Natural Language Processing (NLP)

An AI tool to cash-in the ‘big-text’ in life sciences

Feb 2021

MEHTA PARTNERS

Advisors

Table of Contents

1. Natural Language Processing: Transform unstructured text into organized insights
2. NLP can catalyze a range of applications in pharma and life sciences
3. Why should you consider adopting NLP today?
4. Applications dive
    - Drug discovery
    - Clinical trial insights
    - Competitive intelligence
    - Pharmacovigilance
5. NLP-enabled solutions are already gaining traction among BioPharma fraternity
6. Key considerations while adopting an NLP solution
7. MP Group can catalyze your NLP initiative

"Innovation at the moment is exponential. It is a full-time job just to keep track of the innovation that comes out"

Mark Baillie
Novartis

"Healthcare data is growing in an exponential way and our capability of utilizing it is growing in a linear way, creating a huge gap. In, we see that 80% of data is unstructured and untapped.. If humans were to read and understand full data we'll take ages. That is where NLP comes in."

Enrico Santus,
Bayer Pharmaceuticals

"I see NLP as an intellectual asset which opportunistic leaders would either be quick to harness, or late to harness. Harness they shall."

Lior Gazit,
Memorial Sloan Kettering

"..social media as a source to identify adverse events is very challenging and actually costly to monitor input channel because of the very high volume of data…Until now this process has been highly manual and expensive, as you can imagine, and time consuming. The idea of was to move into a solution which would reduce the necessary manual review workload. We could do it using NLP "

Damir Bucar,
Novartis

"NLP provides data that would take tens or hundreds of times longer with tedious manual work. It enables downstream calculations to provide insight. Some work would not have been done or done comprehensively without it."

Eric Su,
Eli Lilly company

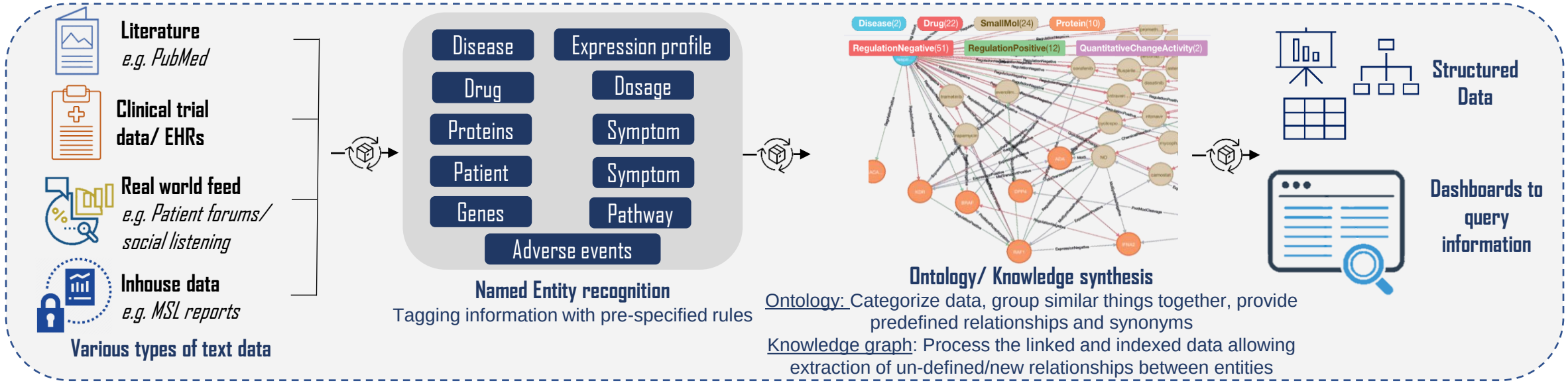
Image Source: Ontotext

Natural Language Processing: an AI solution to Pharma's big text problem

Text analytics powered by NLP can transform unstructured text into organized insights

Pharma has seen an explosion in terms of data. PubMed has more than *30 million citations, clinicaltrial.gov has about 350,000+ studies registered, millions of EHRs*, and numerous other sources like KEGG, patient forums, etc. that are rich source of text data, the surface of which has barely been scratched in terms of extracting insights. **How do we use all this information??**

NLP provides the ability to read or parse human language process and understand large amount of free (unstructured) text into structured data



A protein or target like VEGFA - vascular endothelial growth factor A may also be referred to as MVCD1, VEGF VPF or vascular permeability factor in the literature. An NLP algorithm can be trained to spot all such synonyms and robustly identify all the associated literature, diseases or drugs.

