Drug Discovery Today • Volume 00, Number 00 • July 2020 PERSPECTIVE

ELSEVIER

feature

The upside of being a digital pharma player

Alexander Schuhmacher[1], alexander.schuhmacher@reutlingen-university.de, **Alexander Gatto**[2], **Markus Hinder**[3], **Michael Kuss**[4] and **Oliver Gassmann**[5]

We investigated the state of artificial intelligence (AI) in pharmaceutical research and development (R&D) and outline here a risk and reward perspective regarding digital R&D. Given the novelty of the research area, a combined qualitative and quantitative research method was chosen, including the analysis of annual company reports, investor relations information, patent applications, and scientific publications of 21 pharmaceutical companies for the years 2014 to 2019. As a result, we can confirm that the industry is in an 'early mature' phase of using AI in R&D. Furthermore, we can demonstrate that, despite the efforts that need to be managed, recent developments in the industry indicate that it is worthwhile to invest to become a 'digital pharma player'.

Introduction

Pharmaceutical R&D has used elements of AI for at least the past 20 years. The few fields of application of AI have been predominantly in drug discovery. Most probably because of technical limitations, such as the lack of computer power or the limited availability of large data sets, AI did not generate the expected added value to the R&D process. Moreover, the deep learning revolution only began with the breakthrough of the deep neural network 'Alex net' in 2012. It was recently reported that pharmaceutical companies have now started to reconsider the role of AI and to develop AI business cases for R&D [1]. In particular, we expect four archetypes to develop in the industry in the coming years. We call them 'Conservative Pharma Players', 'Selective AI Explorers', 'Digital Pharma Players', and 'New Analytical Entrants', whereby the latter are technology companies, such as Google, more rooted in data analytics than in healthcare. Here, we analyze and describe the state of AI application in pharmaceutical R&D, and outline a risk and reward perspective toward a digital R&D as the determining factors of the currently varying use of AI by pharmaceutical companies.

State of AI in pharmaceutical R&D

Information from annual company reports, investor relations information, scientific publications, and patent applications of the top 20 pharmaceutical companies by total revenues in 2018 (plus Merck KGaA) show that AI does not yet have market relevance for the core strategies of most of these leading firms. Quite the contrary, because only Johnson&Johnson and Novartis have started to commercialize AI by launching products and services in healthcare, such as chatbots to exchange information in a human-like manner with patients and physicians. For example, Johnson&Johnson is offering one AI-based product to monitor infants' sleeping behavior (Nod^{TM}), whereas Novartis (together with Microsoft) is supporting physicians in disease diagnosing with its product 'Assess MS'. We could identify only 18 AI-related patent applications filed in the last years on behalf of the evaluated pharmaceutical companies. Compared with the thousands of applications that have been filed by leading technology companies, such as Microsoft, Alphabet, or Samsung in recent years, it is apparent that the pharmaceutical industry is only just beginning to apply AI [2]. The overall AI status of the pharmaceutical industry appears to be 'early mature': The activities of all 21 sample companies for the years 2014–2019 are represented by 417 publications (including 398 R&D-related AI publications), 73 company-internal

https://doi.org/10.1016/j.drudis.2020.06.002

research projects utilizing AI technologies, 61 collaborations with AI start-ups, seven research alliances/consortia, 11 acquisitions/investments in start-ups, and two joint ventures.

Here, we analyze what leading pharmaceutical companies are doing in the field of AI in R&D and why, as well as in which specific areas and to what extent firms are active. The question of why pharmaceutical companies are using AI in R&D can be derived from information disclosed in annual company reports and investor relations information. In general, the purpose of applying AI in pharmaceutical R&D is to understand, to predict and to support: (i) understand: covers activities that extend and refine an existing knowledge base. In this context, companies utilize AI to better understand the basic mechanisms of diseases, to explore molecular interactions, or to understand the principal impact of drugs on various types of patient; (ii) predict: refers to research activities in the field of pharmacokinetics, for instance to replace animal tests or to predict drug action in the body. Other predicting activities deal with treatment responses or mortality rates in clinical trials; and (iii) support: relates to the use of AI in target identification, lead finding, drug candidate selection, or patient selection in clinical trials.

