The Role of Artificial Intelligence in Drug Discovery and Development

by Michael Liebman

Artificial Intelligence (AI) is an exciting, growing field. Due to the high and growing number of data, the comprehensive evaluation of information behind data makes AI tools indispensable. In Drug Discovery and Development the application of AI has become important to accelerate progress and enhance decision making in many fields and disciplines of medicinal chemistry, upscaling, molecular and cell biology, pharmacology, pharmacokinetics, formulation development and toxicology. In clinical testing AI has high importance in increasing success rates by enhancing trial design (biomarkers, efficacy parameters, dose selection, trial duration), selection of the target patient population, patient stratification and evaluation of patient samples. The increasing relevance of AI in drug discovery and development is reflected by the growing number of start-up companies specialized in this field, the growing number of collaborations from Pharma with AI platforms, and the high number of articles and reviews reporting current applications, their success and limitations.

In the first part of this article, Michael Liebman focuses on a general overview on AI in drug discovery and development; the second part provided by Yann Gasthon-Mathé and his colleagues from IKTOS (France)—an AI company specialized in drug discovery and development-related AI applications—highlights key points to succeed in AI drug discovery projects.

This feature is an integral part of the first D3SC Newsletter — See https://iupac.org/d3sc-newsletter-sep2021/>

1929: *Japanese biologist and professor Makoto Nishimura created Gakutensoku, the first robot to be built in Japan. Gakutensoku translates to "learning from the laws of nature," implying the robot's artificially intelligent mind could derive knowledge from people and nature. Some of its features included moving its head and hands as well as changing its facial expressions. (reproduced from [1])*

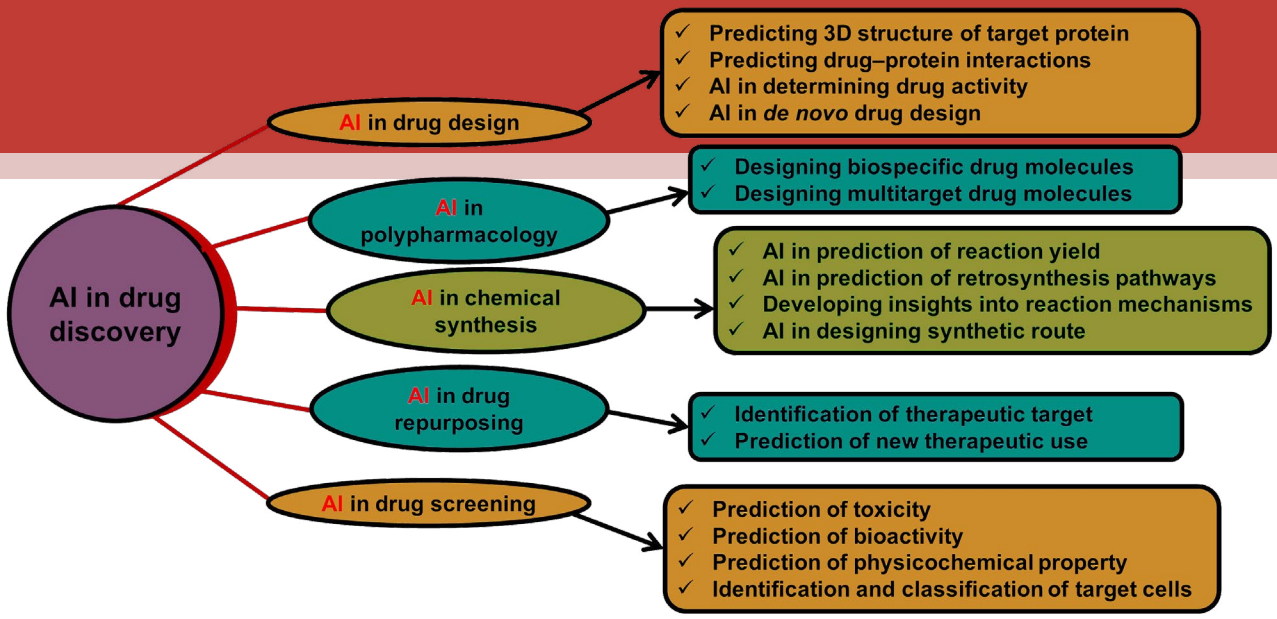
General Overview

Brief highlight of AI key advantages, current status of maturity and its impact on Drug Discovery & Development. A few review references, including a recent article on this very topic of AI in drug discovery and development brings us up to speed on the literature and are cited [1-4].

Two fields, artificial intelligence and drug discovery/development, have long histories that pre-date their current names and definitions. Although they have evolved separately and not necessarily linearly, their paths have more recently begun to converge as significant advances in computational and experimental/technological capabilities provide access to enormous amounts of data and computing power. This commentary provides an introduction to the state of this convergence and what critical gaps remain as challenges still to be addressed.