Knowledge graphs: Network of entities with direct and indirect relationships

Establish relationships between drug, proteins, genes, pathways, mechanisms, adverse effects & more to aid your decisions

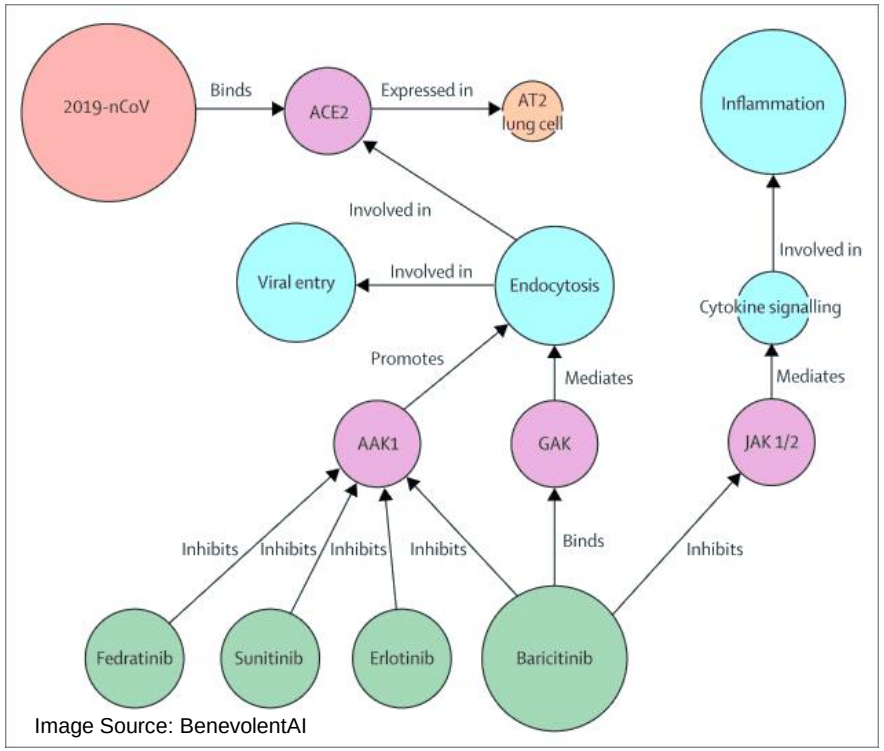
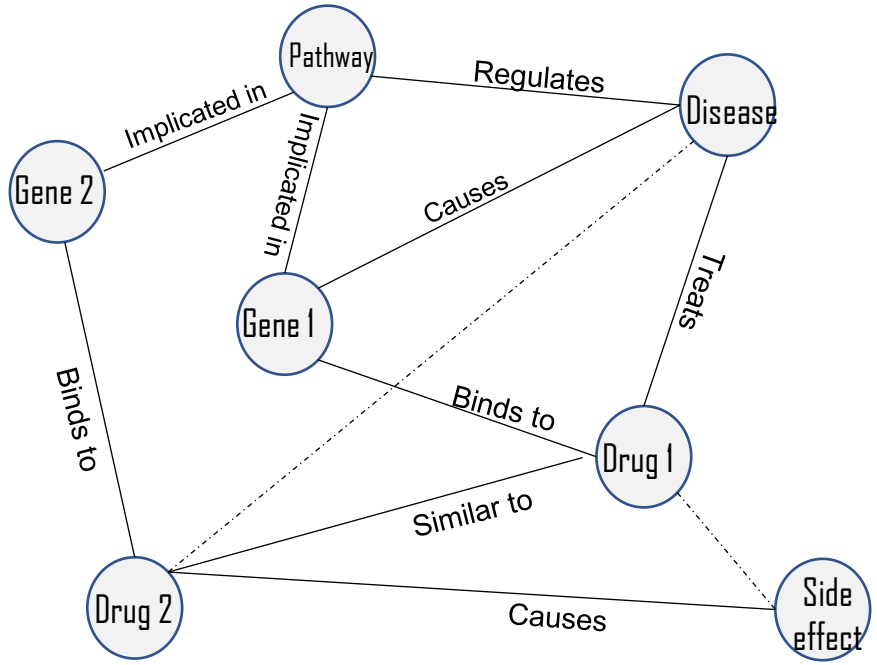


