Continuous Bioactivity Screening and Lead Optimisation in Accelerated Timelines: The Role of Real-Time AI Platforms in Modern Drug Discovery

Dr. Michel Beauregard, Associate Professor of Geomatics Engineering, Université Laval (Canada)

1. Introduction to AI in Drug Discovery

The pharmaceutical design lifecycle has begun to integrate artificial intelligence (AI) in order to aid the complex process of drug discovery. Historically speaking, choosing the appropriate drug constitutes the critical aspect of developing an appropriate pharmaceutical. As diseases have become more complex, the matter of choosing an appropriate therapeutic agent has become even more challenging. Traditionally, pharmaceutical industries have utilized a trial-and-error method coupled with high-throughput screenings to select an innovative compound. This strategy calls for a tremendous amount of time and money to be wasted, with some deaths related to pharmaceutical misuse presenting even after commercialization. Furthermore, the traditional methods have become even more expensive due to a high rate of failure during the approval process. AI provides an alternative technology combining computational power with advanced prediction algorithms. Although these algorithms were created approximately 20–30 years ago, only recently could we achieve the desired level of computational performance and dependencies necessary for their applications to be clinically useful. AI technologies have frequently evolved from simple decision tree methods to complex feedforward neural networks. Over time, AI has been integrated into different steps related to pharmaceutical development, leading to the creation of multi-part processes, including virtual screening, absorption, distribution, metabolism, excretion, and toxicity (ADMET)-based lead selection, medicinal chemistry in silico, as well as, in some cases, predictive toxicity analysis. New techniques such as reinforcement learning have enabled the production of de novo molecular generators, with in vitro biological testing undertaken. There is a growing shift toward the real-time

use of AI in drug discovery applications in response to the evolution and refinement of various tools.

1.1. Overview of Traditional Drug Discovery Processes

Conventional drug discovery processes for chemical drugs consist of a target identification phase, a hit discovery phase, a lead optimization phase, and clinical trials. These different phases can generally be thought of as progressing from highly speculative academic activities to more applied, industry-oriented work. As we progress through the pipeline, some potential drugs wither and die, but our confidence that the remaining will prove viable is expected to grow. However, historically, this relationship has not been quite as direct as the pipeline diagrams might suggest. For example, just because a chemical can serve as a lead, that does not necessarily imply it came from a good biological target, and vice versa. Yet in a process divided by phase, an unrelated secondary breakthrough might require the potential drug candidates to start up the network over, leading to historical inefficiencies in drug development processes and providing few backup options.

Currently, nominating a drug candidate occupies 5 to 10 or more years and costs in the neighborhood of $50 to $100 million. If one were to sweep the slate clean after phase II, perhaps the cost of guessing 500 patients wrong is something like $90 million. These are huge economic failures that guide our concern. The challenge in current drug discovery is often less the brute force of chemistry screening and more the identification of the needles that are both present and unexpected in the available chemical haystack. Scarcity can refer to uniqueness in some assessed way and is viewed as a strategic unique. The current rate of discovery is slow — only about 20 new drugs are approved for sale in a given year, and currently just a few percent of new drugs make it all the way through the testing phase. AI can perform the discovery process faster than chemists by rapidly testing all chemicals for some identified, useful endpoint.

1.2. Emergence and Evolution of AI in Drug Discovery

A Brief History of AI

Far from being entirely new to drug discovery practice, AI and informatics have a long-standing history in the pharmaceutical industry. General-purpose AI methods have been used as early as the 1970s for drug design and synthesis planning, often backed by

attempts at formalizing expert knowledge in chemistry. The most significant shift, however, only took place decades after these early pioneering works. In the early 2010s, data crucial for drug discovery—such as molecular structure and assay results—became abundant due to several technological advancements: an exponential growth in the availability of biological, chemical, and physical data due to high-throughput screening efforts; the increasing affordability of genomic and transcriptomic profiling due to faster sequencing methods and reduced sequencing costs; and the progress in electron microscopy and modern cryo-EM technology. These computational developments were accompanied by the long-term evolution of machine learning methods. The success of efficient non-linear dimension reduction techniques enabled complex, non-linear, and high-dimensional property relationship exploration typical for much of the omics data utilized in drug discovery. Changes in error quantification using specialized loss functions also paid dividends in specific fields.

