Artificial Intelligence and Drug Innovation

Bowen Lou
The Wharton School, University of Pennsylvania, bowenlou@wharton.upenn.edu

Lynn Wu
University of Pennsylvania, wulynn@wharton.upenn.edu

Artificial Intelligence and Drug Innovation

Completed Research Paper

Bowen Lou
University of Pennsylvania
3730 Walnut Street,
Philadelphia PA, 19104
bowenlou@wharton.upenn.edu

Lynn Wu
University of Pennsylvania
3730 Walnut Street,
Philadelphia PA, 19104
wulynn@wharton.upenn.edu

Abstract

We study how artificial intelligence (AI) can influence the drug development process in the global pharmaceutical industry. Despite considerable effort made in developing drugs, pharmaceutical firms experience declines in novelty for drugs they produced. As AI becomes an important general purpose technology (GPT), it could be used to address some known challenges in the drug development process. Using two large-scale datasets that contain detailed historical records of global drug development and patents, we identify AI-related patents to approximate firms' AI capabilities and construct a relatively new similarity-based metric to measure drug novelty based on their chemical structure. We find that AI can primarily affect the earliest stage in drug discovery when tasks are heavily dependent on automatic data processing and reasoning. However, it may not necessarily help with the more expensive and risky clinical trial stages that require substantial human engagements and interventions. Additionally, AI can facilitate the development for drugs at the medium level of chemical novelty more than at the extreme ends of the spectrum. Our study sheds light on the understanding of the roles and limitations modern technology can have on drug development, one of the most complex innovation processes in the world.

Keywords: Artificial Intelligence, Drug Development, Innovation

Introduction

Artificial Intelligence (AI)'s impact on the global economy is increasing and has already transformed innovation and business practices in many industries.[1] Advances in AI including its subfield of machine learning have drastically improved in recent years and has empowered a wider variety of innovative business applications through the use of supervised and reinforcement learning. AI has currently surpassed human doctors in detecting certain cancers, enabled the first self-driving cars, and beaten the best human players in the game of Go and Starcraft.[2] As AI has progressed beyond the narrow domains that have traditionally been limiting its applications (Esteva et al. 2019; Topol 2019), AI is starting to exhibit characteristics of "general purpose technologies (GPT)" (Bresnahan and Trajtenberg 1995; Cockburn et al. 2018; Jovanovic and Rousseau 2005) that are expected to transform the nature and the process of innovation (Bughin et al. 2017).

Despite the promise of AI as a GPT, empirical evidence documents a general decline in innovation and productivity especially in the recent years (Bloom et al. 2017; Gordon 2017; Jones 2010). Medical

[1] According to a report dated in 2018, "How AI boosts industry profits and innovation" by Accenture research, it is estimated that AI will raise the profitability rate by an average of 38% by 2035 and provide an economic boost of $14 trillion across 16 different industries.

[2] https://www.forbes.com/sites/samshead/2019/01/25/deepmind-ai-beats-professional-human-starcraft-ii-players/ and https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/

innovation is particularly important as one of the greatest gains in life expectancy in developed countries has come from the new therapeutic drugs to treat conditions such as heart disease, cancer, and HIV/AIDS (Lichtenberg 1998; Lichtenberg and Sun 2007). Although the investment in pharmaceutical R&D has dramatically increased since 1990, productivity as measured by the novelty of new drugs being approved has declined (Pammolli et al. 2011). The new drugs developed are often incremental improvements upon existing drugs with minimal added therapeutic value, and they played an important role in driving up healthcare costs (Naci et al. 2015). The simultaneous increase in investment in AI and the decline in innovation quality mirrors the IT-productivity paradox post by Robert Solow in a 1987 remark, "we see IT everywhere but not in productivity statistics." Likewise, we can argue we see AI everywhere but not in the innovation statistics (Wu et al. 2018). In this paper, we examine how AI affects the development of new drugs. By focusing on one important industry where innovation is critical for productivity and competitive advantage, we aim to understand some potential drivers in explaining the apparent innovation paradox.

The pharmaceutical industry is extremely competitive with no single firm occupying a market share greater than 6% of the industry.[3] Thus, it is critical to continuously develop novel drugs that can surpass the competition. Although scientists have made considerable effort in identifying the underlying patterns and mechanisms of how a biological target could causes a disease, the drug development process is inherently slow when dealing with biological system, perhaps the most complex system in the world, costing billions of dollars per drug and taking 5 to 15 years to develop (Hughes et al. 2011).

AI could potentially accelerate drug development by efficiently identifying more new drug candidates that human researchers alone cannot identify. Accordingly, both small and large pharmaceutical firms are starting to use AI to aid drug discovery and development. We present two instances of anecdotal evidence of using AI to improve drug discovery from a large pharmaceutical company and a startup. With the goal of developing drugs in silico instead of in the lab, Novartis is actively expanding its AI capability and digital transformation by launching several innovative internal programs. It developed a digital infrastructure system that manages the digitization of medical records and clinical trials[4] and use predictive analytics and advanced AI to propose new treatments for diseases.

Start-ups are also using AI to facilitate drug discovery. BenevolentAI based in UK has already created its own AI platform for analyzing academic publications, patents, clinical trials and patient records. Deep learning[5] is applied to produce a knowledge graph for hypothesis generation and validation, and to infer the interplay between numerous chemical or biological entities such as genes, proteins, diseases and drug candidates (Fleming 2018). Based on these intelligent methods, new drug candidates have already been proposed to treat amyotrophic lateral sclerosis (ALS) (also known as motor neuron disease, MND), glioblastoma and Parkinson's disease, among others, and are believed to shorten the time for drug development at early stage by 60%[6]. The acceleration in ALS drug discovery is substantial as AI has aided the identification of 5 new compounds, of which, three are promising after extensive tests, and one works exceptionally well[7].

To understand how AI can facilitate the development of new drugs, we measure AI capability by using a large-scale global patent database, and measure drug developmental stages and chemical novelty of drug candidates using the Informa Pharmaprojects dataset. Following the computational methods in chemical informatics, we use a newly developed method to measure novelty based on the chemical structure of the drug and compare how it differ from prior drugs (Krieger et al. 2018). This measurement of novelty is

---

[3] https://www.hardmanandco.com/wp-content/uploads/2018/09/global-pharmaceuticals-2017-industry-stats-april-2018-1.pdf. Pfizer was the largest player that accounts for 5.5% share in the global pharmaceutical drugs market in 2017, followed by Novartis, Roche and Johnson & Johnson.

[4] Novartis Seeks Hidden Cures in Machine Learning, AI, https://www.informationweek.com/big-data/big-data-analytics/novartis-seeks-hidden-cures-in-machine-learning-ai/d/d-id/1332269

[5] Why we need to use AI for life not just lifestyle, https://benevolent.ai/blog/why-we-need-to-use-ai-for-life-not-just-lifestyle1

[6] Great News for Big Pharma: BenevolentAI Uses New Funds to Prepare ALS Drug Clinical Trials, https://healthcareweekly.com/big-pharma-news-benevolentai/

[7] This AI unicorn is disrupting the pharma industry in a big way, https://www.wired.co.uk/article/benevolent-ai-london-unicorn-pharma-startup

based on the chemical structure at the time of the drug candidate's initial development, and thus it would not conflate with ex-post measurement of success (e.g. getting FDA priority review or eventual market size).

To the best of our knowledge, our study is the first to systematically and empirically examine the linkage between AI capabilities and drug development. We document the change of drug development before and after a firm acquires AI capabilities and track how they affect different stages of the drug development process. We find that AI can primarily support the development of very early-stage molecular drugs, more specifically, during the discovery and pre-clinical research stage. We do not observe the significant impact of AI on any later stages during the three intermediate phases (Phase I/Phase II/Phase III) in clinical trials as well as the final stages for FDA approval or market launch. We further explore the effect of AI on drug novelty and find that AI is best at discovering drug candidates with medium level of novelty. For firms with similar research and development capabilities, we also find that those with AI capabilities tend to develop more novel drugs than those without AI capabilities.