Image Source: BenevolentAI

Example of a drug discovery knowledge graph that led to identification of Barcitinib as a repurposed drug for COVID-19

The network contains relationships between entities and allows intuitive querying of data. Such networks can identify hidden insights from the vast universe of literature. The knowledge graphs can be built for drug discovery, KOL insights, pharmacovigilance, competitive intelligence and more.



Validated relationships

Possible/Inferred relationships

NLP-based text analysis consists of several processes, including information retrieval, information extraction, lexical and semantic analysis, pattern recognition, tagging and annotation (sometimes without a need for knowledge graphs), and data mining techniques, such as association analysis and visualization.

NLP offers a range of applications for the BioPharma industry

Text mining can reduce man-hours needed to analyze documents for drug discovery, regulatory or competitive insights, pharmacovigilance and more

Regulatory insights

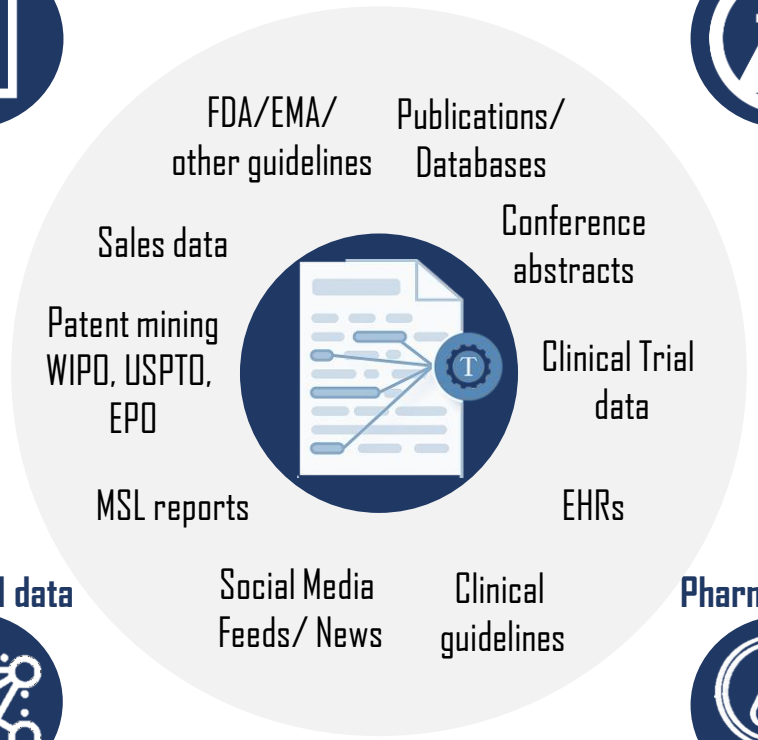
Continuous monitoring of regulatory requirements to provide risk management and surveillance by integrating both external and internal information (*e.g., FDA Warning letters, RTQs,, BLA review reports*)

Drug discovery

Contextualize millions of data points from publications, internal data, clinical trials, etc. to extract causal relationships between drugs, proteins, diseases for target identification, repurposing, quick literature digest etc.