AI in the Pharmaceutical Landscape

The acceptance of established methods such as deep learning and support vector machines in drug discovery was slow, with researchers often turning to well-accepted classical methods for practical applications in pharmaceutical pipelines. Out-of-the-box implementations of state-of-the-art AI methods often yielded unsatisfactory results, mostly due to the extraordinary complexity of the compounds considered. Demonstrable success cases are rare but increasing and center around specific proof-of-concept case studies. The lack of successes was not due to a fundamental failure of AI methods to identify lead compounds. Instead, it was due to insufficient generation of high-quality compound-related data. AI methods thrive on data, and true success stories in the field provide a clear indication of the positive impact brought by large amounts of high-quality data, proving that AI and big data have the power to revolutionize experimental projects.

Over time, AI has been raised from an experimental external tool to a method of essential competition for insight into drug development. AI performance is still relatively weak, with slow rates of full technological adoption, indicating how hard it is to substitute, aggregate, or hybridize existing experimental assays full of molecular features. Nonetheless, AI is a tangible investment for enabling high-throughput content and big data learning, the value of which is expected to incrementally revolutionize the

entire drug discovery process. In the future, AI can change drug discovery from cellular processes such as target function, target-disease linkage, and target safety to systemic solutions addressing the identification of new medicines and the implementation of new business models. This review focuses specifically on the application of analysis of molecular information such as structure-action, pharmacokinetics, and virtual screening.

2. Machine Learning Techniques in Drug Discovery

Machine learning and data-driven technologies play a critical role in modern drug discovery. These methods can be categorized into three types: supervised learning, unsupervised learning, and reinforcement learning. Through comparative analysis, supervised learning can predict the impact, outcome, context, or reaction of the treatment, while unsupervised learning can recognize the urgent or latent variables by analyzing patterns or clustering. Reinforcement learning is primarily applied to optimize decision problems, including multi-objective conflicts that are challenging to solve. However, the boundaries, overlaps, and exchange of these three methods are emerging in recent developments. It is clear that these machine learning techniques can be integrated directly or indirectly, enhancing the power of the AI approach toward enabling precision medicine.

There are two major challenges in choosing machine learning algorithms for developing the medicine we are going to apply. The first one is that there are usually only small numbers of labels because of the limitations of cost, time, and space for accuracy in most medical research. The second one is that there are always many biases in the given records of patients, which is very unauthentic. Supervised learning can answer the question of "what can this data do?", which is the most popular machine learning algorithm among these three types. Supervised learning is adept in predictive modeling for the outcome, interaction, or reaction of a drug. As for reinforcement learning, we can use it for the decision to optimize the dosage and formulation of the drug.

2.1. Supervised Learning for Predictive Modeling

Supervised Learning for Predictive Modeling: Historically, predictive modeling has played and continues to play a substantial role in optimizing drug discovery workflows. Supervised learning is a part of predictive modeling that involves the utilization of a dataset with pre-identified outcomes, hereby called a labeled dataset. These labeled datasets can vary from simple to complex and can encapsulate various biomedical,

genomic, chemical, and pharmacological domains. Typical applications of predictive modeling in drug discovery and development include the early stages of identifying specific potential drug candidates, predicting their biological activity, as well as toxicity, adverse events, and bodily distribution. Supervised learning for predictive modeling typically involves one or more modeling cycles for preprocessing the data to select relevant molecular or phenotypic features, transforming the data, training the model, and evaluating the performance. The quality of the model is often controlled by the feature selection method and preprocessing techniques to improve the prediction results given solely the information that is important. Supervised learning can be categorized into three main stages: regression, classification, and ensemble learning. Several performance measurements have been used to assess or rank the models; however, the current major challenge is to accurately evaluate the models' performance and identify the best set of predictive parameters. Several real-world examples, such as the identification of potential inhibitors, vaccine adverse event prediction, multi-target anti-cancer drug prediction, and pharmaceutical development, have resulted in successful model applications with ongoing clinical trials. Moreover, the predictive models are continually being refined to incorporate or exclude potential signals or targets due to off-target effects to better drug development.

2.2. Unsupervised Learning for Pattern Recognition

In the pharmaceutical field, unsupervised learning represents a critical tool to fill the gap due to missing data and to capture and characterize structures that could represent hidden or unexplored relationships. One of the main emerging applications of unsupervised learning in drug discovery is de novo drug design, which helps to identify novel drug compounds starting from a pool of imaginary molecules associated with different drug-likeness properties. This unsupervised learning problem is related to the identification of groups of small molecules with similar structures or properties. Clustering, k-means, or hierarchical clustering is consequently used to group together small molecules that are structurally similar, a strategy that is employed to infer possible and improbable drug candidates based on their bioactivity.

