Leveraging Artificial Intelligence (AI) and Machine Learning (ML) to Support Regulatory Efficiency – Current Progress

FY22 GDUFA Public Workshop

Sub-Session 2C: Artificial Intelligence and Machine Learning for generic drug development and assessment

Meng Hu

Division of Quantitative Methods and Modeling
Office of Research and Standards
Office of Generic Drugs
CDER | U.S. FDA

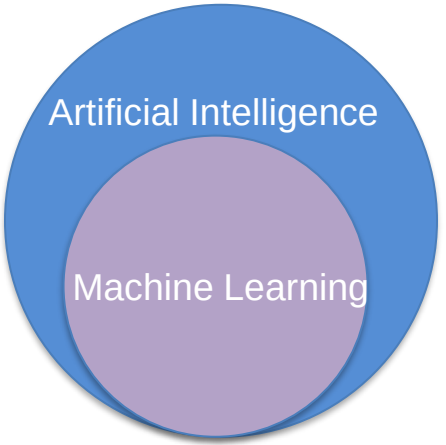
May 9, 2022

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

What is Artificial Intelligence (AI)?

- "*It is the science and engineering of making intelligent machines, especially intelligent computer programs*." by **John McCarthy** in 2004



- AI is transforming many areas of everyday life:
  - ATM (automated teller machine)
  - Smartphone (e.g., face recognition)
  - Autopilot
  - Chatbot
  - Personalized recommendation

AI offers opportunities to facilitate generic drug development and regulatory assessment

- Development of automation tools:
  - Enhanced efficiency (e.g., saving time)
  - Improved consistency (e.g., reducing human error)
  - High-quality deliverables

- Utilization of advanced data analytics methods:
  - Promoting business intelligence
  - Supporting regulatory assessment

AI offers opportunities to facilitate generic drug development and regulatory assessment

- Development of automation tools:
  - Enhanced efficiency (e.g., saving time)
  - Improved consistency (e.g., reducing human error)
  - High-quality deliverables

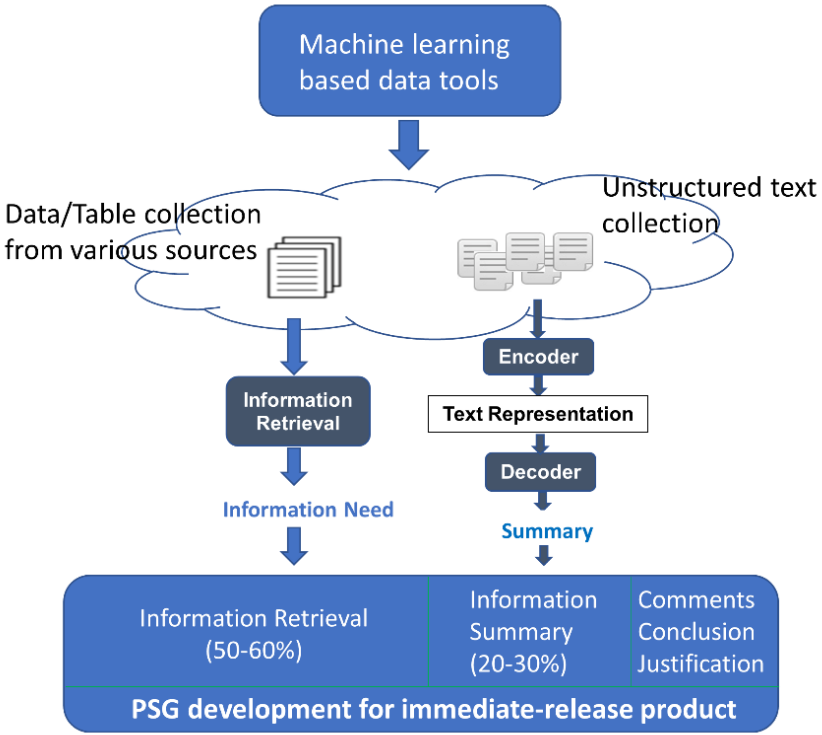
- Utilization of advanced data analytics methods [1-5]:
  - Promoting business intelligence
  - Supporting regulatory assessment