Competitive insights

Automated mining of literature, patents and social media to extract first mention of new targets, drugs or pathways along with the associated organisations



Clinical Trials

Rapidly extract and analyze information such as trial site, selection criteria, study design, patient no. and characteristics from all databases and utilize the information to design clinical trial protocols or data analysis

Real world data

Derive custom, actionable market insights by integrating MSL conversations, medical information requests from HCPs, literature, conference abstracts, news, and social media

Pharmacovigilance

Detect new safety signals by automated ingestion of literature/case reports and complementing with structured data from Adverse Event Reporting systems, such as FAERS, EudraVigilance, and VigiBase

Additionally, NLP powers the chatbots and several other healthcare applications

Why should you consider adopting NLP today?

NLP can significantly reduce time and enhance robustness of extracting insights from big text data

NLP has already revolutionized industries like travel, marketing etc., and is at the tipping point to transform Pharma/life-sciences. Reinforcement learning and intelligent cognitive search using Natural language understanding (NLU), that allow intuitive questioning of data, are driving advancements very rapidly

Real time tracking: Keep a real-time track of newly published literature, news, clinical trials etc., and integrate the knowledge into already existing knowledge graphs to identify key insights and evidence to support your research or strategic decisions

Robust hypothesis support: Connect all disparate data sources to extract information in a common dashboard to support your hypothesis with evidence generated by identifying relationships and assertions that would otherwise remain buried in the mass of textual big data

Reduce Man-hours needed to mine data: NLP allows automated extraction of articles or information for the specific queries and can help create custom read lists while filtering the noise or irrelevant information. This can significantly reduce the time and cost spent in manually scavenging the information though the web

End-to-end intelligent systems: Traditional machine learning algorithms deployed for ADMET/Activity prediction or OMICs data analysis can only read the data in form of tables/sheets etc. Often, immense amount of data is present as unstructured texts (papers, conference, doctor notes) and cannot be handled by ML algorithms. Integrating NLP can extract structured information from the text which can be then used to input/train the ML algorithms

Drug Discovery (1/2)

Deep dive

A biomedical literature search precedes every drug discovery project, from identifying specific genes involved in the therapeutic area of interest to understand pathophysiological mechanism, from identification and validation of a molecular drug target to getting details on any hit/lead molecules, etc.

NLP allows to sift through seemingly endless data, while carefully eliminating biased data sets to extract actionable insights

Academic insights
PubMed, conference abstracts, databases (ChEMBL etc), computational studies, structure similarity, pharmacophore designs, etc.

Clinical insights
Targets in clinical trials studies, safety reports

Commercial insights
Preclinical pipeline drugs, clinical trials pipeline, patents, market shares and competitors analyses

Target identification
E.g.: what are targets involved in pancreatic adenocarcinoma?

Gene/Allele-disease mapping
E.g.: what is the MOA by which the gene or alleles of a gene affect the disease phenotype?

Biomarker discovery
E.g.: what biomarkers can help me with patient stratification for targeted/personalized therapy?

Lead identification/optimization
E.g.: Look at chemical diversity and filter out unwanted compounds a for desired target in all literature quickly

Problem: *Astrazeneca* wanted to create a framework to quickly contextualize literature around a set of targets and create a selection framework to reduce the time spent reading papers

Solution: Astrazeneca adopted applications from multiple providers with different strengths to create a holistic solution.

Leveraging one provider, they built a knowledge graph with over 25 diverse data sources extracting 17 different entities (genes, proteins, diseases, symptoms, drugs, side effects, etc.) to define and categorize explicit relationships between identified entities

Leveraging the other provider, they build a query system to extract key information on targets using a custom DEST framework (druggability, efficacy, safety liability and tractability) and create a selection criteria

An agile system that efficiently resolved uncertainties about the targets and early development of drugs using data-driven testing of hypotheses

Case Study

Drug Discovery (2/2)

Deep dive

A biomedical literature search precedes every drug discovery project, from identifying specific genes involved in the therapeutic area of interest to understand pathophysiological mechanism, from identification and validation of a molecular drug target to getting details on any hit/lead molecules, etc.