Today, all leading pharmaceutical companies have at least one research collaboration with AI specialists to gain access to AI technologies. Of the evaluated companies, 12 participate in AI consortia and 11 acquired and/or invested in AI companies, most notably the acquisitions of Flatiron Health and Foundation Medicine by Roche in 2018. Annual company reports and investor relations information indicate that 11 of the companies analyzed are active in alliances, such as the AI innovation of Sweden (www.ai.se/en), Alliance for Artificial Intelligence in Healthcare (AAIH, www.theaaih.org), Accelerating Therapeutics for Opportunities in Medicine (ATOM, https://atomscience.org), Machine Learning Ledger Orchestration for Drug Discovery (MELLODY, www.melloddy.eu), MIT industry consortium of Machine Learning for Pharmaceutical Discovery and Synthesis (https://mLpds.mit.edu), or TriNetX.

By reviewing scientific literature published by the leading pharmaceutical companies in the 2014–2019 period, we gained insight into where the pharmaceutical industry is using AI in R&D (Fig. 1). With regard to the drug discovery phase, we identified 218 scientific publications that can be structured in four categories and seven

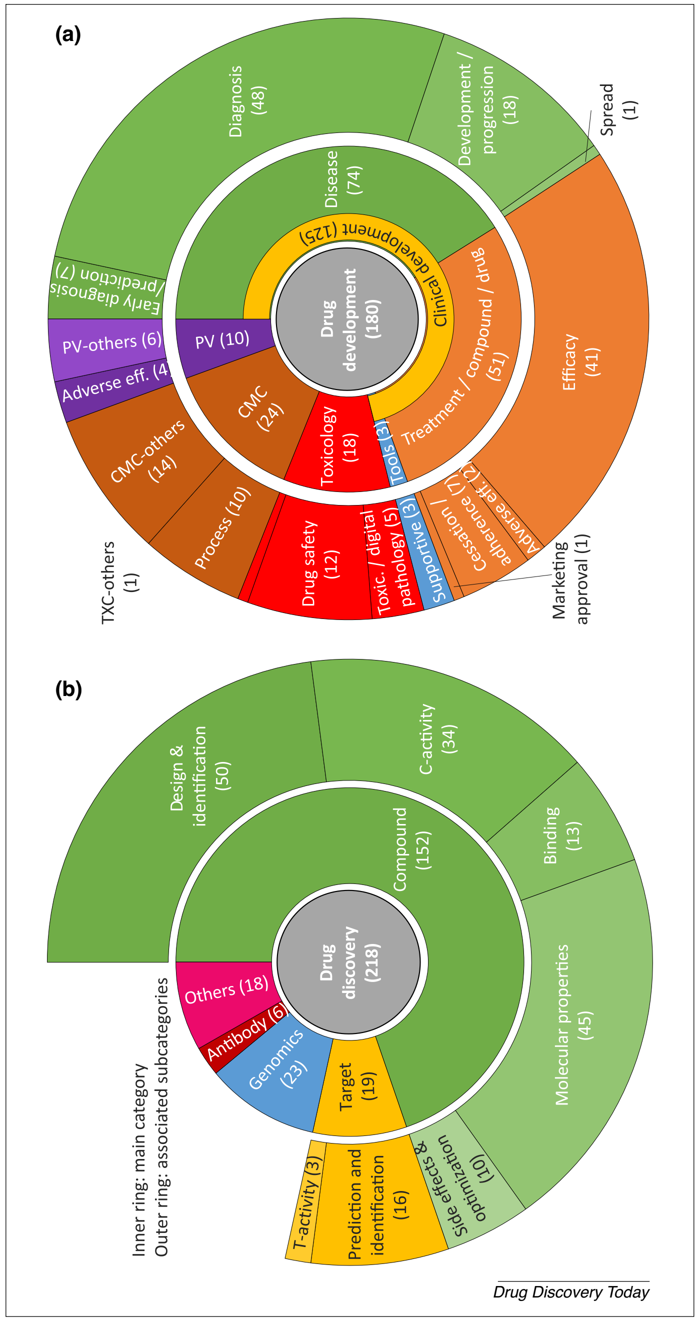

FIGURE 1

Categories and associated sub-categories of artificial intelligence (AI) applications in (**a**) drug discovery and (**b**) drug development for 2014–2019.