Development of automation tools

- Aims:
  - Automating labor-intensive works
  - Enhancing efficiency, consistency and quality

- Challenges:
  - Data sufficiency
  - Pattern recognition
  - AI model selection
  - Dealing with unstructured data

An ongoing effort: Text Analysis and Machine Learning to Facilitate PSG Development (75F40119C10106)



Challenges:

- Evolving layouts of source documents (e.g., drug labeling and internal review documents).
- Need for information retrieval based on semantic understanding.
- Capturing information from unstructured text (e.g., reviewer's analysis/comments in review documents)
- Choosing a proper Natural Language Processing (NLP) model

www.fda.gov
https://www.fda.gov/drugs/generic-drugs/science-research

Sematic understanding-based information retrieval from drug labeling for "food effect"

EXAMPLE 1: NDA 205832

Absorption

Nintedanib reached maximum plasma concentrations approximately 2 to 4 hours after oral administration as a soft gelatin capsule under fed conditions. The absolute bioavailability of a 100 mg dose was 4.7% (90% CI:3.62 to 6.08) in healthy volunteers. Absorption and bioavailability are decreased by transporter effects and substantial first-pass metabolism.

After food intake, nintedanib exposure increased by approximately 20% compared to administration under fasted conditions (90% CI: 95.3% to 152.5%) and absorption was delayed (median t_{max} fasted: 2.00 hours; fed: 3.98 hours), irrespective of the food type.

https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=da1c9f37-779e-4682-816f-93d0faa4cfc9

EXAMPLE 2: NDA 210491

Absorption

After a single dose in healthy subjects in the fed state, tezacaftor was absorbed with a median (range) time to maximum concentration (t_{max}) of approximately 4 hours (2 to 6 hours). The median (range) t_{max} of ivacaftor was approximately 6 hours (3 to 10 hours) in the fed state.

When a single dose of tezacaftor/ivacaftor was administered with fat-containing foods, tezacaftor exposure was similar and ivacaftor exposure was approximately 3 times higher than when taken in a fasting state.

https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=302ae804-37db-44fd-ac2f-3dbdeda9aa4b

Note: keyword searching does not work for this task. For example, searching for "food" will lead to a high false positive rate.

Current progress

- The state-of-the-art Bidirectional Encoder Representations from Transformers (BERT) model was used for this NLP application.
- An NLP pipeline was developed to extract drug product information (e.g., ADME information) from drug labeling with minimal human intervention.
- A paper on automatic ADME information retrieval from drug labeling was published (*Frontiers in Research Metrics and Analytics*) [6].

Discussion

- AI technologies bring opportunities to advance development and regulatory assessment of generic drugs.

- **We need your input and insight on how to take full advantage of this opportunity to facilitate generic drug development.**

- Please join the following panel discussion to share your thoughts/ideas.


FDA
U.S. FOOD & DRUG
ADMINISTRATION

References

1. Hu, et al. Predictive analysis of first abbreviated new drug application submission for new chemical entities based on machine learning methodology. *Clinical Pharmacology & Therapeutics* 106 (1), 174-181

2. Gong, et al. Big data toolsets to pharmacometrics: application of machine learning for time-to-event analysis. *Clinical and translational science* 11 (3), 305-311

3. Gong, et al. Heterogeneous treatment effect analysis based on machine-learning methodology. *CPT: pharmacometrics & systems pharmacology* 10 (11), 1433-1443

4. Hu, et al. Equivalence Testing of Complex Particle Size Distribution Profiles Based on Earth Mover's Distance. *AAPS Journal* 20(3):62

5. Rogstad , et al. Modern analytics for synthetically derived complex drug substances: NMR, AFFF–MALS, and MS tests for glatiramer acetate. *Analytical and bioanalytical chemistry* 407 (29), 8647-8659

6. Shi, et al. Information Extraction From FDA Drug Labeling to Enhance Product-Specific Guidance Assessment Using Natural Language Processing. *Frontiers in Research Metrics and Analytics* 6:670006