An additional unsupervised learning approach utilized by the drug discovery community is that of dimensionality reduction for feature extraction. Here, principal component analysis allows the reduction of the dimensionality of a dataset, keeping

relevant information, to visualize and explore drug candidates with similar mechanisms of action, adverse events, or other relevant characteristics. For example, PCA is used in combination with a clustering algorithm to create assignment maps that guide experts from various domains in annotating the targets, diseases, and phenotypes of interest. Consequently, the combination leads to the advancement of new methodology development in biomarker identification, for target and hit deconvolution, and drug repositioning. However, to induce experts in performing these annotation tasks, the unsupervised learning results need to be explored from an interpretability point of view. Thus, a human-in-the-loop framework is used to increase the opportunity to validate the results and uncover relationships within the biological system that otherwise would have remained hidden.

2.3. Reinforcement Learning for Decision Making

The use of reinforcement learning (RL) is growing in decision-making for drug discovery. RL is a learning framework where agents learn to make optimal sequences of decisions according to their experiences. The system to interact with, RL, consists of agents, environments, and rewards. Agents observe states of the environment and perform predictive actions, changing their states. Every action leads to a reward from the environment. They are continuously learning and interacting with the environment to receive cumulative rewards. In drug discovery, the environment refers to the collection of all the assays, and the agents are experimental design systems. The decisions are what to assay and in what proportion to match with the dynamic environment within the discovery theories and priorities.

RL can maneuver in dynamically active environments, which is best suited for drug discovery. It brings a comprehensive perspective in decision-making with real-time feedback from these experiments. It is proven for a wide range of applications in the discovery process, ranging from optimizing the formulation of drugs, clinical trials, and drug products to molecular synthesis. By learning recurrently from activities, RL can improve the policy or strategy by adapting to the environmental cues in every transitional state from day-to-day experiments. RL systematically refines the strategy by continuously factoring in the feedback from every action. At the same time, RL has some challenges. One of the fundamental challenges is the need for large data efficiency and computational resources. To date, the adoption of RL in drug discovery is less explored.

However, RL fosters not only structural growth in the discovery process but also a higher success rate in today's virtual, complicated world. Success stories include an innovation from the synthetic chemistry domain, where the use of RL catalyzed the optimization of test conditions for palladium-catalyzed C-H functionalization to receive 2.6 times the previously highest yield.

3. Real-Time Applications of AI in Drug Discovery

Real-time applications of artificial intelligence are beginning to transform drug discovery cycles. In high-throughput screening automation, time can be saved by screening libraries of billions of compounds computationally, looking for those that can interact with proteins implicated in disease. Predictive toxicology assessments can provide a real-time picture of potential efficacy and safety of compounds, predicting potentially dangerous effects in early stages such as cardiotoxicity and hepatotoxicity. Importantly, lead optimization and preclinical candidate assessments can also be accelerated. Techniques such as molecular docking can significantly limit the selection of a small number of potential compounds and those chemical modifications. The effect is to turn a 'hit' into 'lead' compounds, the initial building blocks of potential new drugs.

Currently, around 10,000 proteins have been validated as having links to disease. This presents an urgent need to identify more potential drug candidates, as currently only about 15% of these proteins are considered to be well 'druggable'. In order for these real-time AI processes to provide significant value, it is necessary to be able to deploy these deep learning algorithms at an extremely large scale, while ensuring high computational efficiency. The scalability and efficiency of these algorithms is particularly important for drug discovery and development – a cycle that currently takes an average of 10 to 15 years. In just the last 5 years, many companies have found ways to use AI to accelerate rounds of drug discovery.

3.1. High-Throughput Screening Automation

One of the first steps required to search for drug candidates, which exhibit a biological effect relevant for treating a specific disease, is considered to be high-throughput screening. This innovative method was introduced as a tool in the late 1980s and early 1990s. High-throughput screening enables scientists to run millions of potential molecular tests in less than a week. These tests can be used to identify the active compounds in the minimum amount of time. High-throughput screening is an

integration of chemistry and statistics with biology, which is widely accepted and performed in modern laboratories for drug discovery. High-throughput screening has been enhanced by robots and improvements in computer software. The 21st century can be described as biologically enlightened, where researchers and scientists are finding new ways of understanding things on a molecular scale. In the search for compounds that can be used in healthcare, high-throughput screening has paved the way for other screening technologies of relevance. Automation ensures that reliable data are produced time and time again, avoiding errors made by humans. The use of state-of-the-art robotics has introduced new levels of efficiency and precision into this drug discovery approach.