Overall, these findings suggest that while AI has the promise of becoming an important GPT for a variety of applications, it can facilitate drug development only in the earliest stages in the development process where tasks are performed to substantively enhance human ability for addressing an extremely complex biological system that heavily relies on automatic data processing and reasoning. However, it does not seem to have an effect on helping the drug through the expensive and risky clinical trial stages that require substantial human engagement and intervention. In addition, we show that AI is suitable for developing drugs at the medium level of novelty, suggesting the limitations of using AI to develop drugs at all novelty levels. By examining the drug development process, one of the most innovation-intensive and complex processes, these insights also contribute to the growing debate of a simultaneous decrease in innovation productivity and the growing use of advance technologies. While AI can provide many benefits as a GPT, applying it to areas it cannot support could lead to mismanagement of AI and accordingly, contribute to the apparent lack of productivity gain from AI investments.

Theory and Literature Review

A considerable body of literature has abundantly documented the impact of digital transformation and technology in healthcare. The research on Health IT (HIT) often focuses on its efficacy on various healthcare services. HIT has been linked to reduced costs and improved quality in patient care through the use of large enterprise healthcare IT systems, such as personal health record (PHR), electronic medical record (EMR) systems, and clinical decision support systems (CDSS) (Agarwal et al. 2010; Hillestad et al. 2005; Murdoch and Detsky 2013). Agarwal et al. (2010) provides an overview of HIT as a key for improving healthcare services and outcomes, such as lowering mortality rates (Amarasingham et al. 2009; Devaraj and Kohli 2000; Devaraj and Kohli 2003) and improving patient safety (Aron et al. 2011; Parente and McCullough 2009). Goh et al. (2011) examines factors influencing the adoption and diffusion of HIT and its impact on delivery of healthcare service. While the effect of IT on healthcare services have been extensively studied, limited attention has been paid on how modern IT, especially recent advances in AI, affects drug product development.

Developing drugs is perhaps one of the most expensive processes in the world, costing about an average of $2.6-billion for a typical drug, with 90% of drug candidates failing to achieve regulatory approval from the FDA. This innovation process requires deep understanding of a complex biological system with up 25,000 genes generating millions of proteins that can interact with each other and with other cell types (Pisano 2006). Managing this complexity is primarily why it is difficult to developing new drug candidates (Dougherty and Dunne 2012). While the earlier attempt in digitizing the human genome to manage the complexity was touted for its potential in delivering new therapeutic treatments, it has not lived to the expectation in part due to the inability to effectively use data analytics tools. However, modern machine learning applications can substantially ease the process of identifying complex and anticipated interactions and can thus address some known challenges associated with the pharmaceutical innovation (Lo et al. 2018; Schneider 2018; Vamathevan et al. 2019). It is important to identify at which stage of the drug development process can AI have the most effects.

AI and Drug Development Processes

Drug development typically has 5 phases. The first is the discovery and pre-clinical trials stage where drug candidates are proposed to address certain biological targets. The next three phases (Phase I/Phase II/Phase III) involve human clinical trials. Lastly, if the drug passes these trials, it would gain the FDA's approval and would be launched into the market. While the failure rate at each of the stages is high, the reasons for the failure at each stage differ. During the discovery and preclinical trial stage, the bottleneck is in identifying drug candidates that can recognize and modify their targets to achieve therapeutic effects (Gashaw et al. 2011; Hughes et al. 2011). Because of the long and expensive process in developing a new drug, the pharmaceutical industry has fully recognized the importance of improving the efficiency of the drug discovery process, and the need to do things differently. AI has the potential to aid the drug discovery process in two important ways: 1) AI automatically and intelligently collects, digests, analyzes and detects complex patterns in the existing data about biological system; 2) AI can search systematically to generate more hypotheses for drug testing that scientists may not be able to process on their own (Gil et al. 2014; Vamathevan et al. 2019).

As information about the chemical structure of drug targets and existing drugs is digitized, they can be inferred from their common knowledge representation and interrelated concepts to discover new drug candidates. AI can be used to facilitate this process by expanding the space for potential drug-target pairs (Dougherty and Dunne 2012; Pisano 2006). Machine learning, especially supervised learning, can be particularly helpful because it requires only input-output pairs to be specified to make accurate predictions without the need to specify the mechanism or pathway of how a drug candidate can treat the target. AI can find hidden linkage and patterns in a vast amount of digital data that would have been extremely labor intensive or impossible for human to identify. Accordingly, AI can accelerate the process of identifying drug candidates and overcome the barriers in computation and data management that hamper the discovery of new drugs (Vamathevan et al. 2019). The resulting drug candidate suggested by AI could then serve as a useful starting point to examine whether it can address the disease.

This insight is backed by recent evidence that pharmaceutical firms, such as Novartis and BenevolentAI (introduced earlier), are actively using machine learning to uncover hidden patterns and generate new hypotheses about the relationships between certain molecular compounds and possible new treatments for certain health conditions. For example, to search for new cancer therapies, Genentech adopted a machine learning and simulation system from GNS Healthcare. The system can analyze a diverse set of patient data to uncover new pathways and personalized drug targets to individual patients.[8] The startup Atomwise pioneered the development of deep learning models to optimize drug designs and substantially shorten the process of discovering new drugs. It has excelled at the early stage of the discovery process by screening more than 10 million compounds every day to predict the bioactivity of small molecules (Wallach et al. 2015). It has received over $50M in early stage venture capital funding in a short amount of time.

Overall, these examples show AI can expedite hypothesis generation for drug-target pairs by identifying novel patterns in large data. Once a list of potential candidate targets is identified, AI can also help sift through which molecules possess suitable characteristics to become a drug candidate. This includes the optimization of chemical structures to improve drug properties such as toxicity and metabolism, which is costly and data-intensive (Lo et al. 2018; Schneider 2018). Similarly, AI system can reduce the human errors in this discovery process that has proven to be a bottle neck in the drug discovery process (Gil et al. 2014). By facilitating both the discovery and verification of drug candidates, we expect AI to have a strong effect on the discovery and pre-clinical trial stage.

Hypothesis 1: AI has a positive effect on the development of drugs prior to the clinical trials.

After the discovery and preclinical stage, clinical trials are used to verify both safety and efficacy of the drug on actual human subjects (Junod 2014). Compared to the pre-clinical stage, clinical trials require far more human interactions and communications to meet different requirements set by the FDA.

[8] https://www.gnshealthcare.com/gns-healthcare-announces-collaboration-to-power-cancer-drug-development/

The primary reasons for failures in clinical trials are the drug candidates from earlier development stage lack the necessary efficacy and safety when used in human (Harrison 2016). This is common in clinical trials in part because biological system is extremely complex with plethora of unpredictable interactions (Dougherty and Dunne 2012). Accordingly, it is difficult to infer how human subjects would interact with the drug even when it has shown promise in the early stage. This is further compounded by differences in the varying demographics and health conditions of the patients. As stated in (Northrup 2005), what is not understood about the human body and how it functions is far greater than what is understood. The lack of knowledge about human body hinders the usefulness of AI to infer about complex interactions in human body. Furthermore, much of the decisions in clinical trials would still need to rely on human judgement and clinical experience. This type of knowledge is not only difficult to digitize but its incomplete nature further impedes the use of AI in clinical trials. Accordingly, the innovation process for finding a drug candidate at this stage goes beyond computational complexity that AI can support.

Moreover, a misunderstanding in correspondence with the FDA about a pre-clinical trial can put a program at risk. Each phase of a clinical trial also requires a patient pool and a customized dosage. Drug developers need to communicate with the FDA and submit documents about the evidence of drug safety and efficacy (Junod 2014; Sacks et al. 2014). These activities often require extensive human communications and thus cannot be easily automated using AI. As AI cannot yet replace human judgement, clinical experience and communication that are critical for the clinical trials and the FDA approval stage (Agrawal et al. 2018), we expect a limited effect of AI at the later stages of drug development.

