Artificial Intelligence Community of Experts

The Challenge

AI in life sciences is moving from hype to reality

Many questions remain with respect to:

- Best practices
- Ethics
- Regulatory guidelines
- Data quality
- Transparency

The Value Proposition: A Cross-Industry Collaboration to Develop Best Practices, Frame Regulations, Understand Opportunities and Pitfalls

AI and ML Community of Interest established in 2017

- New idea incubation: DataFAIRy, Ontology Training
- Drug Repurposing datathon won BioIT 2020 Innovation Award
- DDT paper on GMLPs in 2021 + new paper in progress
- Delivered 40+ webinars (11 in 2023, 3 more in planning for 2024)

Takeda Roche Zifo Boehringer Ingelheim eurofins novo nordisk abbvie AstraZeneca SERVIER ELSEVIER

Good Machine Learning Practices (GMLPs)

The Challenge

AI is a useful tool, used in many use cases in drug discovery, medical diagnostics, and patient management

We need to learn how to properly use AI as a tool

The Value Proposition: Define the Best Practices for AI use in drug discovery research

Brainstorming since 2019

- DDT paper on GMLPs in 2021 positioned the Pistoia Alliance as a thought leader in GMLP
- Later US FDA GMLP Guidelines are similar to the Pistoia Alliance recommendations
- New 2023 paper in progress: text mostly written
- More specific and actionable recommendations than in 2021
- Most likely destination “Computers in Biology and Medicine” (impact factor 6.7)
- Authors from Roche, Eurofins, Takeda, and academia
- Sponsored by Takeda

Takeda

Large Language Models (LLMs) in Life Sciences

The Challenge

There is an interest in LLMs and their use in drug discovery R&D, but the technology is poorly understood, and there is no consensus yet on what a realistic pre-competitive project in this space could be

The Value Proposition: Define the Best Practices for AI use in drug discovery research

Emergent new project idea - brainstorming since May 2023

- Initial interest limited by lack of clear and highly specific scope
- Proposals from Abbvie and the Pistoia Alliance exist, + Novo Nordisk, Boehringer Ingelheim promised
- Abbvie allocated funding; Roche, AZ, Novo Nordisk, ZS promised
- In-person meeting set for November 13th in Boston, MA

Regulation of AI and ML in the US and the EU

The Challenge

Emergent regulation:

- US FDA + UK NHS guidelines for Good Machine Learning Practices
- EU is planning the EU AI Act in 2023
- US is planning a similar act too

What is the impact on R&D?

The Value Proposition: Identify experts and learn

Webinar scheduled for January 23, 2024

- Frederik van den Broek, Senior Director, Professional Services and Consulting, Elsevier
- Koen Cobbaert, Senior Manager - Quality, Standards & Regulations, Philips
- Sophie Ollivier, Chief Data Officer R&D, Servier
- Gideon Rosenthal, Head of Research, Data Science Group

Additional activities TBD

NLP Use Case Database

The Challenge

The potential value of NLP methods for automation and insights is significant. However, the full benefits of NLP are often not fully realized.

This results in higher costs, lower value realization and disillusionment with a valuable technology.

The Value Proposition: Create a Database of Shared Insights to Increase Probability of Success of the Application of NLP Methods in R&D

This will create a bottom-up qualitative Natural Language Processing (NLP) Use Case Database, to allow NLP practitioners in pharma companies to share successes and failures with their peers. Narrowing down successful use case scenarios will lead to less experimentation and higher success rates for new NLP initiatives.

Deliverables include:

- Qualitative NLP Success Failure Database with 50+ use cases (done)
- Annotations of NLP use case methods and success/failure criteria
- Collaborative insight into why NLP use cases may fail or succeed with an industry-wide view (ongoing)

AstraZeneca Takeda Roche

Good Machine Learning Practices (GMLPs)

ELSEVIER

Drug Discovery Today

Volume 26, Issue 5, May 2021, Pages 1107-1110

Feature

Best practices for artificial intelligence in life sciences research

Vladimir A. Makarov [1], Terry Stouch [2], Brandon Allgood [3], Chris D. Willis [4], Nick Lynch [5]

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/j.drudis.2021.01.017

Highlights

- Use of artificial intelligence (AI) and machine learning (ML) methods in life sciences research calls for best practices.