The adoption of high-throughput screening within the pharmaceutical industry has been outstanding. Over the past decade, there has been a revolution in robotics in the field of automation. This is well known since the last fifty years, three stages in robots' automation technology have been observed. High-throughput screening workflows work like this: a filter is set with numerous chemical ingredients or combinations of ingredients that may bind to a target. The goal is to identify the targets. Robots drop dyes, one per tiny saucepan, onto layers of cells growing in tiny saucepans. The dyes fluoresce in the presence of the target molecule. The robot measures how bright the cells become; only plucking those little pots that really light up. Each drug company has its own chemicals repository and search engine, which are highly guarded trade secrets. New drugs take a long time to develop and a lot of money. Not every promising compound resulting from high-throughput screening assays in a drug candidate can be due to safety and efficacy issues. Besides this drawback, new compounds usually have fewer side effects than a drug used. In addition, the human body frequently deactivates medication or its effectiveness weakens.

3.2. Predictive Toxicology Assessment

A significant area where AI systems can have a big impact on the drug discovery process, especially in real time, is in the prediction of the potential toxicity of drug candidates. Understanding the potential safety risks that may be associated with new drug candidates is essential to prevent petering out partway through the clinical trial process due to concerns, particularly regarding potential long-term safety issues. Predicting the risk of toxicity more closely links the identified biology with the

development candidate, as the eventual toxicity observed in the clinic is due to the adductive reactive metabolite binding to the target. Modellers in predictive toxicology may consider – within the in silico domain – both the chemical-related and the biologically related activities of the reactive metabolites as part of the risk assessment.

Several in silico approaches are now being used to predict the concentrations of molecules and their metabolites resulting from following different routes of entry. These models could also provide information on where different molecules are mostly accumulating, providing insights in human health studies. This can only be done if the model has been developed with broad applicability. Historically, in silico approaches have been developed to define structure-activity relationships. The structure-activity relationship is defined as a relationship between the chemical structures of compounds and the biological activities of the corresponding compounds. Quantitative structure-activity relationship models used for the prediction of human intestinal absorption of drugs are regarded clearly by many regulatory agencies for use in drug submission documents. When determined to be applicable, the models are very efficient in predicting whether there is a risk of oral absorption.

Data-driven approaches utilizing large data sets are increasingly used to develop predictive models for toxicological endpoints. Up to date, the primary route for such an in silico approach, as with many other areas of use, is for the screening of drug candidates rather than for regulatory assessment and decision-making. Given the complexity of utilizing such approaches, utilizing sex and age in the models is thus far from use in the clinical environment.

In conclusion, applying a range of in vitro approaches lends itself to AIs. Advances in toxicogenomics, toxicophenomics, and high-content screening techniques pursuant to AI are ideal for developing models that can support the prediction of adverse outcomes and may eliminate the need for species extrapolation. Recently, there has been considerable progress in computational models to assess the toxicity of a drug compound to skin, liver, and kidney. In addition, these human in vitro data can be integrated for pharmacokinetic modeling to obtain doses for an analysis of mechanisms of action and key events. The data available on skin decomposition in an in vitro battery, the design of read-across approaches, and the integration towards biological output data is rapidly expanding.

Estimates of the low cost of basic in vitro toxicity assays for skin and anatomic characteristics to transfer in vitro toxicity into real-world human exposures show that the models can model compound relationships from structural and computational descriptors. Predictive toxicology assessment is a critical strategic requirement to facilitate the transition of drug candidates from toxicology species into their first-in-man clinical trial. Initiating clinical trials in human volunteers is very expensive, and candidate failure at this stage is far from ideal, both in terms of wastage of money, intellectual effort, and sacrifice of volunteers. By assessing chemical structures and biological activities in real time, the risk can be mitigated, the dose levels can be set, and the risks to the potential study participants can be deemed acceptable.