Pharmaceutical firms are using AI in drug discovery (a) within the research and development (R&D) value chain in four specific categories overall (represented by the segments of the inner ring). The main categories are labeled 'compound', 'genomics', 'target', and 'antibody'. There are seven associated subcategories (the segments of the outer ring). In drug development (b), pharmaceutical firms are using AI in four main categories (the segments of the inner rings): 'clinical development', 'toxicology', 'chemistry, manufacturing and controls' (CMC), and 'pharmacovigilance' (PV). There are 16 subcategories (the segments of the outer ring). The categories were identified and compiled by manually reviewing the abstracts of all scientific publications listed in the publication database SciFinder (https://scifinder.cas.org/) for the 21 investigated pharmaceutical companies examined for the period 2014–2019. The search term 'artificial intelligence OR deep learning OR machine learning' was used. Figures in brackets represent the number of publications identified per category. Following abbreviations were used: T-activity, target activity; C-activity, compound activity; PV-others, pharmacovigilance, others; and CMC-others, chemistry, manufacturing and controls, others.

Please cite this article in press as: Schuhmacher, A. et al. The upside of being a digital pharma player, Drug Discov Today (2020), https://doi.org/10.1016/j.drudis.2020.06.002

associated subcategories: (1) 'compound', (2) 'genomics', (3) 'target', and (4) 'antibody'. For example, the category 'compound' covers all activities that are related to the identification, design, and synthesis of a compound in drug discovery, the prediction of its activity, the prediction of its binding behavior and molecular properties, and the prediction of adverse effects and optimization strategies. With respect to drug development, we structured the 180 identified publications into four categories of AI use in R&D: (1) 'clinical development', (2)

