How Big Pharma Adopts AI To Boost Drug Discovery

Oct. 8, 2018 by Andrii Buvailo

(Last updated 08.10.2018)

The type of artificial intelligence (AI) which scares some of the greatest minds, like Elon Musk and Stephen Hawking, is called “general artificial intelligence” -- the one which can “think” pretty much like humans do, and which can quickly evolve into a dangerous “superintelligence”. There is a notion that it might be invented in the nearest decades, but today we are definitely not there yet. The AI which is making headlines these days is a “narrow artificial intelligence”, a limited type of machine “intelligence” able to solve only a specific task or a group of tasks. It can’t go anywhere beyond specifics of the problem for which it is designed, so apparently, it will not hurt anyone in the nearest time. But already now it can provide meaningful practical results on those narrow tasks, like natural language processing, image recognition, controlling self-driving cars, and helping develop new drugs more efficiently. With the ability to find hidden and unintuitive patterns in vast amounts of data in ways that no human can do, AI represents a considerable promise to transform many industries, including pharma and biotech.

The interest in AI-driven solutions for early stage drug discovery is growing steadily among biopharma leaders with a projected market volume reaching $10B by 2024 (for AI-based medical imaging, diagnostics, personal AI assistants, drug discovery, and genomics). The last couple of years were marked by a wave of new R&D collaborations between key biopharma players and AI-driven companies, primarily startups. Let’s see who is doing what in the biopharmaceutical AI landscape.

(Since most AI-driven companies use a mix of different approaches and rely on interdisciplinary sources of data for their modeling work, the below classification of AI use cases is illustrative):

1. **AI for drug target identification and validation**

In June 2017, Genentech announced a research partnership with GNS Healthcare to identify and validate novel cancer drug targets using the company's proprietary REFS casual machine learning and simulation AI platform. According to GNS Healthcare website, their platform is the first commercially available
AI-driven tool that automates the transformation of diverse streams of biomedical and healthcare data (such as longitudinal electronic medical records (EMR records), next generation sequencing, and other ‘omic data) into mechanistic, computer models that are representative of individual patients.

A month later GlaxoSmithKline (GSK) formed a collaboration with a Baltimore-based AI-driven company Insilico Medicine, to explore how Insilico’s AI capability can help in the identification of novel biological targets and pathways of interest to GSK. According to Insilico’s website, the company is using a multi-stage highly iterative in silico drug discovery process involving the application of generative adversarial networks and reinforcement learning algorithms. The process is described as a closed feedback loop containing stages like data mining, hypothesis generation, lead compound identification and optimization etc., allowing for a gradual improvement of the overall output prediction result over time and with more data available.

2. AI for target based and phenotypic drug discovery

Very active in the AI space, GSK signed a $43 M drug discovery collaboration with U.K.-based AI-driven startup Exscientia to identify small molecules for ten selected targets across undisclosed therapeutic areas. Using a rapid “design-make-test” cycle Exscientia is able to design new molecules using AI-system, employing as well phenotypic and high content screening data, and assess their potency, selectivity and binding affinity towards specific targets. The projects will be heavily supported by Exscientia’s big-data resources—from its medicinal chemistry and large-scale bio-assays.

A month earlier Takeda announced a multi-year research partnership with AI-driven drug design company Numerate to develop new clinical candidates in oncology, gastroenterology, and central nervous system disorders. Numerate plans to apply AI-based modeling at every stage of the process -- from hit finding and expansion through lead design/optimization -- to absorption, distribution, metabolism, and excretion (ADME)/toxicity predictions. As stated on the website of Numerate, their AI-platform is able to work with data points obtained from different studies -- from high-content, low-throughput phenotypic assays as well as high-throughput screening, structure-based design and traditional computational methods. Trained with versatile information, the AI-system can probe very large chemical spaces and identify the most promising drug candidates.

Numerate is also involved in a recently announced partnership with Servier -- it will focus on the design of small molecule modulators of ryanodine receptor 2 (RyR2), a highly challenging target identified as
important in cardiovascular diseases.

Another drug discovery focused company advancing artificial intelligence technology is Atomwise, which uses deep convolutional neural network AtomNet to empower structure based drug design. Atomwise was founded in 2015 and since then established a number of confidential collaborations with high profile organization, such as pharma giants Merck, and Abbvie, as well as Stanford University, Duke University School of Medicine and so on. Atomwise just announced a new evaluation agreement with Pfizer to demonstrate its ability to identify promising drug candidates for up to three proteins of choice by Pfizer.

The company’s AI-based system can learn to recognize protein and ligand structures and interactions starting from the simplest features up to full objects, basically “teaching itself a chemistry course”. Further, it can model bioactivity of small molecules and chemical interactions and identify new molecules for the targets with previously unknown modulators. AtomNet outperformed traditional docking approaches, as was shown on benchmark examples.

Structure-based drug discovery (SBDD) approach complemented by the use of AI (deep learning) technology is the primary focus of a partnership between Korean biopharmaceutical company CrystalGenomics and AI-driven technology company Standigm aiming at discovering and developing novel drugs to treat cancer, rheumatoid arthritis and liver related diseases.