3.3. Lead Optimization and Molecular Docking

Lead optimization, the reiterative process of improving drug candidates, is critical for successful rational drug design. Lead optimization aims to increase the potency, selectivity, and physicochemical and pharmacokinetic properties of each lead compound to develop new small molecule drugs. Through physicochemical optimization and bioactivity optimization, a compound enters the drug development cycle when it sufficiently meets efficacy, selectivity, and safety requirements. Since high-throughput screening identifies tens of thousands of hits that must undergo lead optimization, a significant pipeline bottleneck is the lead optimization phase. Molecular docking plays a critical role in this direction by simulating the molecular interaction between a drug-like molecule and protein, RNA, or DNA receptors to predict binding affinities between small and large molecules with their biological target, as well as visualization for molecular interaction studies. For example, after eight months of large-scale screening and more than 20 compounds synthesized and evaluated using a nanomolar ATP site with human kinase PIM1, recommended compound lists with target validation data were obtained within days. Similarly, molecular docking was used to optimize the lead with a molecule of bactericidal topamine and gained access to the high micromolar gyroxipen at the predicted Kd of the co-crystallized solution within two weeks of the descriptive insight. Recent advances in computational optimization and molecular docking tools have made it possible to improve current drug candidates, both through in-depth structural studies of reported inhibitors and relatively rapid virtual high-throughput assessment of large compound libraries generated from larger diversity sets.

A similar reinforced virtual discovery hybrid approach is used to improve leads with high nanomolar Kd on kinase and email association.

4. Challenges and Limitations of Real-Time AI in Drug Discovery

Real-time AI in drug discovery holds enormous promise, as it enables timely predictions, design, and synthesis of molecules. However, several challenges and limitations hinder its industrial-scale implementation. A fundamental barrier is the quality and availability of data. Suboptimal or biased datasets can limit and potentially hamper the predictive power of AI models. Biased models might reduce the diversity of compound libraries and, overall, impair the robustness of the identified hits and/or leads. Furthermore, the development and application of AI models in general are obstructed by concerns over their interpretability, transparency, and black box nature. Industry stakeholders must be confident that the models are able to provide meaningful explanations for their predictions or classifications. The existence of biases may limit this interpretability.

Moreover, AI algorithms should be robust, secure, and protected against takeover and malicious intent. For example, biased AI could easily be exploited by hackers to overthrow ML models, presenting a security risk. A fundamental challenge is related to concerns about the deployment of biased AI models in decision-making. These concerns have implications ranging from fairness and anti-discrimination perspectives to ethical considerations. AI methodologies should be developed in a transparent state. Differences in model performance and associated inputs for different patient populations or subgroups will have to be determined and validated. This is essential because it has been noted that an algorithm trained on a specific group as the population may not work acceptably on a group from another population. There are further concerns regarding regulation, as the rapid evolution of AI methods may outpace the adaptation of current guidelines and frameworks that inform drug development. These limitations lead to a lower likelihood of time and cost savings and entail a higher attrition rate in development. Consequently, the implementation of real-time AI in drug discovery can misguide and delay the synthesis and testing of promising drug candidates, thereby leading to an overall increase in development time and cost and a reduction in the success rate of new drug development. Regulatory bodies should outline guidelines to address the issues and incorporate real-time AI into preclinical and

clinical development. Public-private partnerships may be important in addressing these issues to enhance the likelihood that real-time AI methods will be accepted.

4.1. Data Quality and Availability

Quality data is the cornerstone of any successful application of AI to drug discovery. Quality implies that the data is complete and annotated with the required features. Challenges in data quality occur across data types and include datasets that are either incomplete due to missing values or of low quality due to inconsistent or out-of-range values. When using them to train classification or regression models, incomplete or inconsistent data causes issues such as variation outside the determined input range. In addition, biased datasets are common. In particular, public datasets often contain only successes, while commercial databases may be heavily skewed toward a commercial advantage. Another vital aspect for model robustness, transferability, and the generalization of the trained model that is frequently overlooked is the availability of diverse datasets.

Data has been siloed in research publications, stored across lab sample fridges, transcribed in scanned PDFs or non-standard lab logs. Two general strategies are proposed for the management of the data quality problem. The first approach is through the use of formalized data curation and validation techniques, and the second involves collaboration between data creators worldwide to develop data-specific standards with the explicit purpose of guaranteeing sufficient and publicly accessible data. The leadership here often comes from regulatory initiatives seeking to standardize clinical or safety-related data in order to predict health liabilities of potential therapeutics. Initiatives to access large-scale sequencing and other biology-relevant data, combined with crowdsourcing to expand the numbers and types of compounds tested in these analyses, have a substantial impact on the use of the resultant data to build improved AI in drug development. Finally, data-driven methods have high filtering capabilities that are directly dependent on the quality and appropriateness of the input data. Major issues in any of these data aspects are likely to prevent the opportunity of a foray into the uncharted chemical matter and thus affect compound success rates across drug development.