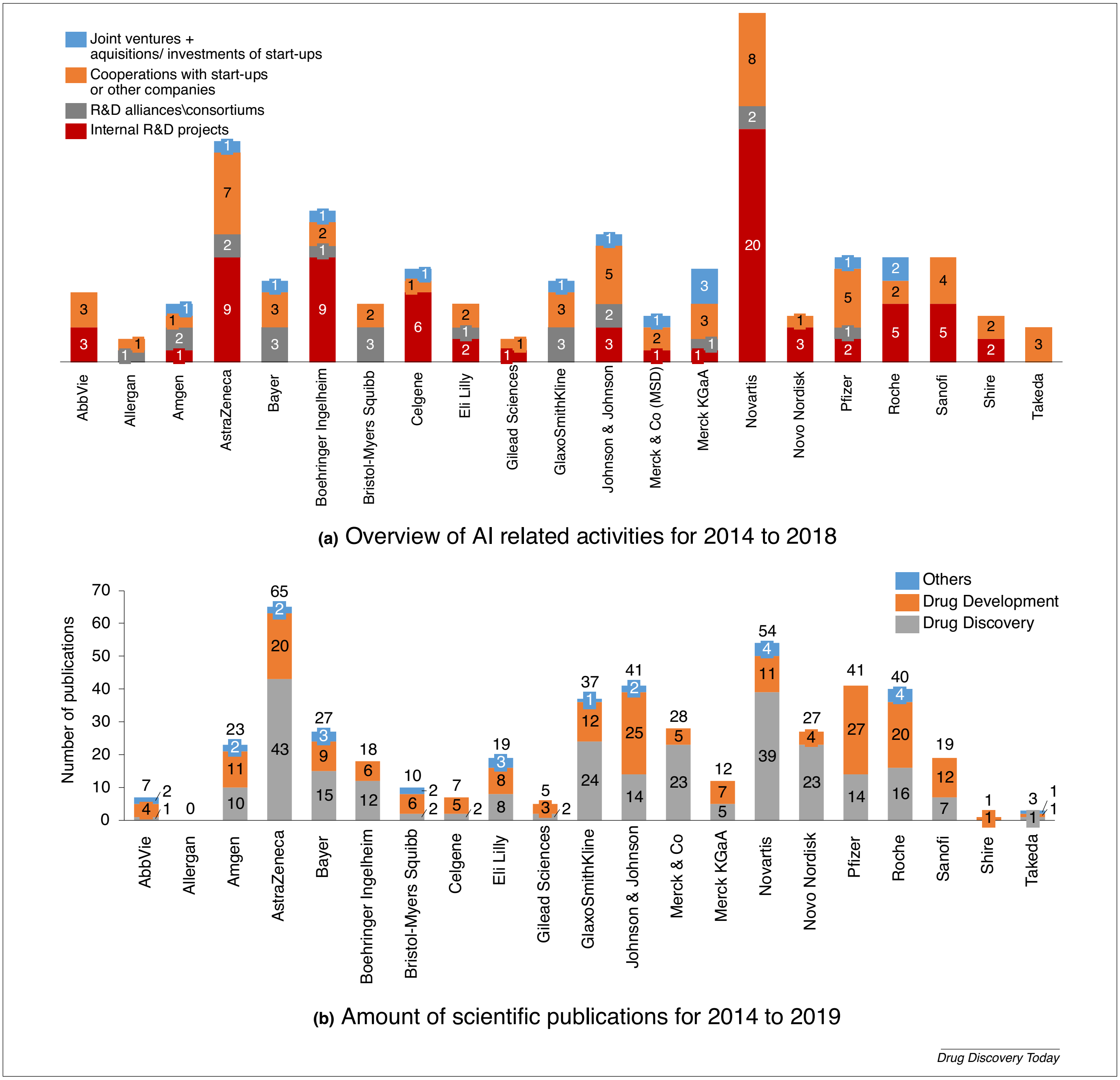

FIGURE 2

Artificial intelligence (AI)-related activities and publications by companies.

Companies focus their AI-related activities **(a)** on four main areas: driving internal research and development (R&D) projects, setting up R&D alliances or consortia, starting collaborations with start-ups or other companies, and investing in or acquiring start-ups or building joint ventures. Noticeable here is that the identified activities of AstraZeneca and Novartis amount to nearly twice the average of their competitors. The figures shown are based on the annual reports of the companies and their investor relations for the period 2014–2018, using the search terms 'artificial intelligence', 'deep learning', or 'machine learning'. The year 2019 was excluded for reasons as stated in the legend accompanying Fig. 3 in the main text. The total number of scientific publications relating to AI per company **(b)** broken down into drug discovery and drug development, in the period 2014–2019 indicates that AstraZeneca and Novartis were the most active in this regards. The category 'others' comprises publications that could not be assigned clearly to either drug discovery or drug development. As mentioned in the legend to Fig. 1, only publications found in SciFinder database were considered.

Please cite this article in press as: Schuhmacher, A. et al. The upside of being a digital pharma player, Drug Discov Today (2020), https://doi.org/10.1016/j.drudis.2020.06.002

'toxicology', (3) 'chemistry, manufacturing and controls', and (4) 'pharmacovigilance', which cover 16 subcategories. We conclude from this that the pharmaceutical industry is applying AI along the entire R&D value chain, from drug discovery to drug development.

As illustrated in Fig. 2, there is a wide variation in how much the leading pharmaceutical companies are doing with respect to AI in R&D. Two notable outliers in this context are Novartis and AstraZeneca. With, respectively, eight and seven active collaborations, 20 and nine reported internal research projects, as well as 54 and 65 scientific publications relating to AI, these two firms are by far the most active pharmaceutical companies in the AI field (Fig. 2). We also categorized the state of AI by contrasting the total revenues of the pharmaceutical companies (an indicator of their financial power) with the number of scientific publications they issued relating to AI in drug discovery and development in the 2014–2018 period (Fig. 3). Again, Novartis and AstraZeneca appear to be more involved than might be inferred from their financial position in the AI field, whereas peers, such as AbbVie, Gilead Sciences, Sanofi, and Takeda, appear to be doing less than expected.

Risk and reward perspective of a digital pharma R&D

In our view, the factors determining the varying use of AI by firms are the R&D strategies, the therapeutic areas, and the fact that pharmaceutical companies usually face significant risks when navigating towards a digital R&D. In this context, a qualitative distinction among the types of risk that pharmaceutical R&D decision-makers are facing can be made between preventable, strategic and external.

First, external risks are related to complexities that are beyond the control of a company, such as the dynamic regulatory environment or that AI talents expect high market rates of remuneration for their services [3].

Second, pharmaceutical companies need to overcome preventable risks, which are internal risks that could be managed and avoided by the company, such as the costs of a digital R&D, inflexible processes, the social dynamics of a digital transformation, and incomplete and inhomogeneous data records. Subcategories thereof are (1) technical; (2) performance; (3) control; (4) ethical; and (5) economic risks. The typical technical risks are: a mismatch of the AI systems with the business requirements, a lack of clear ownership and classification of data, and a lack of data integrity. This often implies the threat of a lack of confidence in the new system and, consequently, some of the potential of AI might be 'left on the table.' The typical performance risks associated with AI are discussed herein.