The term "artificial intelligence" (AI) was promulgated by John McCarthy in 1956, although the concepts stem from Greek mythology (Talus and Pygmalion), from Hebrew folklore (Golem) and later the alchemists, Paracelsus (Homunculus) and Roger Bacon. Although mathematical logic, critical to AI as we know it today, was developed in the 20th century by Russell and Whitehead and Hilbert, it originated with formal reasoning among the Chinese, Indian and, Greek philosophers. Today we recognize Babbage, Lovelace, Shannon, Turing and von Neumann, Minsky, McCarthy, Newell, and Simon in the recent family tree of AI.

We can also separate "drug development" into two main components: drug discovery/design/preclinical optimization and clinical drug development. The former's history extends to ancient times when natural products, *e.g.* herbals, minerals, animal substances, were used and continues to the highly refined



Role of artificial intelligence (AI) in drug discovery. AI can be used effectively in different parts of drug discovery, including drug design, chemical synthesis, drug screening, polypharmacology, and drug repurposing. (reproduced from [3])

experimental and computational approaches applied today. The latter largely reflects the introduction of regulatory processes and procedures that help to insure safety and efficacy and then leads to commercialization. Drug discovery initially involved trial and error and practitioners who could remember what worked and what did not. For many centuries it existed in conflict with religious views until being reborn in the Renaissance period, with pharmacology being born in the late 19th century. Drug design has evolved from the early experimental efforts to modify and improve natural products through medicinal chemistry to incorporate computational approaches in the 1970's. Clinical development may have started with Avicenna (1025), progressed to Lind's scurvy trial and Flint's first placebo trial in 1863, but it was not until 1943 that the first double blind controlled trial (Patulin for common cold) and until 1946 that the first randomized curative trial (Streptomycin for TB) were carried out. Today, regulatory agencies, epidemiologists and biostatisticians are involved in trial design for clinical development, and the marketing groups greatly impact the commercialization planning and effort.

AI in Drug Discovery/Design/Preclinical Optimization

Drug discovery/design/preclinical optimization focuses on 1) the identification of targetable processes or molecules that bear responsibility for a specific clinical condition, and sometimes in a specific patient population, and 2) the identification/design/preclinical optimization of an agent to modulate this modified physiology. AI methods are being applied to both these steps.

Target/Process identification typically stems from the collection and analysis of large data sets, *e.g.* population-based studies of clinical observations sometimes linked with genomic data. The application of ontologies and knowledge graphs to link disparate data provides an ongoing opportunity to continually update existing data sets and/or integrate new data sources into the analysis process which focuses on stratification of the disease and of the patient. Machine learning and deep learning methods are being applied as clinical observations are being enhanced with data from digital health applications, *e.g.* Fitbits, *etc*, that can provide almost continuous data feeds that require complex feature extraction. Diagnostic image processing, *e.g.* x-ray, MRI, pathology, *etc,* are being enhanced with AI methods to sharpen feature extraction and provide secondary confirmation of manual interpretations.

***1949**: Computer scientist Edmund Berkeley's book "Giant Brains: Or Machines That Think" noted that machines have increasingly been capable of handling large amounts of information with speed and skill. He went on to compare machines to a human brain if it were made of "hardware and wire instead of flesh and nerves," describing machine ability to that of the human mind, stating that "a machine, therefore, can think." (reproduced from [1])*

Natural language processing (NLP) is being used to "read and interpret" clinical notes. Knowledge graph methods are being used to integrate and represent the combined data/information and support potential analysis to suggest potential molecular or process/pathway targets.

Drug selection can involve *de novo* design, screening of the deck molecules for a Hit or modification of existing molecules, Hit to Lead optimization of pre clinical parameters such as PK/PD including both small molecules and peptides/proteins, *etc* as well as re-purposing of existing molecules for new indications and also potential combination therapies. As it focuses on both efficacy and safety, drug selection needs to consider potential drug-drug interactions, risk for adverse events and potential for differential response among diverse patient populations. The application of AI methods enables greater opportunity for computational screening of potential drug candidates as it can significantly increase the feature space used to define specific properties for inclusion or exclusion. Machine learning approaches have the potential to learn from existing data from existing compound libraries, including high throughput screening results, to help identify critical features in both categories and to enable ongoing, rapid computational screening of large compound databases using these features. Such analyses can also "infer" potential features for further molecular modification through medicinal chemistry approaches. Ongoing screening of existing libraries provides opportunities for re-purposing of compounds for new indications and/or development of combination therapies that can compensate for potential side-effects or establish multi-target treatments. It is notable that one of the first applications of AI methods in drug discovery focused on reverse engineering of existing patents and generation of novel synthesis pathways. Interestingly, one of the AI companies, Exscientia, announced the first AI-designed Immuno-Oncology drug to enter clinical trials (https://www.exscientia.ai/news-insights/exscientia-first-ai-designed-immuno-oncology-drug-trial).

AI in Drug Development

Drug development includes two major phases: the first, pre-clinical testing through clinical trials and the second, submission for regulatory approval. AI has been introduced more recently into the developmental phases, primarily to enable the aggregation, organization and analysis of "big data" with the goal of improving trial performance and regulatory approval.