NLP allows to sift through seemingly endless data, while carefully eliminating biased data sets to extract actionable insights

Academic insights
PubMed, conference abstracts, databases (ChEMBL etc), computational studies, structure similarity, pharmacophore designs, etc.

Clinical insights
Targets in clinical trials studies, safety reports

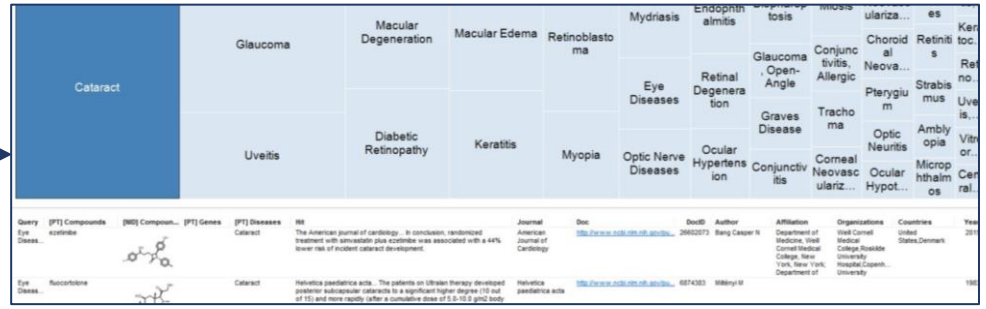
Commercial insights
Preclinical pipeline drugs, clinical trials pipeline, patents, market shares and competitors analysis

Target identification
E.g.: what are targets involved in pancreatic adenocarcinoma?

Gene/Allele-disease mapping
E.g.: what is the MOA by which the gene or alleles of a gene affect the disease phenotype?

Biomarker discovery
E.g.: what biomarkers can help me with patient stratification for targeted/personalized therapy?

Lead identification/optimization
E.g.: Look at chemical diversity and filter out unwanted compounds a for desired target in all literature quickly



Problem: *Roche* medicinal chemists were spending too long searching for key drug-related information buried in the mass of published literature, patents and internal sources
Solution: Roche developed an NLP system augmented with ChemAxon's chemical annotation and name-to-structure tools, to extract and organize compound/target/disease relationships using simple user interface for searching. The solution not only offered a comprehensive literature search but also saved a lot of time saving them a lot of
The solution can be customized to extract patent information, details like KOLs and important centers, and more granular information about diseases and drugs

Case Study

Clinical insights

Deep dive

The strict regulatory environment of clinical trails mandates an efficient documentation therefore, making it a rich source of data. Although some information in trial reports is well structured and searchable using keywords, much of the key information lies in unstructured text. Moreover, managing the such extravagant documentation is a time and cost intensive exercise.

NLP allows to automate entries at several points and to extract useful information from diverse clinical text, such as clinical notes, radiology reports, and pathology reports for several applications.

Patient to Trial | Trial to Site Matching

Data used: ICD-10 codes, EHR, and unstructured clinical data, including doctor's notes, pathology reports, operating notes and other important medical data, Trial sites and databases

Process: Data is processed to extract key elements such as symptoms, diagnoses, treatments, test results, genomics, socio-economics and more while masking protected health information

Outcome: Researchers, Pharma cos and doctors can use these multidimensional patient profiles (or site profiles) to find and compare patients (or sites) suitable for the study

In a pilot study conducted by Mayo Clinic in Rochester, Minnesota, IBM's Watson for Clinical Trial Matching system increased the average monthly enrolment for breast-cancer trials by 80%

Clinical Trial Design

Data used: Journal papers, drug labels, clinical trial data and regulatory guidelines

Process: Data is processed to suggest optimized protocols for faster and efficient trials with insights and recommendations for analytics, Inclusion / exclusion, study design, outcome measurements, statistical analysis, etc.