'Old data' are used and 'nothing new' is generated, because AI systems usually use historical data sets. AI can create a kind of 'black box,' whereby an AI system produces an outcome without knowing how it arrives at the result and the conclusions. Or inadequate information is used so that the result of applying AI might resemble a correlation even though the outcome is not significant. These examples imply the threat of misinterpretation and bias, which can have important consequences, such as selecting the wrong patient population for a trial or making incorrectly informed project and investment decisions. Next, standard AI use runs the risk of being 'out-of-the-box', lacking sufficient configuration and specifications when

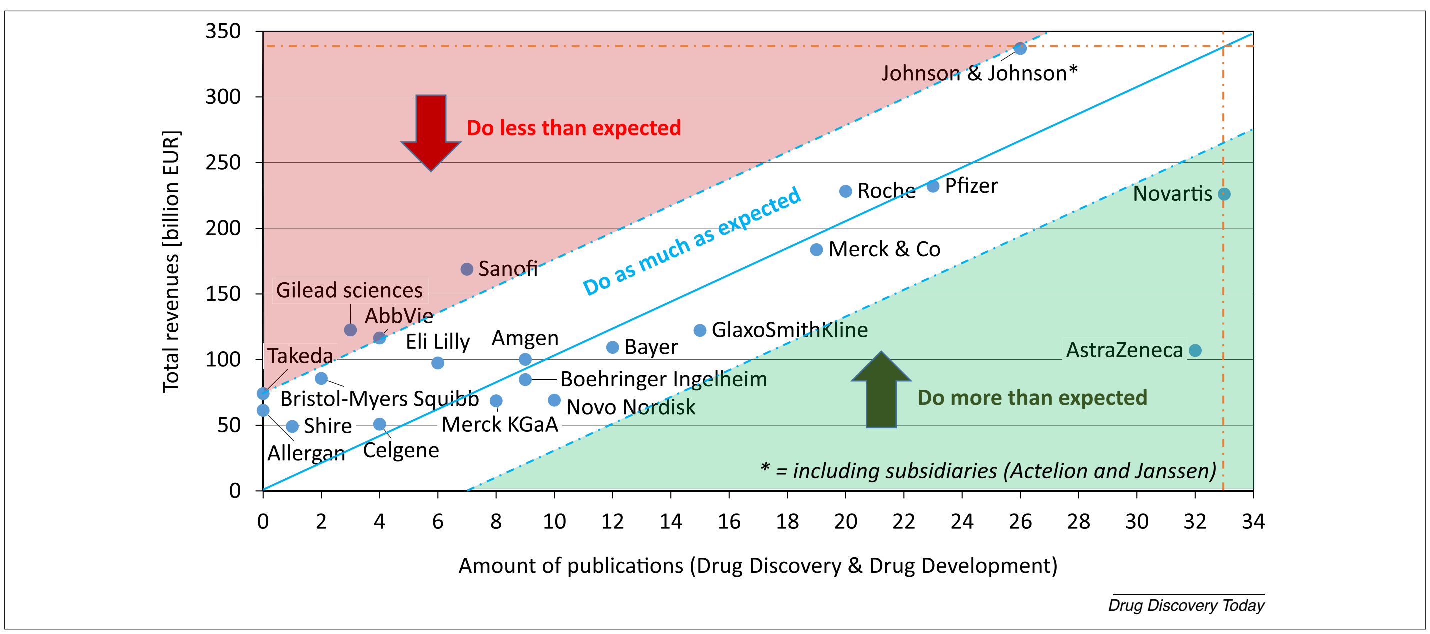

FIGURE 3

Status of artificial intelligence (AI)-application in research and development (R&D) of leading pharmaceutical companies for 2014–2018.
The use of AI in R&D of the leading pharmaceutical companies in our sample can be classified by plotting total revenues against the number of scientific publications for the period of 2014–2018. As a reference, a straight (the solid blue) line is drawn from zero (no revenue, no publication) to the coordinate point on the chart formed by taking the maximum total revenues and the maximum number of publications identified in the sample. The standard deviation of the total revenues from this reference line determines the border (the broken blue lines) of three groups, which represent companies that 'do less as expected' (e.g., Gilead Sciences or Sanofi), 'do as much as expected' (e.g., Boehringer Ingelheim or Pfizer), and 'do more as expected' (e.g., AstraZeneca or Novartis). Total revenues were taken from the annual reports of each company and the number of publications was taken from the SciFinder database (further details, see legend accompanying Fig. 1). The period 2014–2018 was selected, because annual reports for all of the companies in the sample were available for these years. In 2019, Allergan, Celgene, Gilead Sciences, and Shire did not issue an annual report because they had been acquired in the same or the previous year. The 2019 annual report of Takeda was pending at the time of the analysis because its fiscal year ends in March.

Please cite this article in press as: Schuhmacher, A. et al. The upside of being a digital pharma player, Drug Discov Today (2020), https://doi.org/10.1016/j.drudis.2020.06.002

implemented, and might not be tailored to the specific goals and needs of a company (control risk). This implies the threats of neglected value propositions and flawed business cases. Furthermore, the use of AI to make decisions that lead directly to action or the application of AI in making completely autonomous decisions raises ethical risks. According to a recent PwC Global CEO survey, 84% of CEOs agree that AI-based decisions need to be explained if they are to be trusted [4]. Otherwise, there are threats of a loss of trust and the rejection of AI-based outcomes. Finally, pharmaceutical companies need to deal with the economic risks associated with AI in R&D.

Third, strategic risks relate to the business challenge of taking higher risk for higher reward. For example, because AI has the potential to be a game changer, the path towards a digital pharma R&D might be very costly, without even knowing whether AI in R&D will materialize and keep its promises.

