Democratizing Artificial Intelligence For Pharmaceutical Research

Dec. 19, 2018 by Andrii Buvailo

Over the last five years the interest of pharmaceutical professionals towards machine learning (ML) and artificial intelligence (AI) has measurably increased -- while only one “AI-related” research collaboration involving “big pharma” appeared in the news in 2013, the number of such events increased up to 21 in 2017 alone, involving some of the top pharma players like GSK, Sanofi, Abbvie, Genentech, etc.

Needless to say that the application of AI for various pharmaceutical research tasks has become a widely discussed topic at practically all industry conferences and symposia. Besides, there is a long list of industry events specifically focused on AI in life sciences -- and they are crowded with top leaders from “big pharma”.

The key reason AI has taken pharmaceutical industry by storm is obvious: technology giants like Google, Apple, Amazon, Facebook, Microsoft, managed to demonstrate striking practical feasibility of the technology in the areas of natural language processing, text, image and video processing etc (one obvious practical example is a cohort of personal assistants --- Alexa, Siri, or Google Assistant -- you can try one of them right away and see how it blows your mind with what it can do). There is no better proof of concept than a demonstrated, measureable, practical result.

A lot has been said about how transformative AI is, and I believe biopharmaceutical leaders have successfully passed a stage of “do we actually need AI in our organization?” and are starting to usher into the next stage: “how to practically adopt AI at scale?”.

Tactics vs Strategy

While outsourcing AI-driven research from specialized vendors in a form of joint projects and research collaborations will remain a suitable mode of action, as a relatively simple and almost risk-free solution to try new things -- with little initial investment on the side of pharma most of the commitment lies on the shoulders of outsourcing partners -- it may only serve as a short-term tactical maneuvering.

Strategically-wise, pharmaceutical companies will inevitably have to focus (some already focused) on trying to build internal AI-powered research workflows and business processes and, ultimately, create own core know-how in what relates to applying AI to in-house research datasets.

I presume, as was with the digitalization of pharmaceutical industry over the last half a century (proliferation of personal computers and Internet, corporate digital systems (ERPs, CRMs, ELNs etc, cheminformatics and bioinformatics tools, e-commerce, etc), it is democratization of AI-technology and its components, that will be a key driver in the upcoming “AI-zation” of pharmaceutical research.

Making AI accessible to (bio)pharmaceutical researchers

In the context of technological progress, democratization is defined as making new tech accessible to the wider community of professionals, beyond just tech domain experts. The process of technology democratization is driven by the standardization of parts and modules, specialized tools, architectures, common platforms, user interfaces, designs, processes etc. It involves creating industrial standards for training personnel and crystallizing best use practises across various types of professionals.

Let’s summarize some of the current trends and available resources, facilitating the adoption of AI technologies by pharmaceutical organizations of various sizes:

1. Specialized hardware

Training complex machine learning (ML) models as well as using them in practice requires an unprecedented amount of computing power.

While most of the personal computing we do daily can be well served by general-purpose processors that can handle all kinds of tasks, the rise of machine learning, especially deep learning, pushed hardware companies towards building specialized chipsets that are custom built for ML tasks. With Google’s Tensor Processing Unit (TPU) and NVIDIA’s DGX-1, life science industry can enjoy powerful hardware built specifically for complex machine learning projects.

One illustrative example, although coming from the medical field, is a powerful NVIDIA DGX-1 AI supercomputer which has already been deployed in healthcare facilities like Massachusetts General

Hospital, to provide nearly one petaflops of processing power so that doctors can immediately perform comparison of a single patient's tests and medical history with billions of data points from a large population of other patients. Similar kind of implementation can be expected at R&D facilities of large drug discovery and development organizations, having substantial funds to allocate for infrastructure development.

2. Highly scalable cloud computing platforms

Going even further, large technology companies are increasingly offering public cloud services such as Amazon Web Services (AWS), Google Cloud Platform, and Microsoft Azure to name a few. Renting such services at a fraction of a cost of setting up own infrastructures, developers and researchers in drug discovery organizations obtain suitable and scalable resources optimized for building sophisticated ML projects in any area of pharmaceutical research. Such resources are easily affordable by smaller companies, startups and even research labs.

One illustrative example is a recent startups Bigfinite (raised $8.5 M in 2017), which used Amazon’s cloud infrastructure to build a number of AI-driven analytics components for various drug discovery and development tasks, meeting highest pharma compliance standards (GxPs).

Another example, although coming from the medical field, is the implementation of various ML-based models and scenarios in the cloud environment of Microsoft Azure Machine Learning. Professor James Thomas from the University College London is using Cortana Intelligence to quickly develop and deploy to the cloud various AI-driven tools. According to him, the deployment of his tools is as easy as pressing a button, without the need of setting up a server and configuring complex environments. After that, AI-based services become available to a wider community of colleagues and users as web-services via suitable and easy-to-use application programming interfaces (APIs). This illustrates how easy it is to build custom research workflows and analytics systems using ML-compatible public clouds.

