The Impact of Artificial Intelligence Tools in Pharmaceutical Research

Business Strategy, Ethics, and Society

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia – Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Brendan Cox

Spring 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

William Stafford, Department of Engineering and Society

Introduction

Drug development is dominated by large, multinational pharmaceutical companies (*How AI Reduces the Cost and Time of Drug Discovery and Development – NaturalAntibody*, n.d.). Commercializing a novel pharmaceutical product requires an enormous upfront investment that is prohibitively high for many small, under-resourced new entrants into the pharma space. A benchmark study conducted in 2014 asserts that the cost of bringing a new drug to market has nearly doubled since the early 2000s, rising to an astonishing $2.5 billion (News, n.d.). Beyond the immediate financial burden, drug development is also an extremely risky and time-consuming venture. It can take decades for a product to go from lab to market and even then nearly 90% of drug candidates are not granted FDA approval (*How AI Reduces the Cost and Time of Drug Discovery and Development – NaturalAntibody*, n.d.).

Successfully navigating these challenges can still reap great financial rewards, which is the long-term strategy of successful pharma giants. AbbVie's arthritis drug Humira, for instance, saw sales revenue of nearly $21 B in 2021 (*The Top 20 Drugs by 2021 Sales*, n.d., p. 2). These "blockbuster" drugs and the immense revenues they return are a cornerstone of the modern pharmaceutical industry. The reliance on patenting blockbusters to sustain profitable operations creates a demanding environment for new entrants and a lack of healthy competition in the pharmaceutical industry. Novel drug development is difficult to pursue outside of large pharmaceutical companies that can endure significant costs, risk, and time because of the immensely profitable existing drugs in their portfolio.

Pharma's barriers to entry, however, do face a new challenge in the 21st century: artificial intelligence. These algorithms can parse immense troves of data, including information on chemical compounds and structure, animal models, and patient details. AI tools can not only help identify potential physiologic targets for future drugs, but even what molecule might be best suited for such a treatment (*Measuring the Risks and Rewards of Drug Development*, 2018). Recent use cases of AI have as much as halved development and discovery time by facilitating entirely *in silico* research, where pharmaceutical products are investigated through computational models that are much more efficient than traditional physical tests (Masson, 2022). Faster and cheaper drug development has the potential to dramatically alter standard industry practices. This report will characterize the impact AI tools may have on accountability, monopolization, and business ethics in the pharmaceutical industry.

Literature

Relevant technologies to the application of AI in drug development can be divided into two broad categories: predictive and application based. Predictive technology refers to the use of AI before any drug development occurs. Among others, predictive technology includes data mining, virtual drug screening, generative chemistry, pharmacogenomics, toxicity predictions, and drug-target predictions (*AI Is about to Remake the Pharmaceutical Industry – POLITICO*, n.d.). Application-based AI technology refers to uses after a drug has begun development or testing. This includes clinical trial optimization, clinical data analysis, pharmacovigilance, personalized medicine, and applications related to automating laboratory procedures or manufacturing (*AI Is about to Remake the*

Pharmaceutical Industry – POLITICO, n.d.) (Vanhaelen et al., 2020). In either case, the widespread adoption of AI-enabled processes in pharma will change its competitive landscape, impacting the established oligopoly and the behavior of pharma giants.

One of the most powerful tools employed by pharma companies in the realm of predictive AI is called Computer-assisted Drug Design (CADD). CADD, as an umbrella term that includes multiple machine learning (ML) or artificial intelligence (AI) models, allows researchers to create an artificial, computational representation of a new drug and model its properties. This can provide information not only about molecular properties such as selectivity, distribution, absorption, bioactivity, side effects, and excretion, but can help researchers optimize drug attributes entirely *in silico*.(Vanhaelen et al., 2020)

CADD methods interact with vast troves of data to achieve their predictive power. Some of the most crucial databases used for CADD are called drug-protein interaction databases (DPIs). These contain information about the known relationships between existing drugs, their chemical structure, and the molecular mechanisms through which they interact with a patient's proteins. Protein-protein interactions (PPIs) databases are another common class of database with a similar function. PPIs focus on the relationships between various known proteins, rather than on biomolecules in general. Semi-supervised machine learning techniques have been successfully applied to both types of databases to predict new drug interactions and potential side effects without ever physically testing a new pharmaceutical(Vanhaelen et al., 2020). Outside of DPIs and PPIs, pharma AI is trained on available patents, publications, clinical trials, research grants and a score of other sources of pertinent information. The large protein databases, however, are a

convenient place to dive a little deeper into what utilizing biologic datasets looks like for semi-supervised learning.

