE FIRM

Interview Quotes: Model Transparency

Models must be ‘explainable’ to fulfil evaluative, auditing, regulative and ethical requirements.

Scores? Forget it.

HEAD OF DATA CURATION AT BUY-SIDE FIRM

Simply receiving an output won’t work, as you do not know the process. Makes the need to understand important.

EX-QUANT AT BUY-SIDE FIRM

I’ll never trade on someone else’s score. You buy to be exposed to. To understand what’s happening. Once I understand, then I’ll build my own.

EX-PORTFOLIO MANAGER AT BUY-SIDE FIRM

We use built-in analytics. It’s not the work of the portfolio manager. Transparency is a very big thing there.

EXPERT RESEARCHER AT BUY-SIDE FIRM

Multilingual model is important – that can be an issue with data vendors. It’s more relevant for regulatory projects. We’re not willing to accept analytics from data vendors. But people don’t mind where you get the data from.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

Still see complaints about it being non-transparent, e.g., can’t see the articles. Well, it’s 1 billion. We can’t show or QA every item.

FOUNDER AT DATA VENDOR

The challenge with external vendors can be the transparency.

EX-PORTFOLIO MANAGER AT BUY-SIDE FIRM

Bias. Fairness. Statistical bias. That’s a whole different ball game. … We focus on ‘transparency’. Looking at model features, not just categories.

HEAD OF SEMANTIC TECHNOLOGY, ANALYTICS AND MACHINE INTELLIGENCE AT SELL-SIDE FIRM

Certain things – ads are useless. Text has nothing to do with the content. Dupes. Yes. Store what you can and let the client decide. Don’t just give a score, give us everything that was used. If you removed dupes, how many?

EX-PORTFOLIO MANAGER AT BUY-SIDE FIRM

Interview Quotes: Community

A broad, sophisticated and active community has evolved around NLP in our customer base.

Different forums. Ultimate guidance comes from senior innovation forum – heads of businesses.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

Hugging Face – we're using that.

BUSINESS & TECH ANALYST • DATA SCIENTIST AT SELL-SIDE FIRM

Marketing was a driver as well. Clients want to hear we are leveraging NLP.

EX-QUANT AT BUY-SIDE FIRM

If you want your algorithms picked up by groups like us, then publish them. Nobody wants a score, but they're happy to get it from GitHub or a published paper.

HEAD OF DATA CURATION AT BUY-SIDE FIRM

We have a Community of Practice channel – 500+ people. Not everyone is hands-on.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

We're decentralised. We don't share too much between groups – research IP. We do our own thing.

HEAD OF MARKET INTELLIGENCE RESEARCH AT BUY-SIDE FIRM

There is a group within the AI Centre of Excellence – a Data Governance group who follow up and monitor. It's how the whole group pitches their value. To make sure leadership are aware that the work they are doing is translated to business value.

VP, DATA SCIENCE AT BUY-SIDE FIRM

It's interesting as of today because we keep creativity – working with various editors. We have different values, teams and needs but have a centralised community. We share work and all the different experiences.

CHIEF DIGITAL INNOVATION OFFICER OF GLOBAL COVERAGE & INVESTMENT AT SELL-SIDE FIRM

But there is also a virtual Data Science Community of Practice. So they are also contributing when they have time – one or two hours per day. Combination of us plus virtual team.

ML/NLP ARCHITECT AT SELL-SIDE FIRM

We use a lot of open source. Everything we do is open source. More and more open source than it used to be, e.g., Hugging Face set of libraries.

HEAD OF DATA CURATION AT BUY-SIDE FIRM

Additional Resources

LSEG
LABS

Resources

AI/ML SURVEY

The Rise of the Data Scientist: Machine Learning Models for the Future
https://www.refinitiv.com/en/resources/special-report/refinitiv-2020-artificial-intelligence-machine-learning-global-study

EVENTS

Deep Learning & BERT – How Google's Language Model Could Transform Finance
https://solutions.refinitiv.com/LearnItAllLab-BERT

How to Improve the Accuracy of Your Financial Language Models
https://solutions.refinitiv.com/LearnItAll-BERT2021

NLP for Capital Markets 101
https://solutions.refinitiv.com/LearnItAllLab-NLP

BLOGS

The Three AI/ML Trends to Watch in 2021
https://www.refinitiv.com/perspectives/ai-digitalization/the-three-ai-ml-trends-to-watch-in-2021/

Four Ways to Apply NLP in Financial Services
https://www.refinitiv.com/perspectives/ai-digitalization/four-ways-to-apply-nlp-in-financial-services/

LIVESTREAM

Natural Language Processing Trends 2021
https://www.youtube.com/watch?v=xHK6QiCg9_w

LSEG LABS PROJECTS

Financial Language Modelling
https://www.refinitiv.com/en/labs/projects/financial-language-modelling

SentiMine
https://www.refinitiv.com/en/labs/projects/sentimine

ESG Controversy Prediction
https://www.refinitiv.com/en/labs/projects/esg-controversy-prediction

Global Infrastructure API
https://www.refinitiv.com/en/labs/projects/global-infrastructure-api

DATA

Data Catalogue
https://www.refinitiv.com/en/financial-data

Data and Tools

Data Exploration Tool

www.refinitiv.com/en/labs/projects/data-exploration-tool

Developer Community

developers.refinitiv.com

Intelligent Tagging

permid.org/onecalaisViewer

Reuters News Archive

www.refinitiv.com/en/products/world-news-data

Transcripts Data

www.refinitiv.com/en/financial-data/company-data/events/earnings-transcripts-briefs

Refinitiv® Data Platform APIs

developers.refinitiv.com/en/api-catalog/refinitiv-data-platform/refinitiv-data-platform-apis

Thank you

If you have any thoughts or feedback you would like to share, please do contact us at:

refinitivlabs@refinitiv.com

Finally, if this report has been of interest, you may also be interested in our **2020 Artificial Intelligence & Machine Learning Survey**.

LSEG Labs were previously called Refinitiv Labs. We changed our name after LSEG completed the acquisition of Refinitiv.

LSEG
LABS