Notwithstanding all these risks that companies face on their AI journeys, the use of AI has the potential to provide great rewards by: (i) supporting *de novo* small-molecule design [5]; (ii) better predicting links between drug structure and activity, which can result in more efficacious and safer drugs [6]; (iii) better performing 3D protein structure simulation [7], which can increase the effectiveness of drug research; (iv) finding biomarkers, such as for invasive breast cancer, Alzheimer's disease, or other neuropsychiatric disorders [8,9]; (v) predicting the blood–brain barrier permeability of drugs [10]; (vi) predicting cancer drug responses [11,12]; (vii) predicting unknown drug–drug interactions [13]; (viii) predicting the risk of kidney disease/injury [14,15]; (ix) predicting the responses of patients with cancer to chemotherapeutic drugs [16]; (x) analyzing digital images or videos, and recording the presence of pain [17]; (xi) predicting mortality to forecast the risk of early death from chronic diseases [18]; (xii) analyzing mammograms 30 times faster than clinicians, and with 99% accuracy [19]; (xiii) enhancing the prediction of adverse drug reactions [20]; (xiv) diagnosing diseases in their initial stages [21]; (xv) analyzing cells in real time, allowing the development of targeted therapies based on the diagnoses of the specific genomic mutation that occur [22]; (xvi) predicting patients who would likely skip appointments [23]; (xvii) facilitating drug repositioning [24]; (xviii) streamlining clinical trial operations [25]; (xix) reanalyzing large trials or data sets for clinical signals of efficacy or safety in new indications [26]; and (xx) analyzing markers and predictors of drug response to identify patients who can benefit most from treatment [27].

These examples demonstrate the potential of a digital pharma R&D: data analytical and algorithm-based R&D can be used to improve the speed, focus, and quality of drug discovery and development. Besides drug discovery, which has traditionally been a field for the application of AI, translational and clinical development will profit from the application of AI and data analytics in the future. For example, predictive analytics using real-world data can help set up hypotheses, which can be verified through prospective investigations, likely increasing their success rate. Second, AI can revolutionize operational study conduct. This includes predicting enrolment and improving the identification of study centers. Analysis of big data sets enables better defined (stratified) patient groups, ultimately helping to develop precision-targeted therapies, which increase success rates. By analyzing data over time and using real-world data, better insights can be delivered, predictions improved, and trial designs adapted and optimized. Through flexible trials, efficiency can be improved, and drugs could well be approved earlier, because targeted trials allow for smaller patient samples. AI can help reduce trial sizes only if those sites are selected where it is more likely that endpoints can be detected. Better selection of trial sites means more efficient trials with fewer patients screened and lower costs [28].