There are hundreds if not thousands of protein databases, but the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) is a good demonstration of their scope and complexity. Containing over a terabyte of structural data for proteins, DNA, and RNA, RCSB PDB is a critical source for AI/ML development. With this data, companies or researchers working to develop a model for drug discovery would apply a classification algorithm along with a labeling strategy to create a model capable of predicting protein targets. The labeling strategy essentially classifies a protein based on its potential as a target, which is information that may be derived directly from the database or found in wider literature.

Deep learning techniques such as generative adversarial networks (GANs) are a new approach that goes beyond semi-supervised learning into an area referred to as AI imagination. Here the power of CADD becomes truly astounding. GANs are able to generate new objects with desired properties and have been used in several cases to generate patentable *de novo* drug candidates with clinical promise (Dara et al., 2022). These techniques differ from previous approaches which rely on explicit chemical knowledge being built into the software or model. Deep learning approaches are able to derive their own interpretations and insights(Dara et al., 2022). This makes them very powerful tools for designing original products. Other deep learning architectures such as recurrent neural networks (RNNs), variational autoencoders (VAEs), and adversarial encoders (AAEs) have been used in addition to GANs for *de novo* drug development.

Using these advanced models, researchers can more quickly process potential drug candidates, identify physiological drug targets more easily, and investigate relevant molecule characteristics like toxicity or bioactivity completely *in silico*. By replacing physical experiments with digital and AI-driven tools, researchers can discover and investigate drug compounds in a more efficient, cost-effective way. Pharma.AI, a platform designed by In Silico Medicine to support *silico* drug discovery, is a great example of the end product that can be created using datasets like RCSB PDB. The Case section will dive further into the impacts of this platform, but it represents the culmination of several technological stages: successfully aggregating and using big data, implementing neural network models, and validating those datasets and models, among others. However, developing standards for the validation of *in silico* drug candidates is a significant challenge. Without a standardized, trusted, and efficient way to benchmark or compare generative chemistry models, it will be hard to fully utilize these predictive tools.

Application based AI approaches are somewhat less technically complex, mostly providing value in the form of data integration. A good example of the power of application-based AI tools is in the design and implementation of clinical trials. The standard format of a clinical trial does not always facilitate optimal research where heterogenous patient groups, complex therapies, and difficult patient retention or monitoring contribute to a high rate of failure. Clinical trials often span across multiple clinical locations and manufacturing sites, require near constant communication with patients, and generally involve an immense data management workload. AI can mitigate some of these common problems by seamlessly integrating data collection, management, storage and even trial

design itself (*Intelligent Clinical Trials*, n.d.). The utility of AI in the design and conduct of clinical trials stems from its ability to process large amounts of data and integrate that information across locations or databases. Analyzing and extracting patterns from current and past clinical trials, post-market surveillance, and other programs or research could relieve significant burden on the company responsible for bringing a drug to market.

Another important area is in pharmaceutical manufacturing. AI can help use historic data to predict maintenance needs on the manufacturing line, supply chain issues, and inventory forecasting to help optimize production rates. In these and other ways, it can be an indispensable tool in operations management.

Whether used as a predictive tool for the discovery of new drug candidates or an aid in clinical trial data management, AI will change the way pharma companies compete. A cheaper and faster path from lab to market may enable smaller companies to successfully develop new products outside the umbrella of big pharma. A company's competitive capability has been modeled through three intrinsic factors: intangible assets, personnel, and tangible assets. Tangible assets refer to things like data, computer infrastructure, and algorithms (University of Connecticut et al., 2021). These are assets that a firm owns that enable it to conduct useful activities in the AI field. Personnel relates directly to the technical knowledge of a company's employees. For AI-driven drug research, there is an important balance between computational and scientific talent that maximizes AI-driven research potential. Intangible assets are a catch-all for any other factors that may be relevant to a firm's success in utilizing AI platforms. It can include anything from company culture, strategic aims, or other knowledge assets (University of Connecticut et al., 2021).