Outcome: Trial designers can understand how strictness of its eligibility criteria etc., might affect outcomes such as cost, length or participant retention, and in some cases also get an initial draft of the trial design

Eli Lilly has implemented multiple NLP solutions to extract key data and assist the design of trials and reports. They are already witnessing a reduction in cost, time and errors along with improved precision and recall of data extracted

RWE and Clinical trials

Data used: EHRs; patient-reported outcomes such as forums, social media, MSL reports, insurance claims and billing data

Process: Extract the terms related to desired outcomes from unstructured and variable information repositories to create a well-structured data set that is amenable to easy querying and comparison

Outcome: RWE about adverse reactions (ARs) was used to complement or augment the information on ARs reported during clinical trials to the FDA and listed on drug labels

Astrazeneca, BMS, Novo Nordisk, Pfizer and many other pharma companies have either adopted or build internal solutions to mine RWE and augments several applications

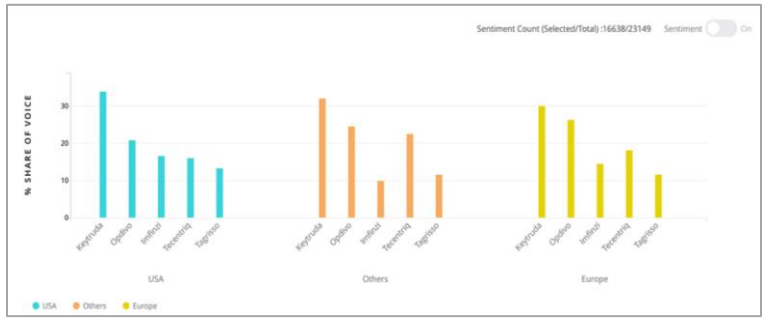
Competitive and Commercial insights

Deep dive

Company executives usually crawl the web everyday to understand opinion and engagement towards the brand, scan the competitive landscape to identify opportunities and drive strategic decisions.

NLP automates real-time monitoring of market insights and create structured dashboards to drive strategic decisions. It can be used for a variety of applications like market landscape including fragmentation or market share, competitor insights like patent fillings or deals, sentiment analysis around drugs in patient forums or congresses/conferences and more.

Sentiment analysis

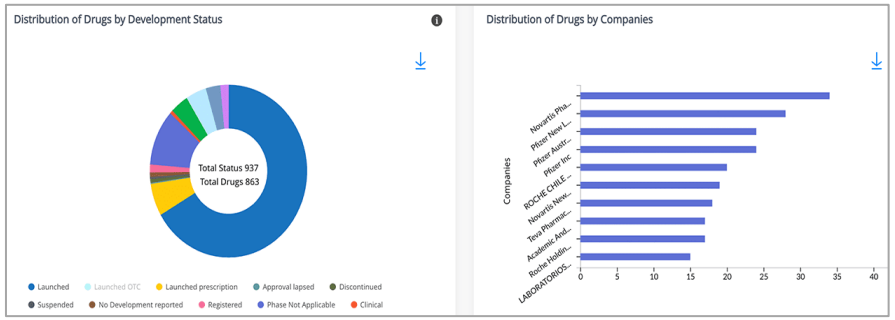


Extract insights about patient sentiments across social media and online forums by continuously monitoring the opinions and engagements around the brand.

Analyze the competitive landscape to identify opportunities to drive positive engagement

Use the same analytics tool to monitor post product launch success

Market insights

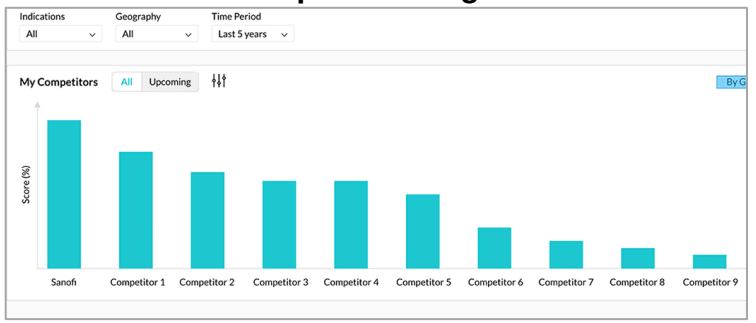


Compare market shares and track changes of competitors by continuous tracking of upcoming launches and performance of marketed drugs.