From earlier examples of applying machine learning for target based drug discovery, it should be also noted a collaboration between Switzerland-based contract research organization THERAMetrics and a computational drug discovery company Cloud Pharmaceuticals aiming to develop novel drugs against Orphan CNS Diseases. Cloud Pharmaceuticals uses Quantum Molecular Design(sm) platform to develop small molecules and peptides with desirable activity on the selected targets.

Update: Cloud Pharmaceuticals recently announced a drug design collaboration with GSK to assist the pharma company with its AI-expertise in designing novel and safe small molecules against biological targets of interest.

3. AI for dealing with biomedical, clinical and patient data

A notable opportunity for AI models to shine in the area of drug discovery is using biomedical and clinical data to draw unintuitive insights about drug candidates, or even attempting to model the whole biological systems to identify novel pathways, targets and biomarkers. Earlier this year, Santen Pharmaceuticals, a
Japanese leader in the ophthalmic field, entered a strategic research collaboration with TwoXAR -- an AI-driven biopharmaceutical company. Instead of using molecular modeling techniques, TwoXAR works with real world biomedical data including gene expression measurements, protein interaction networks, and clinical records. By examining billions of points of information, TwoXAR’s AI platform is able to determine what is relevant, and what is noise, leading to a set of associations indicating the effectiveness of certain small molecules.

In the case with Santen, TwoXAR is aiming to find new drug candidates for glaucoma treatment and for that the company is screening large catalogs of molecules, associated with known data, such as protein structures, binding affinities etc. The data are then linked with molecular changes in glaucoma to derive unique disease-drug associations.

Another company focusing its AI-driven efforts on biology, rather than chemistry is Berg Health, advancing deep-learning screening of biomarkers from patient data and “multi-omic” modeling approaches. The types of data Berg is feeding to its AI algorithms includes not just genome data, but also the proteome, metabolome, and the lipidome of the biological samples to unravel the complex biological networks playing roles in diseases. That, in turn, can likely help identify medications for specific patient populations and, on the other hand, sift through the drug candidates that are likely to fail.

Berg recently announced its first major pharma collaboration with AstraZeneca focusing on identification and evaluation of novel approaches to treat Parkinson’s disease and other neurological disorders. Following the deal, Berg will use its AI-driven drug discovery platform to explore a selection of chemical fragments provided by AstraZeneca to find promising drug candidates.

In May 2017, BioXcel Corporation, a privately held biopharmaceutical company pioneering the application of big data analytics and machine learning-based AI for drug development, announced the launch of InveniAI -- artificial intelligence and big data platform company focused on providing its two R&D engines, EvolverAI and PharmGPS®, to pharma and biotech partners. While the first tool is designed to augment human intelligence with artificial intelligence for analyzing and visualizing big (biomedical, clinical, research) data and identifying new patterns and insights, the other tools helps understand how drugs work and which of them are useful in satisfying unmet medical needs.

BioXcel, in its turn, has a long history of big data and machine-learning-driven collaborations with pharmaceutical companies, including repurposing program with Takeda, Alnylam (RNA interference (RNAi) therapeutics), Axcella (discovery and development of amino acid biologics), and Centrexion (Collaboration in Chronic Pain research).
4. AI for polypharmacology discovery

While “one target one disease” has been a dominating paradigm in drug discovery for years, it is becoming obvious that many diseases are too complex to be efficiently cured within this paradigm. A multitarget drug discovery approach is a promising way to make more efficient medicines.

With this in mind, Sanofi put a $274 M deal with AI-driven Exscientia in 2017 to discover and develop bispecific small molecules that treat diabetes and its comorbidities. Exscientia role will be to come up with pairs of targets, related to glucose control, NASH, weight management and other diabetes-related areas, and generate bi-specific small molecule ligands using AI-based platform.

Similar multitarget strategy was pursued in the other Exscientia research collaboration with Evotec in 2016 to discover and develop first-in-class bispecific small molecule immuno-oncology therapies. As in the case with Sanofi, Exscientia will provide value to Evotec via its AI-driven platform to purposely design bispecific small molecules that can address multiple targets through a single drug.

Tracking back Exscientia ability to discover multi-targeting small molecules, it is important to note the startup’s announcement in September 2015 about the initial results to deliver a bispecific, dual-agonist compound that selectively activates two GPCR receptors from two distinct families, following the earlier collaboration with Japanese pharmaceutical giant Sumitomo Dainippon Pharma.

5. AI for drug repurposing programs

Drug repurposing is one of the golden mines for AI-based technologies to drive value since a lot of data is already known about the drug in question. Repurposing previously known drugs or late-stage drug candidates towards new therapeutic areas is also a desired strategy for many biopharmaceutical companies as it presents less risk of unexpected toxicity or side effects in human trials, and, likely, less R&D spend.