4.2. Interpretability and Bias in AI Models

Interpretability, or the ability to simplify an AI model's decisions, is becoming an increasingly important issue. No matter how powerful algorithms become, the view into the decision-making process remains a crucial aspect of any AI application. Especially with regard to financially risky activities like drug discovery, stakeholders must be able to understand how a model has arrived at a certain output or decision. Additionally, for AI to be accepted as a tool for decision support in the regulated environment of the life sciences, the models must be interpretable to be useful. Regulators desire transparency and need to understand how a given decision was arrived at before approving a new drug. Regulators also require clear documentation of any model development activities.

Secondly, fairness and bias are hot topics in the field of drug discovery – and artificial intelligence in general – for a reason: the risks are real. "Black box" models are especially prone to biases because it is difficult to ascertain how a model has weighed and prioritized the myriad of features that contribute to its outputs. Several unsettling cases of AI models perpetrating discriminatory outcomes have already surfaced. I will briefly discuss two examples in the drug development space and elaborate on the strategies available to mitigate this risk. The case of a Texas woman whose cancer went undiagnosed by a prognosis algorithm because she was male is an example of the pitfalls of using flawed data to train AI algorithms. Pharmaco-biased AI is a relatively new area of study. Some pharmaceutical companies are hesitant to adopt machine learning approaches for this reason, despite the parallel popularity. Regulatory agencies are also increasingly concerned with the potential for pharma-biased models to become a reality. There have been arguments that it is time for regulation of AI applications to protect users against biased AI algorithms. In the pharmaceutical arena, the use of AI tools will not be company policy until policies of that nature are established in line with existing regulations. Once in place, these data protection regulations will require pharmaceutical companies to adopt AI models in line with the principles of privacy and accountability, that is, with consideration for potential biases at all stages of the algorithm's lifespan. These policy measures are expected to consist of, among other things, recommendations for the internal design of an AI model so that its outputs are clear, logically explained, and "transparent" to the user. They will also provide guidance to users on how they can hold the AI algorithm accountable if it makes a mistake or violates a right to legal protection.

4.3. Regulatory and Ethical Considerations

As with any rapidly advancing technology, the development of AI presents a significant challenge when it comes to regulatory frameworks and ethical guidelines. There are currently no specific guidelines, approval standards, or even criteria to evaluate the safety and effectiveness of the AI-based methodologies used in drug discovery. Unlike drugs, the use of deep learning methods for drug discovery is not developed as proprietary technology, where elements and component parts could be kept confidential. Methods provide more than just insights into how something is built; they provide a fundamental blueprint of how experts in drug discovery with limited data could approach similar problems. Thereby, the tool developed has been open and transparent. However, the utilization of these strategies has its flipside: as a method becomes more transparent, it will likely be scrutinized to a greater extent from a regulatory standpoint by the scientific, pharmaceutical, and health sectors.

There are myriad ethical issues that need to be addressed when considering the use of AI for drug discovery. These include complex issues surrounding patient privacy, informed consent for the data generated and used by predictive algorithms to facilitate drug discovery, normalizing the value of public health data, and the public ownership of the data through the use of AI systems to democratize drug discovery. For the greatest public benefit and to create public trust, policy must be informed by perspectives from across the spectrum of interests and disciplines. Stakeholders are essential to work together to understand the implications of AI in drug discovery to maximize public good. Additionally, incentivizing ethical AI for rewarding strategies that are accountable for being ethical in a transparent manner, and for supporting efforts that educate developers and business ethicists about ethical considerations in AI development, especially at the phases related to drug discovery. AI investments result in substantial costs and risks, as well as potential returns. As of yet, there are no approved drugs developed using AI, and in the absence of appropriate safety, efficacy, and ethical guidelines, it is unclear whether AI methods would encourage the approval of drugs that are less safe or are overly effective. Hence, there is a need to identify the regulatory and ethical implications of these technologies.