Hypothesis 2: AI contributes to drug development at the later stages to a lesser extent than at the earlier stages.

AI and Drug Novelty

While AI can accelerate the discovery of drug candidates at the earliest stage, it is unclear whether the drug candidates they uncovered are incremental or sufficiently novel. Research has shown that discovering novel drug candidates is much harder than discovering incremental "me-too" drugs (Krieger et al. 2018). Although the return of novel drugs in both financial terms and in therapeutic value is substantially higher than "me-too" drugs, they also incur much higher risks. AI can potentially accelerate the process for discovering novel drugs that the pharmaceutical companies desperately need. While it can also help the discovery of incremental drugs, the marginal benefits are likely to be smaller for "me-too" drugs than for novel drugs as firms may have already developed the competency in creating incremental drugs.

AI primarily enables the recognition of hidden patterns within digitized data at a much faster speed than human labor. For example, Watson uses AI to undercover 28 new p53 kinases, a cancer suppressor, in 2 months which would have taken researchers more than 6 years to do. The accelerated discovery is possible because there are already well-known p53 kinases with established functionality, so it is relatively easy for AI to recognize similar patterns in the data to find other types of p53 proteins that match existing patterns. However, when there is no data about a certain drug candidate or the drug candidate is sufficiently different from existing structures, it is difficult to use data to infer the functionalities of the drug, and as a result, AI is limited in facilitating this type of discovery (Wu et al. 2018).

Furthermore, it has been documented that AI lacks the abilities to make decisions based on small data, especially those involving human intuition, creativity and human insights (Simon 1977). Inferences based on small data may heavily depend on more tacit knowledge and "sticky" information that is inherently costly to collect, transfer, and therefore difficult to digitize and use as inputs to AI systems (Henderson and Clark 1990; Nonaka and Von Krogh 2009; Von Hippel 1994). Similarly, AI is also limited in facilitating the development of incremental drugs because firms have already developed capabilities to produce such drugs (Krieger et al. 2018). The application of AI in finding incremental drugs may be too expensive for the returns that firms would get from AI, especially in light of substantial upfront investments and strategic planning for the digital transformation required (Bughin et al. 2017).

Therefore, we expect AI is most effective in developing drugs that are of intermediate novelty. These drug candidates can benefit more from broad searches and linkages of diverse data that AI can facilitate, but AI is limited in producing entirely novel drugs that have minimal precedents or available data, and it is limited in producing incremental drugs that are derivatives of known compounds, a capability pharmaceutical companies have already acquired.

Hypothesis 3: AI is most helpful with the development of drugs in the intermediate level of novelty to a greater extent than in the lower and the higher levels of novelty.

Data and Measurement

Drug Development

We primarily focus on the global pharmaceutical industry, which has a well-established process for developing new drugs. The entire process can take 5 to 15 years to complete. Figure 1 shows the life cycle of drug development. However, despite established procedures, there is substantial uncertainty in the drug development process for several reasons. First, understanding the mechanism of disease requires enormous investments in research and development so that therapeutic solutions could be proposed to develop new drug candidates. Second, each of the stages in the drug development process has a high failure rate and faces different types of risks and uncertainty.

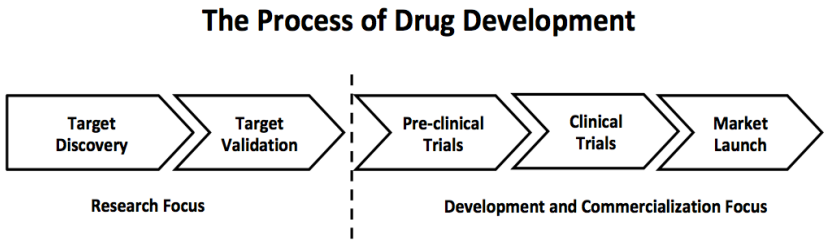
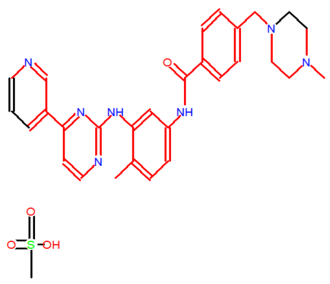


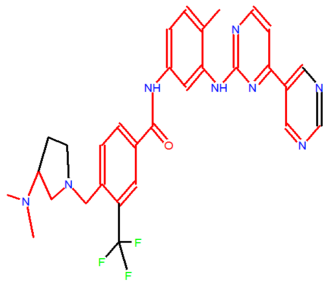
Figure 1. The Life Cycle of Drug Development (Kapoor and Klueter 2015)

We collect drug development data from Informa Pharmaprojects database between 1995 to 2017 (Eklund 2018). This database is the leading source about global drug development and has been widely used in a range of research about the global pharmaceutical industry (Hess and Rothaermel 2011; Kapoor and Klueter 2015; Sosa 2013). It provides a comprehensive coverage of drug candidates and tracks their development stages from the commencement of pre-clinical stage to the ultimate discontinuation or the worldwide market launch. In addition, the data includes the originators and licensees and all other firms involved in the development process. Some drugs may start the development at a small biotechnology firm before moving into the control of larger pharmaceutical firms for clinical trials. Control and development of drug candidates can also change hands through company mergers and acquisitions (M&A). We account for the transfer of drug patents and the associated rights using the Recap database and the Zephr database from Bureau Van Dijk to ensure that drugs in the Informa Pharmaprojects database are correctly matched to the firms responsible for their original development (Eklund 2018). Thus, at any given point in time, we can observe a firm's drug portfolio and pipeline.

Drug Novelty



Drug name: Imatinib Mesylate
Molecular formula: $C_{30}H_{35}N_7O_4S$



Drug name: Bafetinib
Molecular formula: $C_{30}H_{31}F_3N_8O$

Figure 2. Chemical Structures of Imatinib Mesylate and Bafetinib, with their Maximum Common Substructure Highlighted in Red

Using methods suggested in recent research literature in Chemical Informatics (Backman et al. 2011; Cao et al. 2008; Krieger et al. 2018), we measure the novelty in the chemical structure of 13,396 drugs.[9] We follow the "Similarity Property Principle," a central concept in chemistry, which states that structurally similar molecules are more likely to have similar physicochemical properties and biological activities (Johnson & Wiley-Interscience, 1991). We measure drug similarities using a similarity score that is based on the chemical adjacency between the focal drug candidate at its initial development and all previously drugs ever developed[10] (Krieger et al 2018).

First, we calculate the pair-wise similarity score between any drugs, X and Y, which is measured topologically by the "Tanimoto coefficient", which yields the ratio of the total number of atoms in the maximum common substructure (MCS) that appears in both of their structures and the total number of unique atoms[11] (Cao et al. 2008; Krieger et al. 2018; Nikolova and Jaworska 2003):

$$Similarity_{X,Y} = \frac{N_{X\&Y}}{N_X + N_Y - N_{X\&Y}}, \qquad (1)$$

where N_X and N_Y are the total number of atoms in chemical structures of drug X and drug Y respectively, and $N_{X\&Y}$ is the total number of atoms in MCS that appears in both drugs X and Y. Thus, a similarity score of zero means that the two drugs have no common components. A similarity score of 1 indicates that they have the same set of atoms and bonding, although it does not imply that the two molecules are identical because MCS does not take the orientation in space for the molecule into account. The classic example of two drugs with the same similarity score but different orientation is Nexium and Prilosec, where Nexium is a mixture of two version of the same molecule but with differing orientations while Prilosec has only one orientation. Thus, it is possible to have a new drug candidate with a similarity score of 1, meaning this drug differs only in orientation or combinational therapies involving compounds that were previously developed.

---

[9] Our database provides detailed historical records on the development of over sixty thousand drugs. But most drugs that never progress beyond the very early development stage don't have chemical structure information available. Neither do large molecule drugs (known as biologics) that require a more complicated synthesis of substances needed for their manufacturing. Those drugs provided with chemical structures belong to small molecule drugs mainly produced by chemical synthesis. Small molecule drugs make up over 80% of the market (Krieger et al. 2018; Otto et al. 2018).