Good Machine Learning Practices (GMLPs)

1. DS is not enough: domain knowledge is a must
2. Quality data (…and metadata) quality models
3. Long-term data management methodology, e.g. FAIR for life cycle planning for scientific data
4. Publish model code, and testing and training data, sufficient for reproduction of research work, along with model results
5. Use model management system
6. Use ML methods fit for problem class
7. Manage executive expectations
8. Educate your colleagues – leaders in particular
9. AI models + humans-in-the-loop “AI-in-the-loop” (Chas Nelson invented the term)
10. Experiment and fail fast if needed. A bad ML model that is quickly deemed worthless is better than a deceptive model
11. Maintain an Innovation Center for moonshot-type technology programs (this COE is an example of one)

Good Machine Learning Practices (GMLPs)

<u>Pistoia Alliance DDT paper:</u>

1. DS is not enough: domain knowledge is a must
2. Quality data □ quality models
3. Long-term data management methodology
4. Publish everything sufficient for reproduction of R&D work (model code, data, results)
5. Use model management system
6. Use ML methods fit for problem class
7. Manage executive expectations
8. Educate your colleagues – leaders in particular
9. "AI-in-the-loop"
10. Experiment and fail fast if needed. A bad ML model that is quickly deemed worthless is better than a deceptive model
11. Maintain an Innovation Center for moonshot-type technology programs (this COE is an example of one)