5. Future Directions and Opportunities

There are a number of exciting future directions and opportunities for real-time AI-powered drug discovery. The integration of multi-omics data could bring a more network-based, systems-level approach to drug development. Through the use of graph embedding approaches, which transform complicated heterogeneous graph-structured data into simpler and more manageable vector representations, networks can express diverse interactions between drugs, genes, and diseases. Also, a more in-depth understanding might emerge from an increased knowledge of complex biological systems; such a model could report new genes implicated in diseases that are beginning to be dissected based on specific molecular knowledge. Being able to highlight these new genes might, in turn, become a highly valuable avenue for the identification of new drug targets. At the same time, personalized medicine might become a pioneering approach for rare disease research and be developed on the strength of specific rare-subtype transcriptomics and genetics. Recent advances are just the beginning, and by adding other technologies, including 3D miniaturized organoids and primary patient-derived cells, these systems can increasingly come to engage the highly complex and personalized environment of a specific rare disease patient. Despite these exciting avenues, there are a number of challenges that need to be addressed. There is considerable confusion over who is responsible for proving efficacy in novel drug development; is it academic institutions and government-funded projects or the industry? In addition, gaps between academia and industry have not diminished. Academia is seen as a place for investment in initial idea development which, if proven, is taken over by industry development. Also, the high failure rate in drug development raises the question of who should be challenged to ensure the higher probability of translating an AI-generated hypothesis to novel drugs. A potential path to the clinical challenge could be a strong collaboration between academia and industry that supports the very beginning of AI-driven drug discovery development, linking AI approaches to drug repurposing strategies. The application of AI approaches to generating new drug indications has become a burgeoning research field and can be a strong initial challenge for the open drug discovery concept.

5.1. Integration of Multi-Omics Data for Drug Discovery

The integration of various omics data is a transformative research approach in the effort to discover new drugs. The term multi-omics broadly encompasses genomics,

exposomics, transcriptomics, proteomics, and metabolomics among many others, and can provide a comprehensive view of complex biological systems. Each omics approach answers specific questions related to genetic variants, RNA expression, proteins, and metabolites respectively, and comes with specific advantages and drawbacks. By combining diverse data types with increasingly sophisticated AI technologies, researchers can not only find a better target but also a more robust biomarker of the disease state or of drug response. Moreover, multi-omics also unveils emergent, non-linear connections between genes and phenotypes, and therefore has the potential to investigate and reveal synergistic biological effects in perturbation experiments. This new toolbox is now part of research labs, and its application to the development of new drugs is not a distant hope but a reality. There is a growing number of scientific papers showing the success of these approaches as well as case studies in the frame of drug development. Although the potential of multi-omics in drug discovery looks promising, the translation of these results into translational applications that can bring effective new therapies to patients will depend on the harmonization of omics data and its compatibility. Large-scale integrative studies have often encountered challenges related to the disparity in the types of sequencing data or read outputs, the technology platforms, and experimental protocols that generate them, and the file formats and metadata standards used for their submission to repositories. Overcoming these challenges requires a significant amount of development work by the scientific community in conjunction with repositories and literature publishers to define and implement standards for data generation, sharing, and interpretation, as well as solutions for technical interoperability. Applications related to precision medicine, grouping populations or patients by different biological profiles to increase understanding of therapies, and thus allowing treatment regimens to be better tailored, are already being done at a research or clinical research level and have the potential to be translated into clinical practice.

5.2. Enhanced Collaboration between Academia and Industry

The benefits of enhanced collaboration between industry and academia have been recently reviewed. Suffice it to say that the sum of the consortium between academia and industry is greater than its parts, with benefits including better engagement with end users and consumers, exchanging knowledge, resources, and expertise, and accelerating the translational pipeline. Co-investment and co-creation of new knowledge

and tools by these partners should assist in the earlier translation of research findings into clinical applications. The wealth of diverse ideas generated in academia is well positioned for translation, while industry is focused on bringing a medicine to market and providing therapeutic and/or prophylactic options for global use. Industry and academia can also serve important roles in highlighting unmet clinical needs, thus steering research towards areas of high priority.

There are many examples of successful collaboration between industry and academia. In regenerative medicine, we already see successful examples where world-class expertise at academic institutions has been used by industry in partnership. The development of novel breast cancer treatments is an excellent example of how academic expertise was brought together with investment and entrepreneurship in a collaborative environment in this field. An example of a successful public-private-academic partnership is the establishment of a lab at a university, initially by a research council and a pharmaceutical company, followed by close working with other organizations. This has spawned not only new products now growing in the market but also successful spin-outs translating technology to new targets. A number of disease foundations encompassing a group of stakeholders, including patients and funders, are working to fund and develop drugs, and therefore creating strategic partnerships with large pharmaceutical organizations in specific diseases.