First successes in this transformation process towards a digital pharma R&D have been recently reported by IQVIA: a 20.6% increase in the enrolment rate was recorded, 70% of adverse drug reaction case reports could be processed automatically, and the precision rate in screening of a predictive algorithm was 79% [29].

Concluding remarks: the path towards a digital pharma R&D

In our view, leading pharmaceutical companies of today are in an 'early mature' phase of using AI in R&D. The impact of its application is still not proven: AI has not yet contributed to a sufficient extent to R&D efficiency, effectiveness, or productivity. Nevertheless, AI holds promise for the future of pharma R&D and we foresee that more and more pharmaceutical companies might evolve into 'digital pharma players'.

On their path towards a digital pharma R&D and to benefit fully from AI, pharmaceutical executives must decide whether they want to start their AI journey and take the related strategic risks. To fully benefit from it, pharmaceutical companies will need to set up digital strategies, develop related business cases, and provide AI-specific budgets. In this context, decision-makers need to decide how they want gradually to redirect their human and financial resources from traditional R&D units to the new AI-related data science functions. They also need to improve their brand as an employer to become more attractive to data scientists. In addition, they must accept the high wages of data experts and the related high costs for digital upskilling of their R&D functions. In particular, they further need to consider that, in an AI-driven world, the pharmaceutical company with the best data and the best abilities to analyze it will be able to develop the best products. In our view, only leading pharmaceutical companies with extensive R&D portfolios have sufficiently large data sets to enable the systematic use of AI in R&D. Their challenge is to make the information they have (more) accessible and analyzable. Other leading companies with fewer R&D projects might have difficulties to tap into sufficiently large data sets that are useful for AI. Hence, they will need to collaborate with peers, create consortia, or acquire external data. This implies more preventable risks, such as the efforts of coordination or the 'not invented here' (NIH) syndrome. In addition, they need to take the external risk that acquired data are not proprietary (rather commodity) and might result in less innovation. Without exclusive data, pharmaceutical companies will face the unnecessary (external) risk that the digital pharma R&D will miss the expected upside, similar to the discussions that we know from externally acquired compound libraries and high-throughput screening [30].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1 Schuhmacher, A. *et al.* (2019) The art of virtualizing of pharma R&D. *Drug Discov. Today* 24, 2105–2107
2 Gramke, K. (2019) Artificial intelligence. EconSight: https://www.econsight.ch/artificial-intelligence/. [Accessed 27 May 2020].
3 Mullard, A. (2019) An audience with Daphne Koller. *Nat. Rev. Drug Discov.* 18, 576–577
4 PwC (2019) *22nd Annual Global CEO Survey. CEOs' Curbed Confidence Spells Caution.* PwC
5 Zhavoronkov, A. *et al.* (2019) Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 37, 1038–1040
6 Zhavoronkov, A. *et al.* (2019) Artificial intelligence for aging and longevity research: research advances and perspectives. *Ageing Res. Rev.* 49, 49–66
7 Service, R.F. (2018) Google's DeepMind aces protein folding. *Science* . http://dx.doi.org/10.1126/science.aaw2747 Published online December 6, 2018
8 Lee, K. *et al.* (2019) CPEM: Accurate cancer type classification based on somatic alterations using an

Please cite this article in press as: Schuhmacher, A. et al. The upside of being a digital pharma player, Drug Discov Today (2020), https://doi.org/10.1016/j.drudis.2020.06.002

ensemble of a random forest and a deep neural network. *Sci. Rep.* 9, 16927

9 Park, E.K. *et al.* (2019) Machine learning approaches to radiogenomics of breast cancer using low-dose perfusion computed tomography: predicting prognostic biomarkers and molecular subtypes. *Sci. Rep.* 9, 17847

10 Miao, R. *et al.* (2019) Improved classification of blood-brain-barrier drugs using deep learning. *Sci. Rep.* 9, 8802

11 Chang, Y. *et al.* (2018) Cancer Drug Response Profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci. Rep.* 8, 8857

12 Mucaki, E.J. *et al.* (2019) Predicting responses to platin chemotherapy agents with biochemically-inspired machine learning. *Sig. Transduct. Target. Ther.* 4, 1