These criteria will be used in predicting the effects of AI on the competitive landscape of the pharmaceutical industry, specifically on the potential for new entrants to compete against established giants.

Case- In Silico Medicine

In Silico Medicine, a leading company in computational technology for drug development, offers a salient demonstration of the power of AI in pharma with their drug candidate INS018 055. This compound, intended to treat idiopathic pulmonary fibrosis, is the first fully AI generated drug to reach Stage II clinical trials– where its efficacy is tested on humans (Philippidis, 2023).

The repercussions of In Silico Medicine's success are enormous. While there are many other AI-driven drug candidates currently in the approval pipeline, INS018 055 is the only candidate entirely AI-generated and with an AI-discovered biological target (Philippidis, 2023) In Silico Medicine was able to bring INS018 055 from concept to clinical testing with unprecedented speed, low cost, and precision. It is a proof of concept for AI-driven research that will validate a distinctly new way of approaching drug development.

In Silico Medicine started developing INS018 055 in February 2021 on their Pharma.AI platform, which incorporates several software tools developed by In Silico Medicine. These tools include PandaOmics, which identifies potential drug targets for a given disease, and a tool for automated de-novo drug design called Chemistry42. Chemistry42 utilizes generative AI technology, as discussed in the literature, to create new drug-like molecules with specific properties (Medicine, 2023). These two platforms are the core software tools

that In Silico Medicine is hoping to validate with the successful approval and use of INS018 055 in treating IPF.

The speed with which these tools have enabled In Silico Medicine to bring forward viable drug candidates cannot be understated– it is this feature above all else that makes the inauguration of technologies like these so transformative. Chemistry42 enables researchers at In Silico Medicine to find novel molecules in as little as a week (Medicine, 2023). Their IPF drug candidate, for which work only began in 2021, is expected to be brought to market within the next few years. Through traditional methods, bringing a novel drug from discovery to market takes well over a decade on average (Amoranitis, 2023). The period from discovery to clinical trials averages 4-7 years on its own, whereas INS018 055 completed Phase I clinical trials within 2 years of the drug's discovery (Medicine, 2023). Despite the rate with which INS018 055 has navigated early research and clinical trials, the most important area of improved efficiency cannot be so easily measured. These technologies are expected to dramatically increase the rate of drug discovery itself, not just the speed at which discovered drugs are brought to market. Only time will be able to illuminate the magnitude of AI's impact in that regard, but this is where the true power of AI-driven research lies.

Lower costs go hand-in-hand with the rapid new pace AI-driven drug research has set for the field, with the overall cost of developing INS018 055 a meagre $1.8M (Field, 2023). To put that figure in context, anticipated costs for Phase I clinical trials (which INS018 055 has already completed) are roughly $14M on average (Zhou et al., 2023). That number does not account for the progress INS018 055 has made in Phase II trials or the often significant

initial costs incurred in bringing a drug from concept to clinical testing. There is a new cost burden for companies moving to AI in maintaining, licensing, or developing the AI tools that enable this extraordinary expense reduction, but the net effect will be an overhaul of traditional barriers that contribute to big pharma's monopolization of novel drug commercialization.

While this case study is focused on INS018 55 and In Silico Medicine, it is important to recognize that this company is only one part of a rapidly growing, cutting-edge industry. Today, the market for AI in pharmaceutical research is estimated at $1.4B but is expected to more than double to $3.7B by 2027 (*AI In Drug Discovery And Development — Will It Live Up To The Hype?*, n.d.). Companies at the forefront of this emerging field have an enormous opportunity for significant market capture and profit. There are several competing firms in this space, some of which have their own AI-enabled drugs in the approval pipeline. Although by no means a complete list, notable players include AtomWise, Exscientia, Cradle, and Iktos(Shah-Neville, 2024). The complexity of this large, competitive field needs only to be noted for the purpose of this analysis. A focus on In Silico Medicine and its products serves as a simple, effective framework from which to analyze the wider social and economic consequences of the technology.