[10] We also restrict the previous drugs to be within a certain time range so that our novelty score is not automatically decreasing mechanically as the base of comparison becomes larger over time. Our results using the 5-year range are similar.

[11] Conventionally any non-hydrogen atoms are included for computation.

The similarity metric we use has been widely used to search for similar chemical compounds, screen for related drugs, and digitally quantify certain properties of the chemical compound without testing them on animals (Wallach et al. 2015; Wawer et al. 2014). Although we note that similarity in molecular structure cannot precisely predict functional properties, the chemical informatics research has shown that molecular similarity is on average useful for identifying drug quality and novelty.

Our drug data provides the simplified molecular-input line-entry system (SMILES) codes, which is a chemical notation language mainly designed for digital processing (Weininger 1988). It allows for rigorous structure specification and encodes chemical structures as short ASCII strings, with each component to describe and identify atoms, bonds, rings, branching and other compound shapes. We convert the SMILES codes of each drug to its graph representation and use an MCS-based approach (Backman et al. 2011; Cao et al. 2008) on the graphs to compute pair-wise similarity scores.

The MCS approach also enables us to visually display the common and unique substructures between drugs, instead of simply showing a numerical value as typically shown in standard structural descriptor-based methods. We visualize the structure of two similar drugs Imatinib mesylate and Bafetinib, as well as their MCS highlighted in colors (see Figure 2). Imatinib mesylate, the mesylate salt of imatinib, is a first-generation tyrosine kinase inhibitor for treating chronic myelogenous leukemia (CML). Bafetinib was developed as a more powerful and alternative treatment for patients that have become resistant to Imatinib mesylate. In terms of the size of their chemical structures, they both have 42 atoms in total, with 35 atoms appeared in their MCS. Therefore, their pairwise similarity score is calculated as

$$\frac{35}{42+42-35} = 0.714.$$

After all pair-wise drug similarity scores are calculated, we compute the novelty score for each drug candidate by its maximum similarity score to all previously developed drug candidates. We use the time at a drug's earliest development stage as the basis for comparison. Accordingly, this metric does not conflate with ex-post measure of success such as in receiving an FDA priority or being the first in the market to treat a rare disease. As the pairwise similarity score is between 0 and 1, we define the novelty of drug i as:

$$Drug\ Novelty_i = 1 - \max_{j \in P_i} Similarity_{i,j}, \quad (2)$$

where P_i is all drug candidates that have reached at least the Phase I stage of clinical trials prior to the initial development of the focal drug i (Krieger et al. 2018). Therefore, a novel drug candidate should have a lower maximum similarity score and is likely to possess a more distinct molecule structure from its previous drug candidates[12]. The median of all scores of drug novelty is about 0.5. For measuring the drug novelty at the firm-level, we compute the total number of drugs within each novelty range for each firm in a particular year. We also further refine the aggregate by separately calculating the number of drugs at each stage of the drug development process. We do not use the average of novelty scores of firm's drugs for the case where a firm has not developed any drugs for some years.

Patent Stock

To measure a firm's investment in research and development in drugs, we gather information of global patents from a worldwide patent statistical database PATSTAT[13] created by the European Patent Office

[12] Despite the broad coverage of drug information in the Informa Pharmaprojects database, it's possible that we may still miss drugs at the earliest stage of development that are not recorded in the database. We address this issue by using year fixed effect estimation in our firm-level regression analyses. We also measure the novelty of a given drug by comparing it to early drug candidates that reach at least Phase I within a rolling 5-year window and get directionally similar results.

[13] https://www.epo.org/searching-for-patents/business/patstat.html. We use the 2017 Autumn Edition of the EPO PATSTAT database. The date of data collection from the source patent databases is the end of July for the PATSTAT Autumn Edition. On January 4th, 2018, the products "PATSTAT Biblio" and "PATSTAT Legal Status" from this Edition are combined into the new product "PATSTAT Global."

(EPO). PATSTAT offers bibliographical data for over 100 million patents from 90 patent issuing authorities that include both leading industrialized and developing countries going back as far as the nineteenth century. Each patent record contains a detailed patent application, citations, a title, an abstract, and legal persons (e.g. firms or any organizations) filing the patent applications. It identifies whether the patentees are business enterprises, higher education institutions, governmental agencies or individuals (Du Plessis et al. 2009). It also develops a comprehensive approach to standardize the original name of patentees automatically (Magerman et al. 2006). These global patents serve as critical indicators of innovation activities for all companies in the world.

We match the name of pharmaceutical firms from Informa Pharmaprojects database to patent assignees in EPO PATSTAT. We also adjust the assignee names to represent the original company that filed the patent after accounting for their merger and acquisition activities. Based on these matched pharmaceutical firms, we then retrieve their patent application documents from PATSTAT, and extract filing years, titles, abstracts and citations for these patents. Following the convention in the R&D literature (Griliches et al. 1986; Hall et al. 2005), we use the patent filing year (as opposed to the publication year) because it more closely approximates the time at which the firm produced and had the innovation described in the patent available. We measure a firm's R&D investment when the drug candidate starts the pre-clinical stage using the accumulated stock of patent applications that the firm has applied (Aggarwal and Hsu 2013).

Artificial Intelligence (AI) Stock

We measure AI innovation capability using the patents owned by each global pharmaceutical firm. We investigate various concepts, definitions and fundamentals that support AI and subfields within AI to comprehensively measurement AI innovation (Cockburn et al. 2018; Vamathevan et al. 2019; WIPO 2019; Yao et al. 2010). We use three steps to identify AI patent. We start with the latest international patent classifications (IPC) or Cooperative Patent Classification (CPC) that is linked to patents. As the United States Patent and Trademark Office (USPTO) already provides a clear patent class relating to AI: class 706 for "Data Processing – Artificial Intelligence" that consists of a large set of subclasses including "fuzzy logic hardware," "plural processing systems," "machine learning," "neural network," and "knowledge processing systems," we apply the IPC or CPC concordance for this class to obtain the classification code that is used for classifying global patents. Next, we leverage the content-based information available in our patent documents and construct a comprehensive list of validated words and phrases pertaining to AI and search these terms in both the title and abstract of our patents. Specifically, Cockburn et al. (2018) define three interrelated technological subfields within AI—robotics, symbolic systems and learning—that characterize the evolution of achievements in AI. We also follow a well-accepted ACM Computing Classification System (CCS) that accounts for the dynamic change of AI technologies (WIPO 2019). This method has been used for over 50 years to organize the concept and trends of technologies, which can significantly alleviate the lack of consensus on AI categorizations and avoid subjective classifications. CSS provides three major hierarchies to develop AI-related phrases for classification. They are (i) "artificial intelligence" hierarchy that comprise of AI functional application such as natural language processing, computer visions to simulate human cognitive tasks, and AI techniques used to realize the functions; (ii) "machine learning" hierarchy that unveils numerous learning-based AI techniques; (iii) "life and medical sciences" hierarchy under the "applied computing" category that covers activities concerning intelligent computing for producing medicines. Lastly, we also test on several variants of these keywords in our dictionary, but they do not qualitatively change the classifications of AI patents. The patent classification and content-based methods found 6,182 AI patents by 391 global pharmaceutical firms for our sample from 1995 to 2017. Similar to measuring firm's general patent application stock, we aggregate all the AI patents for which a firm has applied as the firm's AI patent stock. In Figure 3, we plot the average number of molecular drugs developed in a firm (left subfigure) and the average number of novel molecular drugs with similarity scores above the median (right subfigure). We graph them in a relative time scale before and after the time when those pharmaceutical firms are developing their first AI patents (at relative year of zero). Accordingly, the negative relative years indicate the period before firms have AI patents, and the positive relative years show the period when firms are developing AI patents. Overall, Figure 3 shows that there are more drugs being developed at the pre-clinical stage when firms have AI patents than when firms do not have AI patents, and the increase continues to exist for novel molecular drugs.

