The Pistoia Alliance

NLP Use Case Database Project

Project champion: Etzard Stolte, Roche
Project manager: Birthe Nielsen: birthe.nielsen@pistoiaalliance.org

Vision

Roche

Have all of PT's knowledge at your fingertips. Get answers on the information hidden in millions of records across Roche - in plain English and without the need to run a digital PoC.

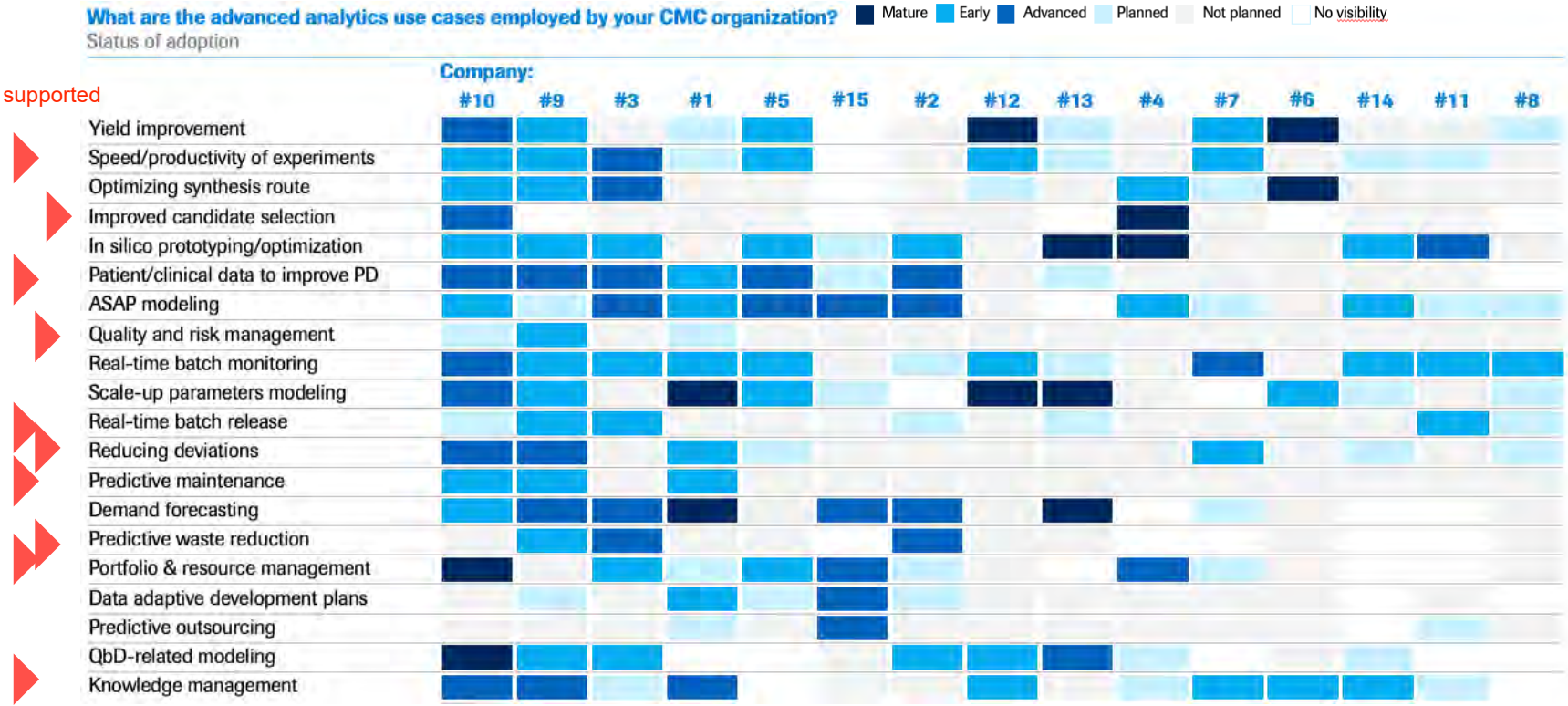
Benefit from advances in Machine Learning to automatically identify, retrieve, correlate and prioritise insights, in order to uncover latent trends, build deeper understanding of causes and increase our efficiency.

Wide range of use cases

theory

Machine Learning (NLP, NER) uses cases in PT (McKinsey 2019)