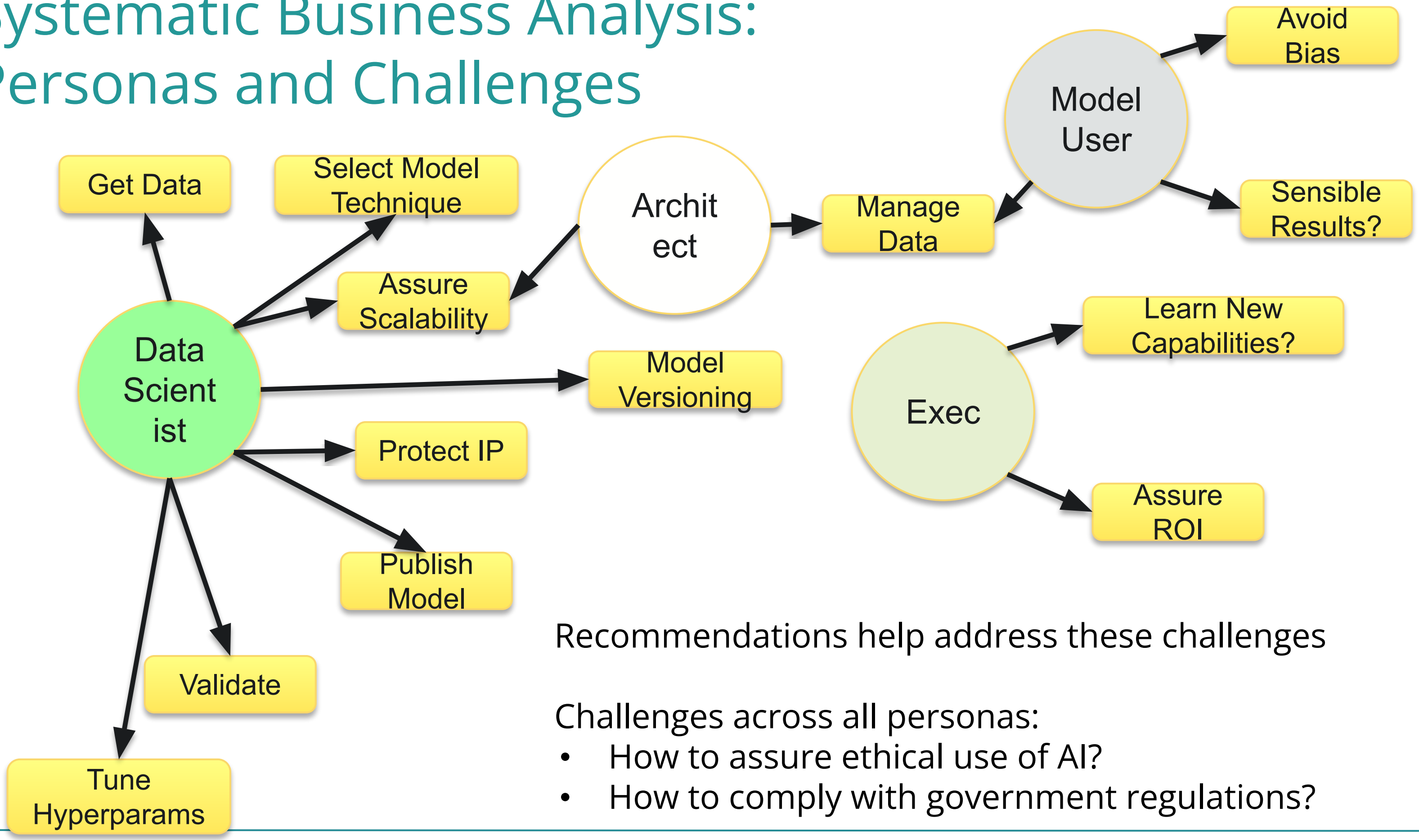
April 2020

<u>FDA GMLP Guidance</u>:

1. Multi-Disciplinary Expertise
2. Good Software Engineering Practices used
3. Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population
4. Independent testing and training data sets
5. Selected Reference Datasets Are Based Upon Best Available Methods
6. Model Design Is Tailored to the Available Data
7. Focus on Performance of Human-AI Team
8. Testing Demonstrates Device Performance during Clinically Relevant Conditions
9. Deployed Models Are Monitored for Performance and Re-training Risks are Managed
10. Users Are Provided Clear, Essential Information

October 2021

Systematic Business Analysis: Personas and Challenges



Recommendations help address these challenges

Challenges across all personas:

- How to assure ethical use of AI?
- How to comply with government regulations?

GMLP Business Analysis Illustration

(truncated to save space)

Persona	Challenges	GMLPs that address these challenges	Author Responsible for Solution(s)
Model User	MS1: How should I manage the data? Includes data protection, versioning, labeling	1. Recommend best practices from the FAIR Toolkit 2. Evaluate data for “fit for purpose”, in particular, for the metadata quality 3. Refer to best practices in Exploration, Cleaning, Normalizing, Feature Engineering, Scaling	Natalja Kurbatova, Christophe Chabbert, Berenice Wulbrecht link to doc
	MS2: How do I avoid bias?		
	MS3: How do I make sure the model produces sensible answers?	1. Set-up a “human-in-the-loop” system. Recommend tools for this, if they exist 2. Set-up business feedback mechanism for flagging model results that do not align with expectations	Chas Nelson MS3 Chas Nelson 21Jul2022
Architect	A1: How to assure scalability of AI systems?		Christophe Chabbert link to doc
	A2: How should I manage the data?	See above for Model User	Christophe Chabbert, Natalja Kurbatova link to doc
	A3: Continuous Integration/Delivery	1. Refer to best practices in DevOps 2. Automated model packaging for ease of production delivery	Elina Koletou link to doc
	A4: How to assure performance (execution speed)?		Elina Koletou link to doc
Executive	E1: How do I learn about the costs and benefits of AI/ML technologies and the limits of possible?	Make recommendations for conferences, training materials, education products, review papers, and books. These must be updated on a frequent cadence	
Data Scientist	DS1: How do I pick a suitable data modeling technique?	1. Make recommendations for methods suitable for specific problem classes; or for auto-ML systems 2. DS should learn the application domain and/or work with domain experts	

Some Less-obvious GMLP Findings

When *developing a model:*

- Compare performance to a baseline (the simplest method possible – think linear regression ☺), and engineer more complex models iteratively

When *using a model:*

- Store data preprocessing details, data and model version details, along performance results (e.g. accuracy and run-time)
  - What is not obvious: tracking these details for production-grade ML models
- Monitor deployed models for performance (e.g. data drift) and adopt a continuous risk assessment and mitigation plan, which may include retraining models whilst being aware of the retraining risks (e.g. catastrophic forgetting)

Across *all stages of model life cycle*:

- Consider ROI
- ROI came out as one the central themes in GMLP brainstorming:
  - What use cases in AI in pharma produce the **greatest return**
  - How to **measure the value of AI** use?
  - These questions have not yet been answered well – opportunity for a paper?