Identify trends like early adopters, fluctuations, trend followers, etc. for physicians/ hospitals to guide decisions for marketing

Identify markets, regions, and treatment centers to prioritize for commercial success

Competitor insights



For each therapeutic area, extract sales, market share, new patent fillings on top of biological/clinical validation, clinical trials status etc. to look at competitor trends and competitor indication analysis

Use predefined/custom logic scoring using various factors to prioritize indications/pipelines based on the insights extracted

Image Source: Innoplexus

Pharmacovigilance

Deep dive

Traditional pharmacovigilance is a tedious process involving manual processing of cases and triaging the information to identify adverse events (AE) and create structured data needed for statistical or regression analysis for signal detection.

NLP allows to automate case intake and processing while uncovering signals faster by analyzing diverse data sets including literature and social listening. NLP/AI based solution are reported to reduce the cost by 25-50%.

Case Intake and Processing

Intake: ICSRs, Intake from call centers for AE reports, Product Quality Complaints (PQC) and Medical Information (MI).
NLP allows easy speech-to-text conversion, processing text data, create structured information and automate the case intake leading to efficient case processing

Triaging, QC and Signal Detection

NLP can generate narratives for cases, extract information on seriousness of the case, identify signals and tag them for manual review or rule/ML based processing, perform automated comparisons with databases like FAERs, Argus, VigiBase etc, for multimodal detection. An intuitively query-able dashboard can significantly reduce manpower and time needed

Documentation and Regulatory Reporting

NLP can help with optimizing storage, search, authoring and managing updates of reports. NLP-assisted classification, indexing and tagging of regulatory/safety documents with real-time updating of all documents can save a lot of time

Intelligent literature and social media monitoring

Auto-detection of relevant articles in literature for individual case safety reports (ICSRs) and processing of social media (Twitter, Facebook, patient forums, blogs or sites like PatientsLikeMe) posts for detection of AEs and early signs

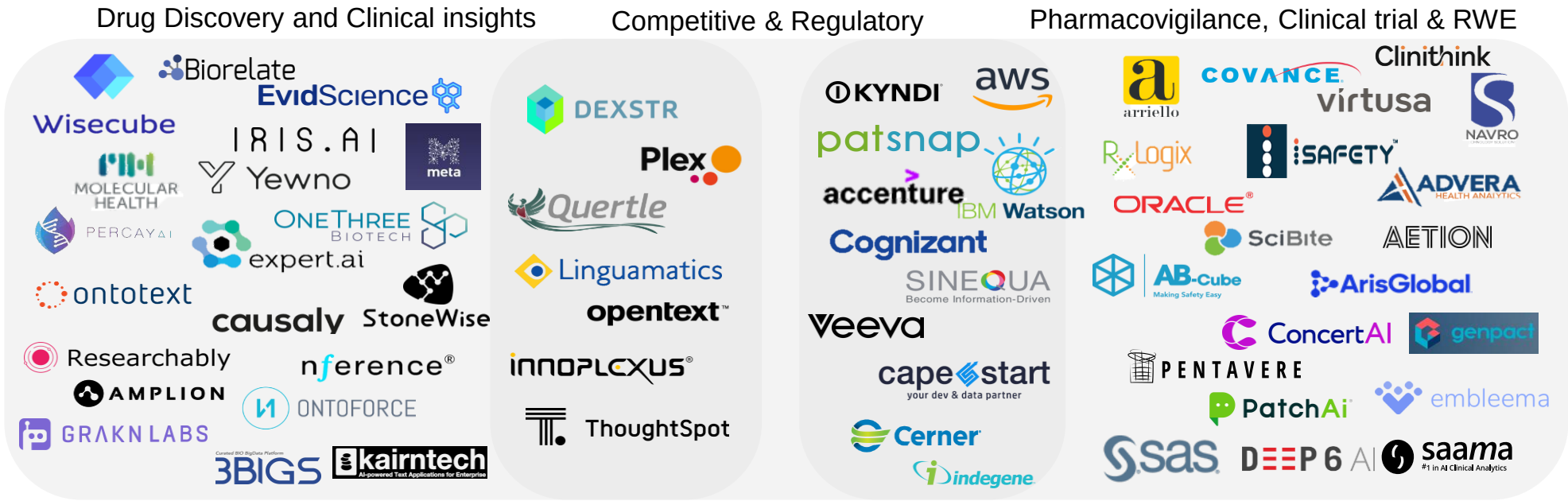
Social listening in several studies has aided early identification of signals