NLP supported



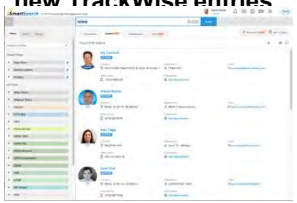
Meaning Based Computing

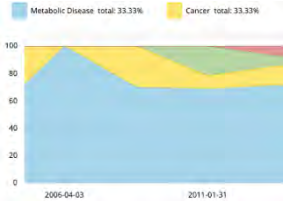
Use Cases types and examples

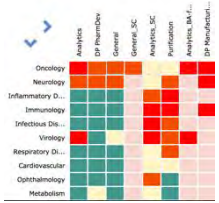
use cases

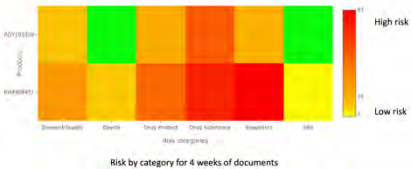
Roche

	Concepts	Trend Detection	Quality Controls	Risk Prediction	Cause Analysis
Type	Automatically extract and correlate concepts using semantic tools	*Interactively explore emerging concepts / trends*	*Monitor operational & content quality measures hidden in un-structured information*	*Real-time classifiers to sense risks across sources, sites & organizations*	*Extract relationships among concepts in a Knowledge Graph*
Examples	• **find documents/experts** • **Automatically add quality tags to lab analysis** • **Show potentially recurring deviations for new TrackWise entries**	• sense patterns in health authority records • early warning signs for batch or release records • off-label use of products	• detect mis-filed documents in submission structures • flag spurious relationships in manufacturing logs	• sentiment analysis of HA ePARs with respect to submission success • emerging issues in batch records	• understand cause / effect relationships in manufacturing deviations • visualise relationships / dependencies inside SOPs







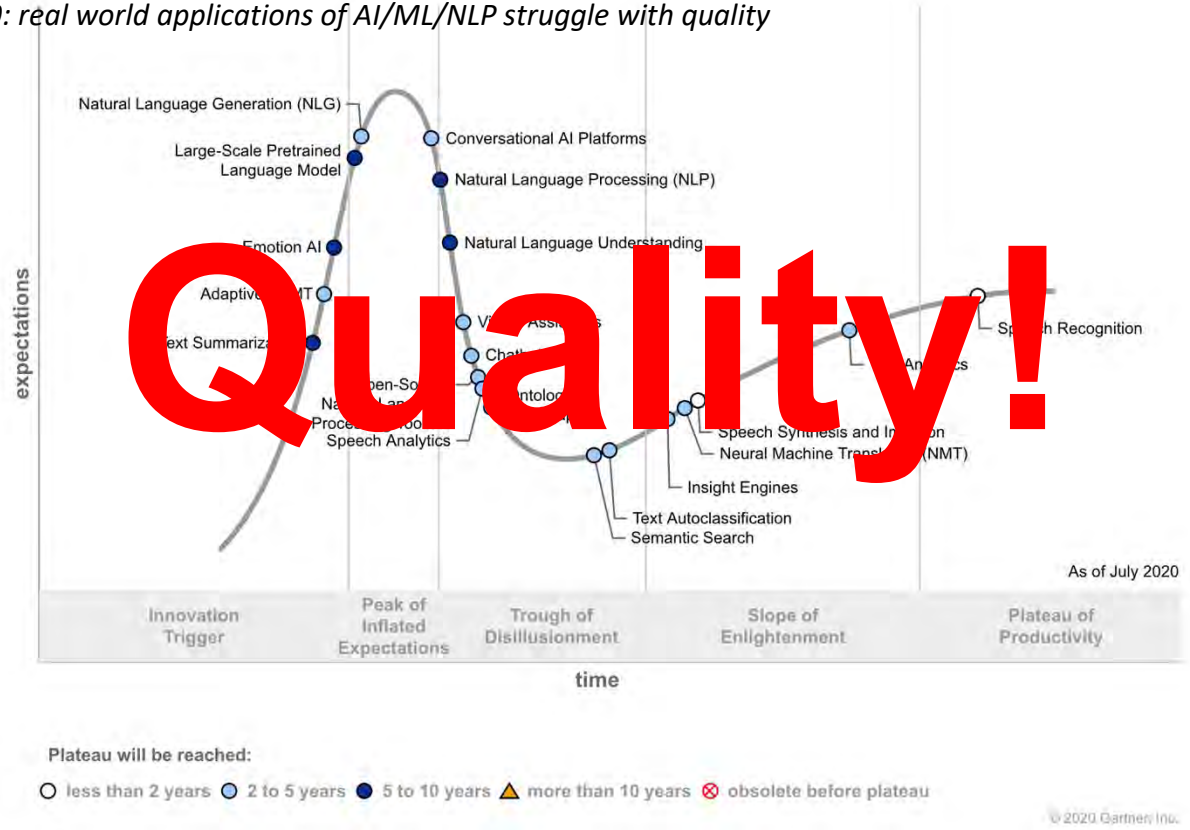


4

Hype Cycle for AI, ML, NLP

Gartner, 2020: real world applications of AI/ML/NLP struggle with quality

reality



NLP

Challenges ? Hell no, it simply does not deliver the quality we need in Pharma

reality

- Pharma is moving beyond buzzwords to value generation
- NLP is more of an art, then engineering

- Observations
  - Hard to determine up front, if a use-case will work
  - “successful” pipelines do not work for slightly different data
  - Manual curation, training sets are exceedingly work-intensive
  - Success seems very use case specific
  - Expensive to find talent that actually delivers the quality we need