Abbvie LLM Proposal Summary

Proposal:
- A winning strategy is creating many special-purpose, finely prompt-tuned small sized LLM models
- Need to agree on the domain where the first special-purpose LLM can be built
- To eliminate hallucinations, these models should refer to the data from a controlled database with a standardized schema. We may populate it with a public data set
- Individual companies can then add proprietary data of their choice to private instances (thus no risk of IP exposure) or collaborate on future projects using the commonality of the schema
- Experts from collaborating companies can work on prompt-tuning the LLM(s) – crowd-sourced reinforcement learning

Outputs:
- *This database schema, optionally pre-filled with public data*
- *Prompt-tuned LLM (start with one, may develop many eventually)*
- *Lessons learned summarized in a white paper*

Why this is a good idea:
- Base for future collaborative R&D between pharma firms or with vendors
- Base for future benchmarks for evaluation and comparison of models
- Brian Martin at Abbvie is already building a prototype and is willing to open-source it

Pistoia (VM) LLM Proposal Summary

Proposal:

- Benchmark for assessment of extraction of terms (named entities) from free text
- Prior benchmark - public CREEDS data set (https://github.com/MaayanLab/creeds) now obsolete
- Prior results on CREEDS 93% accuracy for disease names and drug – disease relations
- Also need to expand the benchmark to additional ontologies – genes, protein names, compounds (IDMP-O?), complex relationship cases (gene to disease to drug)

Outputs:

- *The benchmark collection of examples*
- *Results from a few models – we have a number of NLP and KG vendor companies as members, they may be interested*
- *Lessons learned summarized in a white paper*

Why this is a good idea:

- Benchmark for a common task where LLMs are already used
- Benchmarks are inherently pre-competitive
- Use case not prone to LLM hallucinations
- Base for vendor evaluations
- Project sustainability: can be used for an on-going competition with progressively more complex use cases

NLP project timeline

Q1 2022	Q3 2022	Q3 2022	Q4/Q1 2022/23	Q2/Q4 2023
Roche ideas proposal Round table discussion with interested pharma members	Draft Use Case Database First Use Cases added	Agreed annotation of NLP use case methods and success/failure criteria	Collection of +50 use cases	Collaborative insight – wider view & next steps

Where our NLP database stands today

Named-entity recognition (NER)

Concept Extraction (taxonomy extraction, concept tagging, document classification, semantic search)

Multi-Label Text-Classification

Unsupervised topic modeling

Extractive auto-summarization

Similarity Search

Chatbots/Virtual Assistants/Q&A service

Text Data Augmentation

Text translation

Topic discovery

NLP Document prediction

Natural Language text to Python code generation

Relationship Extraction for Knowledge Graphs

Quality Analytics

Patent Annotation

Please tell me if traditional NLP tasks are mostly solved with the development of ChatGPT, and if NLP engineers are going to lose their jobs?

While ChatGPT has made significant advancements in natural language processing (NLP), it does not mean that all traditional NLP tasks are completely solved or that NLP engineers will lose their jobs. ChatGPT excels in generating human-like text based on given prompts, and it has demonstrated impressive capabilities in language understanding and generation. However, there are still several challenges and limitations in NLP that have not been fully addressed.

Database Content

Use Case

- Deployment type
- Project Level
- Company
- Language
- FTE involved/team
- etc

Data

- Description
- Size
- Issues
- Preparation
- etc

Algorithm

- Summary
- Version
- Liberay
- etc

Outcome

- Learnings
- Success rating

Discussions

- Effort/Benefit ratio
- Non-NLP approach
- Support used
- Lessons learned

Oulook

- Build vs Buy recommendations
- Other applications of technology
- etc

Opportunities

- Expand the database to LLM use cases
- Classify and group use cases
- Obtain ROI information for use cases or broad use case categories
- Identify patterns of success and failure (high/low ROI)
- Publish the findings
- Expand the stakeholder cohort
  - May request input from technology vendor firms for a different view