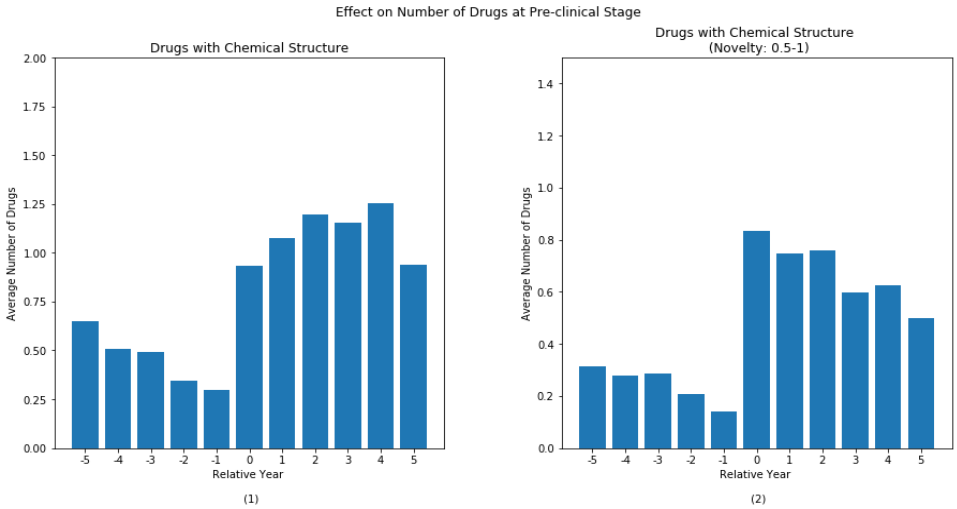


Figure 3. Number of Drugs around Firm's Starting to Build AI Patent Stock

Empirical Strategy and Identification

We merge Informa Pharmaprojects with PATSTAT to compile a firm-year panel dataset that has more than 200 global firms in the pharmaceutical industry between the years 1995 and 2017. Following the best practices in counting the number of unique drugs, we include only drug candidates that contain complete information about their development including progression through the various developmental stages and have advanced beyond the initial pre-clinical stage. However, a drug candidate may stop its development process in any of the clinical trial phases due to detected failure or healthcare regulations.

To estimate the effect of AI stock on drug development for pharmaceutical firms, we first estimate the effect on the number of drug candidates a firm produce in a given year after controlling for a firm's general research and technology investment and firm-specific fixed effects and time dummies.

$$\ln(Number\ of\ Drugs)_{it} = \beta_0 + \beta_1 \ln(AI\ Stock)_{it} + \beta_2 \ln(Patent\ Stock)_{it} + Controls_{it} + y_t + \gamma_i + \epsilon_{it}. \quad (3)$$

The number of drug candidates is highly skewed. Many pharmaceutical firms in our sample have zero drug candidates in a typical year. Firms with AI patents are more productive on average, having developed about two drug candidates in an average year. After accounting for the novelty of the drug candidates, the average number of drugs with sufficient novelty is less than one. Thus, to account for the skewness in the dependent variable, we take the logarithm of one plus the raw number of drug candidates in our main analysis[14]. We also include firm-fixed effects γ_i in the model to control for any unobserved time invariant differences in firm characteristics in AI and innovation investment, and also year-fixed effects y_t to account for temporal shocks and aggregate trends for the development of drug candidates by pharmaceutical firms in our sample. Our main coefficient of interest is β_1 that captures the marginal effect of a firm's continued investment in AI innovation on drug development. Because entrepreneurial exits have been shown to affect organizational innovation outcomes (Aggarwal and Hsu 2013), we incorporate a binary variable about the firm's financial ownership status over years (variable *Public Status* = 1 when the firm is publicly held, otherwise *Public Status* = 0). The ownership status is found in Crunchbase and PitchBook databases. This variable can also be used control for firms' drug development experience since most publicly traded pharmaceutical firms have already successfully developed at least one drug. We also control for firm age and number of employees.

[14] We also use models from Poisson and Negative Binomial regressions for our analyses, and the results are directionally similar.

Next, we focus on AI's effect on the novelty of drugs. Instead of counting all drug candidates, we only use drugs that are sufficiently novel. To determine novelty in drugs, we split the drug candidates into two groups, based on the median value of the novelty scores in our sample (which is 0.5 out of 1 in our drug sample). Those above the median are considered sufficiently novel. We also divide the range of novelty scores into 10 decile increments and compute the total number of drugs within each novelty range. The average of novelty scores is not used because a firm may not have developed any drugs in a particular year.

$$\ln(Number\ of\ Novel\ Drugs)_{it} = \beta_0 + \beta_1 \ln(AI\ Stock)_{it} + \beta_2 \ln(Patent\ Stock)_{it} + Controls_{it} + y_t + \gamma_i + \epsilon_{it}. \quad (4)$$

There are several endogenous factors we need to consider in equations 3 and 4. First, the estimate for AI can be overestimated if pharmaceutical firms with more slack resources choose to allocate technology investments for innovation. We address this reverse causality bias using instrumental variables that are derived from a patent-citation network. In this network, each node is a pharmaceutical firm, and the weight of a link between node A and B is all the citations B has received from A. We create two instrumental variables from the citation network to address endogeneity related to the AI patent stock: 1) the average number of accumulated AI patents produced by the neighboring pharmaceutical firms up to and including the observation year, 2) the average ratio of AI patent stock over all patent stock from these neighbors. Similar to the network-based approaches for instrument construction in (Wu et al. 2017), our instruments rely on the citation flows between firms. The flow can be used to approximate the ease in accessing external innovation. The flow can also satisfy the exclusion restriction as neighbors in the citation network (firms that cite each other) are not necessarily competitors in terms of their products and services. The network neighbors vary across industries and geographical locations and are thus less likely to be affected by common industry or region-specific shocks or competitive pressure. Accordingly, instrumental variables derived from "similar firms" in citation networks are less likely to face the Manski reflection issues that would be caused by instrumental variables using firms in the similar industries or geographical locations (Manski 1993).

Second, there could be selection biases in certain firms choosing to invest more in AI. To address this, we use the Coarsened Exact Matching method to match on the basis of firm's general production of patents, and the control variables used in our analyses. Matching on similar patent stock and general characteristics including ownership status, age and workforce size, firms with AI capabilities produces more of drugs than those without AI capabilities, and the rate of increase is much higher for more novel drugs.

Results

Table 1 shows the summary statistics and the correlation table of all the variables used in the study. We primarily focus on drugs with a known chemical structure in our data. In this sample, we observe a significant decrease in drug candidates as they progress from the pre-clinical trial stage to the last stage involving FDA approval and market launch stage, while the decrease between the pre-clinical and the clinical stage is weaker. We also observe firms with more general patents are more likely to have AI patents as the correlation between AI patent stock and general patent stock is about 0.6.

Table 1. Summary Statistics and Correlation Table

Variables	# of obs.	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10
For ln(Number of Drugs):													
1. at Pre-clinical Stage	4,988	0.24	0.63	1									
2. at Pre-clinical Stage, Novelty: 0-0.5	4,988	0.15	0.45	0.90	1								
3. at Pre-clinical Stage, Novelty: 0.5-1	4,988	0.16	0.50	0.93	0.74	1							
4. at Clinical-Trial Stages (Phase I-Phase III)	4,988	0.19	0.51	0.72	0.70	0.67	1						
5. at FDA Registration and Launched Stage	4,988	0.035	0.19	0.46	0.46	0.42	0.59	1					
6. ln(AI Stock)	4,988	1.37	1.32	0.18	0.18	0.19	0.19	0.24	1				
7. ln(Patent Stock)	4,988	5.24	2.56	0.37	0.33	0.36	0.35	0.31	0.60	1			
8. Public Status	4,988	0.30	0.46	0.25	0.22	0.24	0.34	0.24	0.20	0.36	1		
9. ln(Firm Age)	4,988	2.90	1.17	0.26	0.24	0.25	0.24	0.16	0.27	0.52	0.22	1	
10. ln(Number of Employees)	4,988	8.71	2.30	0.16	0.16	0.16	0.12	0.15	0.079	0.14	-0.18	0.20	1

We then relate firm's innovation capability in AI to the number of drug candidates that a firm generates in a year. To disentangle the effect of AI from general research investment, we control for a firm's accumulated patent stock, financial ownership status, age, and total number of employees. After controlling for firm and year fixed effects in our regression analyses, we find AI patent stock is positively associated with the number of new drugs developed at the pre-clinical stage. Specifically, a one percentage increase in a firm's investment in AI innovation is associated with about 0.03% additional change in the number of drugs (Column 1 to Column 2 in Table 2). However, this effect disappears in the next three phases of clinical trials and the final stage for FDA approval or market launch (Table 3)[15]. This suggests that AI plays a more limited role in developing drugs at later stages than at the earlier stage. Overall these results provide support for Hypothesis 1 and 2.