Moreover, In Silico Medicine and its competitor companies are likely not to be the primary future drivers of drug development. In fact, In Silico Medicine sees INS018 055 and other drug candidates like it as validation for their software products, rather than the main source of profit for the company. Products like Chemistry42 and PandaOmics will be licensed to pharmaceutical companies to facilitate faster, cheaper, and more efficient drug

development. This is a process already underway with an enormous $1.2B licensing deal between In Silico Medicine and Sanofi, a major pharmaceutical company (Masson, 2022). Sanofi is paying In Silico Medicine an upfront $21.5M for access to the Pharma.AI platform and a team of drug discovery scientists. Depending on key research, development, and sales milestones, Sanofi is committed to additional payments of up to $1B (biopharma-reporter.com, 2022). The deal also includes potentially significant royalties for any products successfully developed. This is likely the model that companies developing AI for pharma research will follow. Collaborations and licensing deals with multinational pharmaceutical companies can couple the transformative power of AI-driven research with an existing manufacturing and distribution ecosystem for finished products. In theory, both companies will see financial benefit and an improved capacity to develop, manufacture, and distribute new products.

In Silico Medicine and its deal with Sanofi exemplify a future with a new standard for drug development. Lower upfront costs, a deeper reliance on software, and a faster rate of research will undoubtedly change the pharma landscape. The questions that remain are what this shift means for accountability, monopolization, and ethics in the world of big pharma.

Analysis

With an introduction to the technology being used in AI-driven pharmaceutical platforms and an example of the success of one such platform, the Analysis section will turn to examine how these computational tools will impact society at large. In particular, it will

focus on the ability of pharma giants to exert monopolistic power, how AI could increase accountability in the industry, and finally the social impact of changes in monopolization and accountability. These areas of focus were chosen because of their relationship with competition within the industry. In theory, a more competitive industry as a result of AI disruption could see more distributed market share, higher accountability, and a positive effect on society.

Monopolization:

Lower barriers to entry in the form of relatively inexpensive and risk-minimal *in silico* drug discovery will attract new entrants into the pharma field, driving an increase in competition. However, existing pharma giants will more than likely have the advantage in terms of tangible assets. Better access to high quality data and an ability to financially invest more into AI development are important factors in favor of a persisting oligopoly. Personnel and intangible assets are the two factors based on the business competitiveness framework that new entrants will have a better chance to compete in. Large pharma companies can be governed by a form of corporate inertia that may make them less adept at efficiently adopting cutting edge AI technology. To see a fundamental reallocation of market share in the pharmaceutical industry, new entrants must outmaneuver the established leaders by adopting transformative technological advances faster than their larger competitors.

Despite the better resourced big pharma companies, AI could still serve as a disruptive force reallocating power to smaller players. There are several recent examples of

technologies that have helped close the gap between small companies and the established industry leaders. In the mid-20th century, the computing industry was dominated by mainframe manufacturers, but the advent of personal computers shattered that monopoly by bringing access to computational power to individuals and small businesses. Another example is with 3D printing. Although certainly not as complete an industry transformation, 3D printing has also taken power from traditional methods such as injection molding and distributed it among smaller competitors who can now efficiently make complex and customizable products with significantly less infrastructure. AI will likely serve a similar role in making drug discovery a more accessible venture to those with fewer resources.

Accountability:

A diminished oligopoly may also drive accountability across the industry. A healthier competitive environment reduces the risk of conspiracy between pharma heavyweights and a highly digitized AI-driven drug discovery process will make it much easier to replicate or verify research. Many controversial lawsuits have stemmed from dishonest reporting on the efficacy or function of a drug or the outright suppression of contradictory research (Szalavitz, 2012). Either case of fraud may be easier to disprove in a future characterized by *in silico* research and development.