13 Rohani, N. and Eslahchi, C. (2019) Drug-drug interaction predicting by neural network using integrated similarity. *Sci. Rep.* 9, 13645

14 Ravizza, S. *et al.* (2019) Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. *Nat. Med.* 25, 57–59

15 Tomašev, N. *et al.* (2019) A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572, 116–119

16 Huang, C. *et al.* (2018) Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. *Sci. Rep.* 8, 16444

17 Atee, M. *et al.* (2017) Pain assessment in dementia: evaluation of a point-of-care technological solution. *J. Alzheimer's Dis.* 60 (1), 137–150

18 Weng, S.F. *et al.* (2019) Prediction of premature all-cause mortality: a prospective general population cohort study comparing machine-learning and standard epidemiological approaches. *PLoS One* 14, e0214365

19 He, T. *et al.* (2017) Deep learning analytics for diagnostic support of breast cancer disease management. *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE. pp. 365–368, Institute of Electrical and Electronics Engineers Inc.

20 Raja, K. *et al.* (2017) Machine learning workflow to enhance predictions of Adverse Drug Reactions (ADRs) through drug-gene interactions: application to drugs for cutaneous diseases. *Sci. Rep.* 7, 3690

21 Farooq, A. *et al.* (2017) Artificial intelligence based smart diagnosis of Alzheimer's disease and mild cognitive impairment. *2017 International Smart Cities Conference (ISC2), Wuxi, 2017, pp. 1–4, IEEE*

22 Rutgers University (2019) *Cancer Device Created to See if Targeted Chemotherapy Is Working.* Rutgers University

23 Delvin, H. (2019) Hospital develops AI to identify patients likely to skip appointments. *The Guardian* April 12

24 Xie, L. *et al.* (2017) Discovery of novel therapeutic properties of drugs from transcriptional responses based on multi-label classification. *Sci. Rep.* 7, 7136

25 Goran, M. (2018) *Drug Development Gets Big Data Analytics Boost: With Nerve Live, Novartis is Harnessing the Power of Data for the Benefit of Patients.* Novartis

26 Schieker, M. (2018) Interleukin-1β inhibition with canakinumab associates with reduced rates of total hip and knee replacement (THR/TKR) and osteoarthritis (OA) symptoms: exploratory results from the canakinumab anti-inflammatory thrombosis outcomes study (CANTOS) abstract]. *Arthritis Rheum.* 70 (Suppl. 10), 445

27 Harrer, S. *et al.* (2019) Artificial intelligence for clinical trial design. *Trends Pharmacol. Sci.* 40, 577–591

28 IQVIA (2019) *The Changing Landscape of research and Development.* IQVIA

29 Glass, L. *et al.* (2019) *AI in Clinical Development.* Informa Pharma Intelligence

30 Macarron, R. *et al.* (2011) Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* 10, 188–195

Alexander Schuhmacher[1,*]
Alexander Gatto[2]
Markus Hinder[3]
Michael Kuss[4]
Oliver Gassmann[5]

[1]Reutlingen University, Alteburgstrasse 150, DE-72762 Reutlingen, GermanyReutlingen University, Alteburgstrasse 150, DE-72762 Reutlingen, Germany
[2]Sony Europe B.V., Stuttgart Technology Center, Hedelfinger Strasse 61, DE-70327 Stuttgart, GermanySony Europe B.V., Stuttgart Technology Center, Hedelfinger Strasse 61, DE-70327 Stuttgart, Germany
[3]Novartis Institute of BioMedical Research, Postfach, Forum 1, CH-4002 Basel, SwitzerlandNovartis Institute of BioMedical Research, Postfach, Forum 1, CH-4002 Basel, Switzerland
[4]PricewaterhouseCoopers AG, Birchstrasse 160, CH-8050 Zurich, SwitzerlandPricewaterhouseCoopers AG, Birchstrasse 160, CH-8050 Zurich, Switzerland
[5]University of St. Gallen, Institute of Technology Management, Dufourstrasse 40a, CH-9000 St. Gallen, SwitzerlandUniversity of St. Gallen, Institute of Technology Management, Dufourstrasse 40a, CH-9000 St. Gallen, Switzerland

Corresponding author.

Please cite this article in press as: Schuhmacher, A. et al. The upside of being a digital pharma player, Drug Discov Today (2020), https://doi.org/10.1016/j.drudis.2020.06.002