An illustrative example is a recent R&D partnership between Sanofi and an emerging AI-driven biotechnology company Recursion Pharmaceuticals in 2016 with the purpose to identify new uses for Sanofi’s clinical stage molecules across dozens of genetic diseases. The Recursion’s approach is a “target-agnostic” one, and is based on cellular phenotyping via image analysis using computer vision.
Thousands of morphological measures are thus extracted at the level of individual cells and large catalogs of molecules are screened for the ability to “fix” phenotypic defects associated with each disease. Under the agreement, Sanofi will provide Recursion with a number of small molecules, and Recursion will screen them across its rapidly expanding library of genetic disease models and use machine learning technology to derive promising new indications.

Astellas Pharma Inc. signed a research deal with big data-driven bioinformatics company NuMedii to conduct drug repurposing projects using machine learning techniques. NuMedii’s big data resource includes hundreds of millions of human, biological, pharmacological and clinical data points, normalized and annotated. The company then uses neural network-based algorithms to find novel drug candidates, and biomarkers predictive of diseases, and repurpose existing drugs or drug candidates towards other medical indications.

6. AI for biomarkers development

Development of biomarkers is an important task not only in the context of medical diagnostics, but also in the frameworks of drug discovery and development programs. For example, predictive biomarkers are used to identify potential responders to a molecular targeted therapy before the drug is tested in humans.

One example of a company applying AI-driven modeling for biomarker development is Berg Health. Sanofi Pasteur, a world leader in the vaccine industry, announced last October that it will use Berg Health’s Interrogative Biology® platform and bAIcis® artificial intelligence tool to identify molecular signatures and potential biomarkers for assessing the Influenza vaccine immunological response.

Another notable example of applying AI for biomarker development is Insilico Medicine. In a recent paper published in Gerontology, they present a novel deep-learning based hematological human aging “clock” -- a biomarker for predicting biological age of individual patients. The biomarker model was “trained” using a large dataset of fully anonymized Canadian, South Korean and Eastern European blood test records. The company is also working on multiple other biomarkers using deep learning and incorporating blood biochemistry, transcriptomics, and imaging data.
7. AI for analyzing research literature, publications, and patents

Reading, clustering and interpreting large volumes of textual data is among the most developed use cases for AI-based algorithms. It come in handy for life sciences industry since the number of research publications in the field is growing enormously and it is hard for researchers to sift through vast amounts of data arriving on daily bases to validate or discard research hypotheses.

A research collaboration between pharmaceutical giant Pfizer and IBM’s Watson for Drug Discovery to tackle immuno-oncology was announced last year and became one of the most covered news story of AI application in biopharma sector. This collaboration was aimed to bring the power of AI-driven supercomputer for accelerating analysis and tests of hypotheses by researchers at Pfizer using “massive volumes of disparate data sources” that include more than 30 million sources of laboratory data reports as well as medical literature. There has been a wave of seemingly unsubstantiated criticism for the IBM Watson technology lately, so it is interesting to follow the future press updates about the practical outcomes of this collaboration.

A notable company using AI for scientific data mining, data contextualization and deriving hypotheses, is UK-based BenevolentAI. According to the company’s representative James Chandler: “A new scientific paper is published every 30 seconds and there are 10,000 updates to PubMed everyday’. Navigating through all this information to draw meaningful insights about drug candidates is where AI-based algorithms become indispensable -- this is the type of things BenevolentAI does.

In Nov 2016 BenevolentAI signed a license deal with Janssen Pharmaceuticals, a Johnson & Johnson company, for a sole right over a series of small molecule candidates and patents. The company plans to use its AI-driven platform to accelerate drug discovery and likely find new therapeutic indications for the selected small molecules. According to the license agreement, BenevolentAI can develop, manufacture and commercialize those novel drug candidates in all indications. Since a phase 2b trial has been recently launched by BenevolentAI for a drug candidate from the above partnership -- to treat sleepiness in people with Parkinson’s disease -- it seems like the deal with Jansses is already bringing returns.

Bio-Modeling Systems (BMSystems) is a French company, founded in 2004, developing Computer-Assisted Deductive Integration drug discovery platform (CADI)™ based on heuristic non-mathematical models to generate novel hypotheses from scientific, medical & health data. The platform includes several components for data acquisition & mining, data organization & structuring, an integrative engine, and model representation & visualization tool. According to the companies
presentation, it augments human scientist's capacity (Architect) with artificial intelligence capacity (Engineer). Over the years, the company had a number of undisclosed projects with pharmaceutical companies.

Conclusions: R&D outsourcing and M&A activity will be increasing in "AI for drug discovery" space

With an increasing interest in AI-driven technologies among the leading biopharmaceutical companies, a strategic focus of pharma and biotech businesses will be further shifting towards R&D outsourcing and M&A activity as means to quickly get access to the required expertise and know-hows. Complex nature of AI-based technologies, a need for costly and sophisticated IT infrastructure, a fast pace of progress in the field, and a relative scarcity of highly skilled data science specialists to support specialized machine learning research -- these are some of the key drivers of the ascending outsourcing trend.