Table 2. AI on Development of Drugs at Pre-clinical Stage

DV	(1) ln(Number of Drugs)	(2) ln(Number of Drugs)	(3) ln(Number of Drugs, Novelty: 0-0.5)	(4) ln(Number of Drugs, Novelty: 0-0.5)	(5) ln(Number of Drugs, Novelty: 0.5-1)	(6) ln(Number of Drugs, Novelty: 0.5-1)
ln(AI Stock)	0.038***	0.025***	0.010	0.0073	0.034***	0.020***
	(0.0080)	(0.0083)	(0.0072)	(0.0075)	(0.0067)	(0.0070)
ln(Patent Stock)		0.040***		0.0092		0.045***
		(0.0076)		(0.0069)		(0.0064)
Public Status	-0.11***	-0.11***	-0.059***	-0.060***	-0.088***	-0.093***
	(0.020)	(0.020)	(0.018)	(0.018)	(0.017)	(0.017)
ln(Firm Age)	0.13***	0.11***	0.089***	0.085***	0.083***	0.063***
	(0.013)	(0.013)	(0.011)	(0.012)	(0.011)	(0.011)
ln(Number of Employees)	-0.0076***	-0.0075***	-0.0043*	-0.0043*	-0.0066***	-0.0065***
	(0.0028)	(0.0028)	(0.0025)	(0.0025)	(0.0023)	(0.0023)
Observations	4,988	4,988	4,988	4,988	4,988	4,988
R-squared	0.82	0.82	0.71	0.71	0.80	0.80
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES

Next, we examine how a firm's AI patents can affect the novelty of drugs firms choose to develop, because to reduce risks, pharmaceutical companies tend to focus on the development of incremental "me-too" drugs as opposed to novel drugs with new therapeutic values. Instead of simply counting all drug candidates, we account for the novelty of the drug in Columns 3-6 of Table 2. First, we first split the drug candidates into two groups, based on the median value of the novelty scores in our sample (which is 0.5

[15] We also separately estimate the impact of AI stock separately on the three intermediate phases in clinical trials and observe that the coefficient of AI stock is not statistically significant from zero.

out of 1 in our drug sample). Our results suggest that while having AI patents may not affect the aggregated number of drugs with novelty scores being less than 0.5 (Column 3 and Column 4 in Table 2), it is positively associated with drugs that are above the median in novelty score (Column 5 and Column 6 in Table 2). This suggests that AI is primarily used to develop novel drugs as opposed to incremental drugs. To further explore this phenomenon, we finely divided the range of novelty scores from 0.5 to 1 into 5 decile increments and explore AI's effect on developing drugs in the 5 novelty ranges (Table 4) [16].

Table 3. AI on Development of Drugs at Post Pre-Clinical Stages

	At Clinical Trial Stages		At FDA Registration and Launched Stage	
	(1)	(2)	(3)	(4)
DV	ln(Number of Drugs)	ln(Number of Drugs)	ln(Number of Drugs)	ln(Number of Drugs)
ln(AI Stock)	0.0031	0.00011	-0.0017	0.0026
	(0.0068)	(0.0071)	(0.0034)	(0.0035)
ln(Patent Stock)		0.0095		-0.013***
		(0.0065)		(0.0032)
Public Status	0.096***	0.095***	0.042***	0.043***
	(0.017)	(0.017)	(0.0086)	(0.0086)
ln(Firm Age)	0.059***	0.055***	-0.040***	-0.034***
	(0.011)	(0.011)	(0.0053)	(0.0055)
ln(Number of Employees)	0.00020	0.00023	0.0039***	0.0038***
	(0.0024)	(0.0024)	(0.0012)	(0.0012)
Observations	4,988	4,988	4,988	4,988
R-squared	0.79	0.79	0.63	0.64
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES

Table 4. AI on Development of Novel Drugs with Chemical Structure Information at Pre-clinical Stage

DV	(1) Novelty: 0.9-1	(2) Novelty: 0.9-1	(3) Novelty: 0.8-0.9	(4) Novelty: 0.8-0.9	(5) Novelty: 0.7-0.8	(6) Novelty: 0.7-0.8	(7) Novelty: 0.6-0.7	(8) Novelty: 0.6-0.7	(9) Novelty: 0.5-0.6	(10) Novelty: 0.5-0.6
ln(AI Stock)	0.0080	-0.0070	0.00045	3.75×10^{-5}	-0.0018	-0.0026	0.00090	-8.73×10^{-5}	0.019***	0.015**
	(0.0064)	(0.0066)	(0.00061)	(0.00063)	(0.0019)	(0.0020)	(0.0042)	(0.0044)	(0.0055)	(0.0057)
ln(Patent Stock)		0.047***		0.0013**		0.0025		0.0031		0.013**
		(0.0060)		(0.00058)		(0.0018)		(0.0040)		(0.0052)
Public Status	-0.068***	-0.073***	0.0017	0.0015	0.0028	0.0025	-0.014	-0.014	-0.026*	-0.028**
	(0.016)	(0.016)	(0.0015)	(0.0016)	(0.0049)	(0.0049)	(0.011)	(0.011)	(0.014)	(0.014)
ln(Firm Age)	0.074***	0.053***	0.0020**	0.0014	0.0099***	0.0088***	0.020***	0.018***	0.034***	0.028***
	(0.010)	(0.010)	(0.00095)	(0.00099)	(0.0030)	(0.0031)	(0.0066)	(0.0069)	(0.0086)	(0.0089)
ln(Number of Employees)	-0.0075***	-0.0073***	6.56×10^{-5}	7.02×10^{-5}	0.00035	0.00036	-0.0021	-0.0021	-0.0019	-0.0018
	(0.0022)	(0.0022)	(0.00021)	(0.00021)	(0.00067)	(0.00067)	(0.0015)	(0.0015)	(0.0019)	(0.0019)
Observations	4,988	4,988	4,988	4,988	4,988	4,988	4,988	4,988	4,988	4,988
R-squared	0.43	0.44	0.13	0.13	0.19	0.19	0.54	0.54	0.77	0.77
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

We find that the effect of AI is small and statistically insignificant for novelty value above 0.6. However, at the middle range at around 0.5 to 0.6, we find a significantly positive relationship between AI and number of drugs developed. A one percentage increase in firm's AI innovation stock is associated with about 0.02% increase in the number of drugs with novelty range between 0.5 and 0.6. These results suggest that a firm's AI innovation investment can primarily help the development of medium-novel drugs more than

[16] We also examine the impact of AI on the drug novelty in the decile range from 0 to 0.5, but AI's effect is all insignificant.

completely novel ones or incremental drugs. Robust standard errors are clustered by firms to mitigate the serial correlation among within-cluster errors. Overall, these results support Hypothesis 3.