AI's potential for increased competition in the pharma space does depend heavily on small biotech companies' ability to incorporate these novel technologies. Currently, it is the pharma giants who are best positioned to invest in and capitalize on emerging

computational approaches to drug development. This is exemplified by the Sanofi licensing deal from the In Silico Medicine case study. For the industry to truly shift and see a significant reduction in entry barriers, smaller companies will need to focus on incorporating AI into their process (Foster, n.d.).

Another perspective from which to approach accountability is government regulation through agencies such as the FDA. As a regulatory tool, automation from AI could help reduce the time required for collecting, categorizing, and standardizing data from various records; overall reducing the amount of human involvement throughout this process. Regulatory procedure can be a slow process in pharmaceuticals. Records and data are complex and require regular maintenance and updating. Combining the computational power of AI as a tool for automating data management and analysis with human intelligence or guidance can dramatically relieve regulatory workload. This would maximize time for things like strategic planning of regulatory approvals and investigation into misconduct or flawed research (Patil et al., 2023). In doing so, AI could be a powerful tool for government and public oversight of pharma.

Social Impact/Ethics:

Stakeholders in the improvement of drug development efficiency are not limited to the users of medication but include pharma companies, suppliers, distributors, hospitals, insurance companies, and many others. With so many parties vying to benefit from increased productivity, it can be hard to predict who will be the primary beneficiaries of AI technology. A decrease in the trend towards monopolization and an increase in

accountability efforts, however, point towards a future where the end users of medication see at least some of the benefit brought by these advances in technology. This positive trend for healthcare is dependent on the industry shifting to a more competitive environment, where influence is shared among a higher number of companies. If new entrants are able to capitalize on emerging AI technology faster and better than the current industry leaders, AI will have a positive effect on the pharma industry.

Improvements in the transparency of drug development, efficiency of regulatory agencies, and reproducibility of experimentation are also important features of an AI-driven future that will benefit larger society. AI may create a more accountable industry, not derived just from an increase in the number of competing companies, but also from properties inherent to *in silico* research. Promoting ineffective drugs, covering up unflattering research, selling harmful medications, and violating antitrust laws to charge extortionary prices are but a few examples of recent multi-million-dollar settlements made by leaders in the pharmaceutical industry (Szalavitz, 2012). The vast majority of these crimes involve misrepresenting data, which will be harder to do when physical processes and experiments are replaced with computational models that are more quickly reproduced and validated by regulatory organizations. AI technology has the potential to reduce drug prices, increase competition, and ultimately address some of the primary ethical shortcomings of the pharmaceutical industry and its standard of practice.

Bibliography

AI In Drug Discovery And Development—Will It Live Up To The Hype? (n.d.). Retrieved February 22, 2024, from https://www.drugdiscoveryonline.com/doc/ai-in-drug-discovery-and-development-will-it-live-up-to-the-hype-0001

AI is about to remake the pharmaceutical industry – POLITICO. (n.d.). Retrieved September 29, 2023, from https://www.politico.eu/article/ai-is-about-to-remake-the-pharmaceutical-drug-medicines-industry/

Amoranitis, F. (2023, January 2). Drug development – The four phases. *BioStock*. https://www.biostock.se/en/2023/01/drug-development-the-four-phases/

biopharma-reporter.com. (2022, November 10). *Sanofi taps into Insilico Medicine's AI platform in deal worth up to $1.2bn*. Biopharma-Reporter.Com. https://www.biopharma-reporter.com/Article/2022/11/10/Sanofi-taps-into-Insilico-Medicine-s-AI-platform-in-deal-worth-up-to-1.2bn

Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., & Ahsan, M. J. (2022). Machine Learning in Drug Discovery: A Review. *Artificial Intelligence Review*, *55*(3), 1947–1999. https://doi.org/10.1007/s10462-021-10058-4

Field, H. (2023, June 29). *The first fully A.I.-generated drug enters clinical trials in human patients*. CNBC. https://www.cnbc.com/2023/06/29/ai-generated-drug-begins-clinical-trials-in-human-patients.html

Foster, C. T. (n.d.). *Council Post: Artificial Intelligence, The Pharmaceutical Industry, And Their Future*. Forbes. Retrieved October 24, 2023, from

https://www.forbes.com/sites/forbestechcouncil/2023/05/22/artificial-intelligence-the-pharmaceutical-industry-and-their-future/