Dell EMC has created pre-packaged Machine Learning and Deep Learning Ready Bundles, which essentially de-risk and simplify and accelerate AI/ML projects by pre-integrating all the necessary hardware and software.

3. Standardized deep-learning software frameworks

Currently, there exists a wide range of different software frameworks for building machine learning models -- with a wide range of specific pros and cons, such as compatibility, variety of required programming skills etc, which appears to be an issue in adopting AI by the industry. The cost of building and maintaining ML-based solutions will inevitably drop, and it will become accdessible to a wider community of pharmaceutical professionals, as the tools and platforms gradually standardize around a few dominant frameworks (similarly to the situation with mobile apps development, which became much less expensive as iOS and Android platforms emerged and become the two dominant ecosystems).

Currently, some of the notable open source frameworks for ML-based development include Google's TensorFlow, Amazon's MXNet, Facebook's Torch and Microsoft's Cognitive Toolkit etc.

4. User-friendly tools for ML/AI-developers

Another important step in adopting ML/AI by a wider audience of researchers and life science domain specialists is the availability of flexible and user/developer-friendly tools and platforms to build and customize AI-powered applications for various use cases. Let's take Microsoft Azure ML Studio as an example, which can be used by pharmaceutical professionals to access and develop quite sophisticated ML models though a simple graphical interface. Similar capabilities are being offered by Google and Amazon via their cloud platforms.

Another notable example is Cortex AI studio, a tool designed by Argodesign and CognitiveScale, which offers a straightforward and collaborative way of building AI-powered processes via drag-and-drop user interface.

5. Marketplaces for ML algorithms, open-source resources

While existing demand for the customizable ML-infrastructures is being covered by large technology providers, like Amazon, Google, and Microsoft, there is also a more specific growing demand for specialized libraries and algorithms for numerous small tasks. In fact, there exist marketplaces for the algorithms themselves. I did not find this kind of marketplace specifically for life sciences yet, but here is an interesting example of a general purpose marketplace for AI-based algorithms, Algorithmia, where you can find all sorts of ready-to-use algorithms, say, for face recognition, image processing etc.

6. Ready-to-use AI-powered web tools for life scientists

There is a growing number of ready-to-use AI-driven online tools designed specifically for life science tasks, which can provide substantial value for drug discovery researchers without the need for any advanced coding or ML/AI development skills.

So for example, researchers from the University of Waterloo recently launched Pattern to Knowledge (P2K) online platform, which can quickly predict interactions of biosequences, revealing intricate protein associations in complex environments.

Another useful and free online tool is IBM RXN for Chemistry, an AI-driven service for predicting chemical reactions.

Finally, there is BenchSci -- an AI-driven reagent intelligence platform which supports a powerful antibody search based on the text mining of the research literature (free for academic use).

Surely, there is a whole lot of other useful AI-based tools out there, I encourage you to mention them in the comments.

7. Access to quality big data in chemistry, biology, and medicine

I am listing this items last, but probably, it is the most important one on the list. Well-known “garbage in, garbage out” principle is one of the fundamental notions in machine learning and data science, therefore, having properly curated high quality research data of substantially large sizes is equally important for the successful implementation of AI-driven research processes as the AI-technology itself.

While most pharmaceutical companies have internal proprietary datasets which can’t be disclosed, or are too small and dispersed, there is a growing number of available large databases for training ML models to do various tasks. Here is an example list of 18 life science datasets, including general datasets, image tasks, genome datasets, hospital datasets, and cancer datasets. Chemical databases, suitable for machine learning projects include ChEMBL database (1.8 million molecules with known bioactivities, over 12.000 targets etc), ZINC15 database (over 100 million purchasable drug-like molecules for virtual screening, including), and REAL database (over 680 million synthetically accessible drug-like

compounds, all synthetic routes being validated in thousands of lab experiments).

Conclusion

Democratization of AI-tools, solutions and infrastructures is one of the key drivers of the transformation in the pharmaceutical and healthcare industries. This is largely facilitated by “big tech” corporations, like Google, Amazon, Tencent, Microsoft and others, which provide a lot of on-demand flexibility and resources. Another driver is an open-source community of volunteers and developers, contributing their expertise to developing specialized tools. And finally, there is a plethora of AI-driven startups, focusing on drug discovery and healthcare, who develop state-of-the art platforms and domain-specific solutions for licensing, or for project-based use.

It should be noted, that one of the important benefits of AI democratization is the “standardization” of a typical skill-set of a AI/ML-developer, operating in life sciences, which is important for the emergence of a high quality market of available talent.