Table 5. AI on Development of Drugs with Chemical Structure Information at Pre-clinical Stage, 2SLS and CEM

DV	(1) ln(Number of Drugs)	(2) ln(Number of Drugs, Novelty: 0.5-1)	(3) ln(Number of Drugs, Novelty: 0.5-0.6)	(4) ln(Number of Drugs)	(5) ln(Number of Drugs, Novelty: 0.5-1)	(6) ln(Number of Drugs, Novelty: 0.5-0.6)
ln(AI Stock)	0.38**	0.27**	0.18*	0.048***	0.019*	0.016**
	(0.17)	(0.13)	(0.093)	(0.015)	(0.010)	(0.0081)
ln(Patent Stock)	-0.052	-0.018	-0.032	0.020**	0.017**	0.015***
	(0.046)	(0.035)	(0.025)	(0.0099)	(0.0066)	(0.0052)
Public Status	-0.093***	-0.079***	-0.017	0.045**	0.027**	0.013
	(0.032)	(0.027)	(0.019)	(0.019)	(0.013)	(0.010)
ln(Firm Age)	0.14***	0.084***	0.046***	-0.031	-0.035**	-0.0054
	(0.023)	(0.018)	(0.013)	(0.026)	(0.017)	(0.014)
ln(Number of Employees)	-0.0065**	-0.0058***	-0.0014	0.010***	0.0051**	0.0037**
	(0.0029)	(0.0022)	(0.0015)	(0.0035)	(0.0023)	(0.0018)
Observations	4,988	4,988	4,988	5,464	5,464	5,464
R-squared				0.63	0.61	0.57
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES

To alleviate the endogeneity concern that innovative firms may choose to develop AI capabilities, we use instrumental variables derived from the citation-network as described in the method section. We refer to neighboring firms' (firms that cite each other in patents) AI capability to instrument for one's own AI capability. The associated F-statistics in the first stage pass the weak instrument test. The 2SLS estimates show consistent results that AI capabilities positively affect the number of new drug candidates a firm develops, and the novelty of these drugs is more likely to be medium, ranging from 0.5 to 0.6 in their novelty scores (Column 1 to Column 3 in Table 5). Lastly, we use the Coarsened Exact Matching and continue to find a positive relationship between a firm's AI investment and the number of drug candidates produced (Column 4 to Column 6 in Table 5).

Conclusion and Discussion

In this study, we examine the impact of AI technologies in facilitating the drug discovery process for global pharmaceutical firms. Previous literature shows that the decline in R&D productivity in the pharmaceutical industry is strongly correlated to high-risk innovation investments or the creation of incremental drugs with minimal therapeutic benefits (Pammolli et al. 2011). A recent statistic suggest that only 2.5% of all drug candidates explored in the early research stages survive beyond the pre-clinical stage (Giovannetti and Morrison 2000; Kapoor and Klueter 2015), and the risks of failure in later stages continue to be staggering because of large financial investments. AI technologies including predictive data analytics could potentially alleviate these risks by accelerating the speed and scale at which to firms can process large-scale biomedical or chemical data and potentially facilitating the discovery of new and more novel drug candidates. Using AI patents to approximate a firm's ability in AI, we show that AI can indeed accelerate the identification of novel drug candidates but only at the discovery and pre-clinical trial stage; we do not observe that AI has any effects in the later stages of drug development during phase I-phase III of the human clinical trials nor does it help with FDA approval or commercial launch. This suggests that though AI offers tremendous opportunities and can be applied to all stages of drug discovery and development process (Vamathevan et al. 2019), AI's primary effect on drug development is at the earliest stage which can benefit more from exploratory data analysis in searching for potential drug targets. However, AI has limited effects on developmental stages that involve predicting complex interactions in human body, engaging and tracking human subjects and complying with external health regulations. These results suggest that AI can partially augment human intelligence in drug discovery where analytics is useful, but cannot capture all aspects of the complex biological system that evolves and changes rapidly,

replace human engagement and communicate with patients and regulators that are necessary to complete clinical trials and to comply with FDA regulations (Dougherty and Dunne 2012; Junod 2014).

Future research should further identify the bottlenecks in drug discovery and investigate the extent to which AI, and other technological advances, such as monitoring and tracking technologies in the personal health industry, can reduce these bottlenecks. Note that (Sertkaya et al. 2016) suggests the top three cost drivers cross all three phases in clinical trial are clinical procedure costs, administrative staff costs, and site monitoring costs. Although there are interests in integrating of mobile and wearable technologies into clinical trial programs, the progress has been slow due to regulatory and technological challenges in patient data management. As AI advances continues to accelerate, it is possible that AI could be applied in creative ways to help addressing challenges in the later stages of the drug development process. In addition, our study only examines the novelty of small molecule drugs with known chemical structures and formed by chemical synthesis, but not large molecule drugs (known as biologics) that would require a different novelty measure. Since the share of biologics has increased in the recent years and has started to receive growing interest from the pharmaceutical industry, understanding how AI can affect biologics can also be a fruitful future research direction.

Acknowledgements

We thank Mack Institute for Innovation Management at Wharton School, University of Pennsylvania for generous funding support for this project.

References

Agarwal, R., Gao, G., DesRoches, C., and Jha, A. K. 2010. "Research Commentary—the Digital Transformation of Healthcare: Current Status and the Road Ahead," *Information Systems Research* (21:4), pp. 796-809.

Aggarwal, V. A., and Hsu, D. H. 2013. "Entrepreneurial Exits and Innovation," *Management Science* (60:4), pp. 867-887.

Agrawal, A. K., Gans, J. S., and Goldfarb, A. 2018. "Exploring the Impact of Artificial Intelligence: Prediction Versus Judgment," National Bureau of Economic Research.

Amarasingham, R., Plantinga, L., Diener-West, M., Gaskin, D. J., and Powe, N. R. 2009. "Clinical Information Technologies and Inpatient Outcomes: A Multiple Hospital Study," *Archives of Internal Medicine* (169:2), pp. 108-114.

Aron, R., Dutta, S., Janakiraman, R., and Pathak, P. A. 2011. "The Impact of Automation of Systems on Medical Errors: Evidence from Field Research," *Information Systems Research* (22:3), pp. 429-446.

Backman, T. W., Cao, Y., and Girke, T. 2011. "Chemmine Tools: An Online Service for Analyzing and Clustering Small Molecules," *Nucleic Acids Research* (39:suppl_2), pp. W486-W491.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. 2017. "Are Ideas Getting Harder to Find?," National Bureau of Economic Research.

Bresnahan, T. F., and Trajtenberg, M. 1995. "General Purpose Technologies 'Engines of Growth'?," *Journal of Econometrics* (65:1), pp. 83-108.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., and Trench, M. 2017. "Artificial Intelligence–the Next Digital Frontier," *McKinsey Global Institute,* pp.1-80.

Cao, Y., Jiang, T., and Girke, T. 2008. "A Maximum Common Substructure-Based Algorithm for Searching and Predicting Drug-Like Compounds," *Bioinformatics* (24:13), pp. i366-i374.

Cockburn, I. M., Henderson, R., and Stern, S. 2018. "The Impact of Artificial Intelligence on Innovation," National Bureau of Economic Research.

Devaraj, S., and Kohli, R. 2000. "Information Technology Payoff in the Health-Care Industry: A Longitudinal Study," *Journal of Management Information Systems* (16:4), pp. 41-67.

Devaraj, S., and Kohli, R. 2003. "Performance Impacts of Information Technology: Is Actual Usage the Missing Link?," *Management Science* (49:3), pp. 273-289.

Dougherty, D., and Dunne, D. D. 2012. "Digital Science and Knowledge Boundaries in Complex Innovation," *Organization Science* (23:5), pp. 1467-1484.

Du Plessis, M., Van Looy, B., Song, X., and Magerman, T. 2009. "Data Production Methods for Harmonized Patent Indicators: Assignee Sector Allocation," *Luxembourg: EUROSTAT Working Paper and Studies*.

Eklund, J. 2018. "The Knowledge-Incentive Trade-Off: Understanding the Organization Design & Innovation Relationship," *Academy of Management Proceedings*: Academy of Management Briarcliff Manor, NY 10510, p. 16747.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., and Dean, J. 2019. "A Guide to Deep Learning in Healthcare," *Nature Medicine* (25:1), p. 24.

Fleming, N. 2018. "How Artificial Intelligence Is Changing Drug Discovery," *Nature* (557:7707), p. S55.