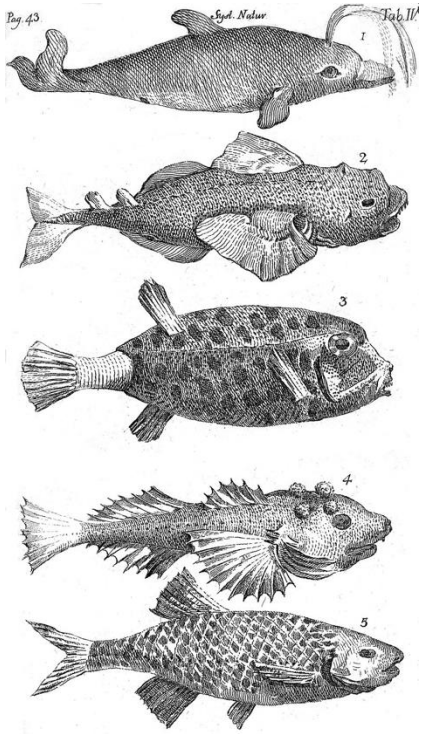
Pistoia Alliance - NLP Expert Group

Value proposition -

proposal

- Share best-practice – get an idea what works & what does not work
- Build database of use cases to be shared in this group; incl. failures

- What we should discuss today
  - Would this be of value to you?
  - Can we come up with a categorization that makes sense?
  - If NLP is an art, can we structure use-cases in a meaningful way?
  - Is a phenotypic classification sufficient (aka like Linne)?
  - Can we deliver on the effort to identify and enter your use cases iteratively into some database?



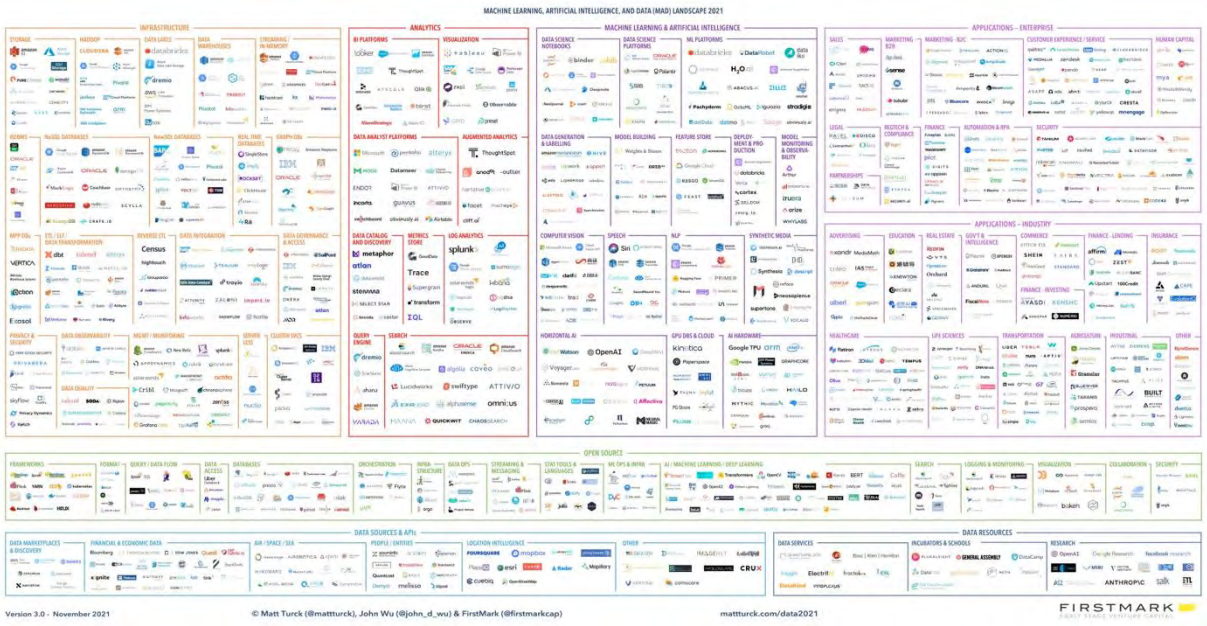
Systema naturæ sistens regna tria naturæ, in classes et ordines, genera et species redacta tabulisque æneis illustrate, Systema Naturae, 1768, Carl von Linné

Pistoia Alliance - NLP Expert Group

What do we **NOT** want to do

worry

- Encyclopedia of NLP methods
- Another NLP discussion group
- Vendor messages, vision documents, general statements
- Evangelism of a particular NLP school of thought or technology
- ???



https://mattturck.com/, November 2021 – Machine Learning & Data Landscape

Pistoia Alliance - NLP Expert Group

Database content

discussion

Roche

“Obvious” attributes

- Use case
  - Company, function,
  - summary
- Data
  - description
  - Data – sample
- Algorithm
  - summary
  - Algorithm – version, library, etc.
- Outcome
  - summary
  - Learning
- ….

possible attributes

- Discussions
  - What went wrong w/ the data, algorithm?
  - How did we make it work?
  - Was it worth fixing the issue?
  - Effort / benefit ratio
- Outlook
  - Would you do this use case again?
  - What needs to change for this to work / work better
  - Are you expecting some
  - …….