Clinical trials involve the identification and evaluation of clinical test sites and personnel that meet specific criteria including: access to adequate trial participants, clinical competency, administrative and technical support that can meet performance requirements and identification of potential trial participants from clinical records. AI methods are being applied to screen potential sites for their history for meeting critical enrollment criteria and compliance, analyzing electronic health records to identify potential candidates who meet trial inclusion/exclusion criteria and integration of real world data that may support trial performance. More recently the addition of digital health monitoring into clinical trial protocols and the need for advanced data processing have also required the addition of machine learning technologies. AI is also providing dynamic indications of participant status as well as identifying early indications of potential adverse events.

Regulatory submission closely follows strict protocols and procedures that are evolving to include both the data digital monitoring systems and real world evidence. The use of real world data may additionally require integration of multiple data sources with ontologies or knowledge graphs as noted earlier in this commentary.

Challenges and opportunities

The complexities of drug design and development provide a natural target for the application of the methods and technologies of artificial intelligence to achieve greater success and at potentially lower cost and shorter time to market. To date, however, the results have been more incremental than disruptive but hold much promise for the future. The technology alone, however, does not directly address some critical challenges which, in turn, present potential opportunities that could enhance clinical and commercial success.

Target selection: How well is the disease/condition diagnosed and stratified, *i.e.* is the phenotype adequately defined? How comprehensive is patient stratification, *i.e.* clinical history, co-morbidities, lifestyle, environment, genomics, *etc.?* Is the target generalizable across the real world patient population, *i.e.* observed diversity?

Drug design/selection: Can we "decode" deep learning to interpret results? All patients reflect co-morbidities and poly-pharmacy, how are these being addressed? How well are pathway modulators being modeled as to individual targets and response?

Clinical trials: How do inclusion/exclusion criteria approximate real world patients? How does this impact commercialization post-approval? Could disease and/or population stratification lead to shorter, more directed clinical trials?

Conclusions

Drug design and development will continue to be an early adopter of new and evolving technologies, both experimental and computational. Among the challenges is whether to apply these technologies to enhance the existing pipeline and processes or is the real opportunity to re-engineer the processes in light of these technologies? Big data, digital healthcare, remote monitoring, and genomics will drive the need to explore how computational and reasoning approaches can be applied to enhance the process both in terms of clinical significance and cost reduction. Artificial intelligence methods hold great promise towards these goals but their success will depend on aligning the right question with the right technology.

References:

1. A Complete History of Artificial Intelligence, May 25, 2021 Rebecca Reynoso, https://www.g2.com/articles/history-of-artificial-intelligence
2. Artificial intelligence in chemistry and drug design, Nathan Brown *et al, Journal of Computer-Aided Molecular Design* (2020) 34:709–715; https://doi.org/10.1007/s10822-020-00317-x
3. Artificial intelligence in drug discovery and development, Debleena Paul *et al, Drug Discov Today*. Jan 2021 Jan; 26(1): 80–93; https://doi.org/10.1016/j.drudis.2020.10.010
4. History of artificial intelligence in medicine, Vivek Kau, Sarah Enslin, and Seth A.Gross, *Gastrointestinal Endoscopy*, October 2020, 92(4): 807-815; https://doi.org/10.1016/j.gie.2020.06.040

Key points to succeed in Artificial Intelligence drug discovery projects

by Quentin Perron, Vinicius Barros Ribeiro da Silva, Brian Atwood, and Yann Gaston-Mathé

Drug discovery and development is an expensive, complex, and time-consuming task [5]. Recently, the development of artificial intelligence (AI) approaches to drug discovery, specifically *de novo* drug design through the use of deep generative models, has triggered a lot of interest in the drug hunter community, especially as an important tool to speed up the process [6].

Since 2017, we at Iktos have worked in collaboration with industry and academia in many different projects, from hit discovery to lead optimization, using AI with ligand and structure-based techniques. In a recently published preprint work, we described the results of a successful collaboration between Iktos and Servier in a late-stage lead optimization project [7]. At this occasion we described, for the first time, the successful application of deep learning to *de novo* design for solving a Multi-Parameter Optimization (MPO) issue in an actual drug discovery project. Using the initial dataset of the project, with 881 molecules measured on 11 biological assays, we built 11 QSAR models and used them in combination with our deep learning-based AI *de novo* design algorithm. We were able to automatically generate 150 virtual compounds predicted as active on all 11 objectives. 20 molecules were selected as the most promising, and 11 were synthetized and tested. Interestingly, the 11 AI-designed compounds that were synthesized and tested displayed functional groups that were either rare in the initial dataset or never tried earlier in the project. Ultimately, one of the 11 AI-designed molecules met all the objectives of the project at the same time, suggesting that this method can propose innovative new molecules to solve MPO

1980: *WABOT-2 was built at Waseda University. This inception of the WABOT allowed the humanoid to communicate with people as well as read musical scores and play music on an electronic organ. (Reproduced from [1])*