Gashaw, I., Ellinghaus, P., Sommer, A., and Asadullah, K. 2011. "What Makes a Good Drug Target?," *Drug Discovery Today* (16:23-24), pp. 1037-1043.

Gil, Y., Greaves, M., Hendler, J., and Hirsh, H. 2014. "Amplify Scientific Discovery with Artificial Intelligence," *Science* (346:6206), pp. 171-172.

Giovannetti, G., and Morrison, S. 2000. "Convergence: The Biotechnology Industry Report," *Palo Alto, CA: Ernst & Young*.

Goh, J. M., Gao, G., and Agarwal, R. 2011. "Evolving Work Routines: Adaptive Routinization of Information Technology in Healthcare," *Information Systems Research* (22:3), pp. 565-585.

Gordon, R. J. 2017. *The Rise and Fall of American Growth: The Us Standard of Living since the Civil War*. Princeton University Press.

Griliches, Z., Pakes, A., and Hall, B. H. 1986. "The Value of Patents as Indicators of Inventive Activity." National Bureau of Economic Research.

Hall, B. H., Jaffe, A., and Trajtenberg, M. 2005. "Market Value and Patent Citations," *RAND Journal of Economics*, pp. 16-38.

Harrison, R. K. 2016. "Phase Ii and Phase Iii Failures: 2013–2015," *Nat Rev Drug Discov* (15:12), pp. 817-818.

Henderson, R. M., and Clark, K. B. 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly*, pp. 9-30.

Hess, A. M., and Rothaermel, F. T. 2011. "When Are Assets Complementary? Star Scientists, Strategic Alliances, and Innovation in the Pharmaceutical Industry," *Strategic Management Journal* (32:8), pp. 895-909.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., and Taylor, R. 2005. "Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, and Costs," *Health Affairs* (24:5), pp. 1103-1117.

Hughes, J. P., Rees, S., Kalindjian, S. B., and Philpott, K. L. 2011. "Principles of Early Drug Discovery," *British Journal of Pharmacology* (162:6), pp. 1239-1249.

Jones, B. F. 2010. "Age and Great Invention," *The Review of Economics and Statistics* (92:1), pp. 1-14.

Jovanovic, B., and Rousseau, P. L. 2005. "General Purpose Technologies," in *Handbook of Economic Growth*. Elsevier, pp. 1181-1224.

Junod, S. W. 2014. "FDA and Clinical Drug Trials: A Short History," *US Food and Drug Administration*.

Kapoor, R., and Klueter, T. 2015. "Decoding the Adaptability–Rigidity Puzzle: Evidence from Pharmaceutical Incumbents' Pursuit of Gene Therapy and Monoclonal Antibodies," *Academy of Management Journal* (58:4), pp. 1180-1207.

Krieger, J. L., Li, D., and Papanikolaou, D. 2018. "Developing Novel Drugs," National Bureau of Economic Research.

Lichtenberg, F. R. 1998. "Pharmaceutical Innovation, Mortality Reduction, and Economic Growth," National Bureau of Economic Research.

Lichtenberg, F. R., and Sun, S. X. 2007. "The Impact of Medicare Part D on Prescription Drug Use by the Elderly," *Health Affairs* (26:6), pp. 1735-1744.

Lo, Y.-C., Rensi, S. E., Torng, W., and Altman, R. B. 2018. "Machine Learning in Chemoinformatics and Drug Discovery," *Drug Discovery Today*.

Magerman, T., Van Looy, B., and Song, X. 2006. "Data Production Methods for Harmonized Patent Statistics: Patentee Name Harmonization," *KUL Working Paper No. MSI 0605*.

Manski, C. F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies* (60:3), pp. 531-542.

Murdoch, T. B., and Detsky, A. S. 2013. "The Inevitable Application of Big Data to Health Care," *Jama* (309:13), pp. 1351-1352.
Naci, H., Carter, A. W., and Mossialos, E. 2015. "Why the Drug Development Pipeline Is Not Delivering Better Medicines," *BMJ* (351).
Nikolova, N., and Jaworska, J. 2003. "Approaches to Measure Chemical Similarity–a Review," *QSAR & Combinatorial Science* (22:9-10), pp. 1006-1026.
Nonaka, I., and Von Krogh, G. 2009. "Perspective—Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory," *Organization Science* (20:3), pp. 635-652.
Northrup, J. 2005. "Vthe Pharmaceutical Sector, V in Lr Burns (Ed.), the Business of Healthcare Innovation." Cambridge University Press, New York.
Otto, R., Santagostino, A., and Schrader, U. 2018. "Rapid Growth in Biopharma Challenges and Opportunities. 2014," *McKinsey & Company*.
Pammolli, F., Magazzini, L., and Riccaboni, M. 2011. "The Productivity Crisis in Pharmaceutical R&D," *Nature Reviews Drug Discovery* (10:6), p. 428.
Parente, S. T., and McCullough, J. S. 2009. "Health Information Technology and Patient Safety: Evidence from Panel Data," *Health Affairs* (28:2), pp. 357-360.
Pisano, G. P. 2006. *Science Business: The Promise, the Reality, and the Future of Biotech*. Harvard Business Press.
Sacks, L. V., Shamsuddin, H. H., Yasinskaya, Y. I., Bouri, K., Lanthier, M. L., and Sherman, R. E. 2014. "Scientific and Regulatory Reasons for Delay and Denial of FDA Approval of Initial Applications for New Drugs, 2000-2012," *Jama* (311:4), pp. 378-384.
Schneider, G. 2018. "Automating Drug Discovery," *Nature Reviews Drug Discovery* (17:2), p. 97.
Sertkaya, A., Wong, H.-H., Jessup, A., and Beleche, T. 2016. "Key Cost Drivers of Pharmaceutical Clinical Trials in the United States," *Clinical Trials* (13:2), pp. 117-126.
Simon, H. A. 1977. "Scientific Discovery and the Psychology of Problem Solving," in *Models of Discovery*. Springer, pp. 286-303.
Sosa, M. L. 2013. "Corporate Structure, Indirect Bankruptcy Costs, and the Advantage of De Novo Firms: The Case of Gene Therapy Research," *Organization Science* (25:3), pp. 850-867.
Topol, E. 2019. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.
Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., and Spitzer, M. 2019. "Applications of Machine Learning in Drug Discovery and Development," *Nature Reviews Drug Discovery*, p. 1.
Von Hippel, E. 1994. ""Sticky Information" and the Locus of Problem Solving: Implications for Innovation," *Management Science* (40:4), pp. 429-439.
Wallach, I., Dzamba, M., and Heifets, A. 2015. "Atomnet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-Based Drug Discovery," *arXiv preprint arXiv:1510.02855*.
Wawer, M. J., Li, K., Gustafsdottir, S. M., Ljosa, V., Bodycombe, N. E., Marton, M. A., Sokolnicki, K. L., Bray, M.-A., Kemp, M. M., and Winchester, E. 2014. "Toward Performance-Diverse Small-Molecule Libraries for Cell-Based Phenotypic Screening Using Multiplexed High-Dimensional Profiling," *Proceedings of the National Academy of Sciences* (111:30), pp. 10911-10916.
Weininger, D. 1988. "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules," *Journal of Chemical Information and Computer Sciences* (28:1), pp. 31-36.
WIPO. 2019. "WIPO Technology Trends 2019: Artificial Intelligence."
Wu, L., Hitt, L., and Lou, B. 2018. "Data Analytics Skills, Innovation and Firm Productivity," *Management Science* (Forthcoming).
Wu, L., Jin, F., and Hitt, L. M. 2017. "Are All Spillovers Created Equal? A Network Perspective on Information Technology Labor Movements," *Management Science* (64:7), pp. 3168-3186.
Yao, L., Evans, J. A., and Rzhetsky, A. 2010. "Novel Opportunities for Computational Biology and Sociology in Drug Discovery: Corrected Paper," *Trends in Biotechnology* (28:4), pp. 161-170.