Pharma/Biotech is swiftly absorbing NLP applications across business units

Several start-ups are pushing to enter the space with only handful having end-to-end capabilities

According to a report, Global NLP in healthcare and life sciences market was valued at US $1.5 billion in 2018 and is expected to surpass US $21 billion by 2026 at a CAGR of 26.8%.

60+ Pharma/Biotechs have adopted solutions across applications. All of the top 20 pharma companies are using NLP based solution for at least one application.



Snapshot of a few players in the space

A lot of 'niche' start-ups have emerged in recent years with specific applications and platforms. Major tech players like IBM and AWS also finding their way-in with generic tools but struggle to understand the complexities of pharma and life science

NLP solution should be specific to the business problem being addressed

Key considerations while adopting an NLP solution

Advanced linguistics & domain expertise

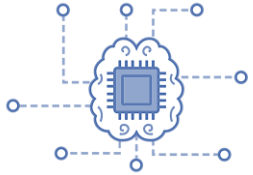
It should be able to recognize life science words and meanings robustly to extract relationships, and use a semantic algorithm to establish, understand and detect the mention of a concept with transparency

Deployment options & security

An option to choose between cloud deployment, an on-premise installation, or a hybrid model for flexibility along with strong cybersecurity measures like HTTPS protocol and HIPAA-compliant servers

Interoperability & Flexibility

An open architecture with a programming interface that supports custom workflows and integration of other analytics tools (like document processing tool) to create a holistic solution. An open search language supporting all NLP functionality is a plus



Integration of new data & ML/DL models

NLP allows structuring of free text which can then be used to push into other ML models to identify patterns. The NLP solution should provide an option to integrate other ML models (like QSAR or ADMET prediction) in the workflows

Scalability & Performance

The platform should be able process large volumes of text data and intake variety of inputs like pdf, ppt, word etc. and to establish robust ontologies and network of the extracted entities, while defining meaningful relationships

Interface and use cases

The NLP solution should have published papers or case studies in the specific areas of interest along with user-friendly UI/UX

MP Group can catalyze your AI/NLP initiative

With over 3 decades of diverse experience and integrated perspective in domestic and global BioPharma, and deep understanding of AI space, MP Group has the capabilities to help establish your NLP initiatives

MP Team will be happy to be an extension of the management team and help with one or more of the below initiatives:

- Asses the internal segments that can benefit from augmentation by NLP platforms and identify the best use cases/solutions for big data/text analytics
- Identify business segments for short-term and long-term benefit from AI interventions and strategize implementation of solutions in a step-wise manner
- Identify partnering or investment opportunities aligning with company's vision
- Technical due diligence to evaluate the NLP platforms best suited for the need

We invite you to write to us -

Viren Mehta
mehta@mpglobal.com

Neel Fofaria
neel@mpadvisor.com

Amandeep Singh
amandeep@mpadvisor.com

MEHTA PARTNERS

Advisors