How AI reduces the cost and time of drug discovery and development – NaturalAntibody. (n.d.). Retrieved September 29, 2023, from https://www.naturalantibody.com/use-case/how-ai-reduces-the-cost-and-time-of-drug-discovery-and-development/

Intelligent clinical trials. (n.d.). Deloitte Insights. Retrieved October 23, 2023, from https://www2.deloitte.com/us/en/insights/industry/life-sciences/artificial-intelligence-in-clinical-trials.html

Masson, G. (2022, November 8). *Insilico strikes gold with Sanofi deal worth up to $1.2B*. https://www.fiercebiotech.com/biotech/amid-biotech-winter-insilico-turns-heat-sanofi-deal-worth-12b-biobucks

Measuring the risks and rewards of drug development: New research from MIT shows that the success rates of clinical trials are higher than previously thought | MIT Sloan. (2018, January 31). https://mitsloan.mit.edu/press/measuring-risks-and-rewards-drug-development-new-research-mit-shows-success-rates-clinical-trials-are-higher-previously-thought

Medicine, I. (2023, January 10). *Insilico Medicine announces positive topline results of the New Zealand Phase 1 trial of INS018_055, an AI-designed drug for an AI-discovered target*. GlobeNewswire News Room. https://www.globenewswire.com/news-release/2023/01/10/2586249/31533/en/Insilico-Medicine-announces-positive-topline-results-of-the-New-Zealand-Phase-1-trial-of-INS018_055-an-AI-designed-drug-for-an-AI-discovered-target.html

News, R. M., Chemical &. Engineering. (n.d.). *Cost to Develop New Pharmaceutical Drug Now Exceeds $2.5B*. Scientific American. Retrieved March 21, 2024, from https://www.scientificamerican.com/article/cost-to-develop-new-pharmaceutical-drug-now-exceeds-2-5b/

Patil, R. S., Kulkarni, S. B., & Gaikwad, V. L. (2023). Artificial intelligence in pharmaceutical regulatory affairs. *Drug Discovery Today*, *28*(9), 103700. https://doi.org/10.1016/j.drudis.2023.103700

Philippidis, A. (2023, June 27). *Insilico's AI Candidate for IPF Doses First Patient in Phase II*. GEN - Genetic Engineering and Biotechnology News. https://www.genengnews.com/topics/artificial-intelligence/insilicos-ai-candidate-for-ipf-doses-first-patient-in-phase-ii/

Shah-Neville, W. (2024, January 5). Five AI drug discovery companies you should know about. *Labiotech.Eu*. https://www.labiotech.eu/best-biotech/ai-drug-discovery-companies/

Szalavitz, M. (2012, September 17). Pharma Behaving Badly | Top 10 Drug Company Settlements. *Time*. https://healthland.time.com/2012/09/17/pharma-behaving-badly-top-10-drug-company-settlements/

The top 20 drugs by 2021 sales. (n.d.). Retrieved March 21, 2024, from https://www.fiercepharma.com/special-reports/top-20-drugs-worldwide-sales-2021

University of Connecticut, Lou, B., Wu, L., & University of Pennsylvania. (2021). AI on Drugs: Can Artificial Intelligence Accelerate Drug Development? Evidence from a Large-

Scale Examination of Bio-Pharma Firms. *MIS Quarterly*, *45*(3), 1451–1482. https://doi.org/10.25300/MISQ/2021/16565

Vanhaelen, Q., Lin, Y.-C., & Zhavoronkov, A. (2020). The Advent of Generative Chemistry. *ACS Medicinal Chemistry Letters*, *11*(8), 1496–1505. https://doi.org/10.1021/acsmedchemlett.0c00088

Zhou, E. W., Jackson, M. J., & Ledley, F. D. (2023). Spending on Phased Clinical Development of Approved Drugs by the US National Institutes of Health Compared With Industry. *JAMA Health Forum*, *4*(7), e231921. https://doi.org/10.1001/jamahealthforum.2023.1921