Challenges and Considerations However, there are barriers, not least the priority of differing value systems in academia and industry and likely differences in the planned impact of research. There can also be a mismatch in IP requirements, with universities developing patents and licensing technology, often where the university will hold a significant upfront share with the academics. This approach is often not palatable to industry. In an almost self-fulfilling prophecy, there might also be issues perceived by industry in that academics are not ready to collaborate and develop commercial research, with only a portion of industry respondents believing that the pace of academic knowledge transfer was either getting better or had remained the same in the past few years. Most of the respondents also reported that the universities are not good at adapting to evolving industry models. A review of approved therapeutics shows that a significant number of agents approved were based upon drug targets first identified from publicly funded research institutions. This study also showed that, of this group,

small molecule drugs were approved at a rate of over 50%, but the translation of research to drug delivery was still a lengthy, complex, and costly process. Key here then is promoting and nurturing a culture of collaboration within research, on initiatives, projects, and networks. There need to be incentives for researchers in institutions to work directly in partnership with organizations, and the skills needed for this 'real-world' research should be taught to students in a higher education environment. Identifying possible ways to bridge the gap between academia and industry can only lead to enhanced drug delivery and development in regenerative medicine.

5.3. Personalized Medicine and Targeted Therapies

In recent years, the field of cancer genomics in particular has made significant progress in characterizing the genetic alterations that drive different types of cancers. Innovations in the area of diagnostics are now making it possible to sequence the genome not only of the tumor but also of the patient. This progress, together with large-scale efforts to sequence highly diverse patient populations, is providing new insights into the role of additional factors that contribute to disease, such as the genetic composition of the tumor microenvironment, the individual makeup of the host that can profoundly influence the response to drugs, and environmental and lifestyle parameters. Using this information, it is now possible to design precision medicine strategies and formulate personalized treatment plans tailored to these factors and the individual patient profile, where drugs and/or combinations are specifically chosen to best suit a given care pathway for that specific patient.

Several approaches are being developed that leverage different types of AI to integrate the above-mentioned omics data. Implementing such a multi-platform biomarker and drug-based approach would have a major impact on patient outcomes. Two of the most significant opportunities of precision medicine in the healthcare industry are its potential to improve treatment paradigms and drug discovery, either through the development of better tailored drugs or a more comprehensive understanding of the basic biology of the target patient population. Although the promise of precision medicine is clear, major hurdles for implementation still exist: since patients may be identified with rare genetic variants, treatments may be rare and costly; studies will require a large amount of molecular as well as clinical data to reach statistical significance; ethical and regulatory issues will face the large-scale acquisition of

molecular data of human subjects. Overall, the development of precision and genomic medicine for stratification and treatment of patients presents modern biology and drug discovery with unprecedented challenges, trials, and potential opportunities poised to revolutionize healthcare.

6. Conclusion

This package of essays has demonstrated the significance of the development of real-time AI and the potential impact on drug discovery operations and new medicine development. AI provides a powerful ability to enhance productivity, quality, and rationality across the drug development process. AI's potential is heavily dependent on data, and the contributions have made it clear that existing barriers are historical, real, and will prevent society from gaining from rapidly advancing technologies. As such, any further development of systems must be preceded by a greater understanding of the value and limitations of AI systems, the underlying science including clarification of the algorithms, and opportunities for validation across multiple modes. Most importantly, the capabilities are only as useful as the questions we ask of them, and systems designed to provide answers must be based on a detailed understanding of the complex questions and stakeholder priorities.

Further research to expand on the limitations and challenges discussed is essential, as is more extensive partnership with a range of different disciplines and stakeholders to address these limitations. These packages of essays set out an agenda for immediate action in these and other areas. In terms of technological development, personalization of AI applications is potentially well developed, but we lack interoperable systems equipped to deliver on personalization with the big-solution benefits. Similarly, with multi-omics unity, existing systems and the state-of-the-art applications are still far away from the promises of the birth and justification for the large bespoke investments. Interoperable systems are to come, and further work is needed also beyond the current state-of-the-art. As these different threads of work develop and are applied to industrialized drug development, we look forward to the next steps of the revolution in pharmaceuticals.