Artificial Intelligence and Social Computing, Vol. 72, 2023, 48–57
https://doi.org/10.54941/ahfe1003273

Machine Reading Comprehension and Expert System Technologies for Social Innovation in the Drug Excipient Selection Process

Evangelos Markopoulos[1] and Chrystalla Protopapa[2]

[1]University of Turku, Turku, 20014, Finland
[2]National and Kapodistrian University of Athens, Athens, 10679, Greece

ABSTRACT

Artificial Intelligence has been used for disruptive but also social innovations in all disciplines and especially in the health sector. This paper addresses a drug formulation challenge and attempts to resolve it with an Expert system (ES) based software architecture that assesses and utilizes drug-excipient and excipient-excipient relationships and results of pharmacopoeia tests data scattered in various forms of documentation and/or scientific literature. The inference engine of the ES operates with rule base and case-based reasoning, powered by Machine Reading Comprehension (MRC) and Natural Language Processing (NLP) technologies that populate and enrich the knowledge base. The MRC and NLP technologies interpret existing drug formulations and pharmacopoeia results to predict and propose potential new drug formulations with indicative quantities, based on their physicochemical characteristics and pharmacopoeia results. The research is based on an extensive literature review, primary research with surveys and interviews, and the analysis of several case studies to indicate the need for the proposed technology and support the system architecture design. Furthermore, the paper presents the pre and post-condition for adopting such technology, highlights research limitations, and identifies areas of further research to be conducted for the optimization of the technology and its contribution to the global economy and society.

Keywords: Excipient, Drug, Formulation, Artificial intelligence, Expert systems, Machine reading comprehension, Natural language processing, Pharmaceutical, Innovation, Technology

INTRODUCTION

The growth of the global population and several unpredicted political, health, and financial crises create an environment of uncertainty in which social innovations can be developed to offer stability in people's lives and create new business development opportunities for the benefit of the economy and society.

One of the undoubted rights of every human being is access to affordable medical treatment. Developing a pharmaceutical formulation quickly enough with secure quality is a complicated process that requires a

science-based systematic approach. Key information such as the properties of the drug substance and excipients, and interactions between the materials, equipment, and unit operations are necessary (Fathima et al., 2011).

Formulations consist of the active ingredient along with specific quantities of excipients, and their development requires specific manufacturing processes and operating conditions. The properties of the ingredients and potential interactions between them must be taken into consideration. This requires exploration in a design space where the connections between the properties are not well-defined. Experts' knowledge is difficult to explain quantify and transmit, so expert systems can be used to capture, analyze, enrich, and utilize this knowledge.

However, the costs and time needed for research and development on new or specialized drugs are not often covered by governmental budgets and initiatives that could make such medicines accessible to all who needed them. Private companies invest tremendous amounts and expect returns on their investments to reduce development costs, shorten development time, improve process design and efficiency, increase productivity, and achieve competitive advantage. The gap between drug availability and its accessibility creates the social need to accelerate the formulation development process and inspire advanced innovations to serve it.

LITERATURE REVIEW

Research indicates that the price of brand-name drugs can drop up to 80% after the commercialization of a new generic that has the same action and can potentially replace them. The global generic drug market value is projected to rise from $311.8 billion in 2021 to $531.84 billion in 2028 (Forecast, 2021; Gallagher, 2022).

Since the excipients form 90% of the pharmaceutical product, the choice of excipients in a generic formulation product is the most critical challenge. Excipients are accounting 0.5% of the total pharmaceutical market with a market value of $4 billion (Haywood and Glass, 2011) and the process of properly choosing them is crucial and time-consuming. Currently, drug formulation development can be briefly described with the following process. Excipients are selected based on the route of administration, physicochemical characteristics, place of action, and the release profile of the active ingredient. Each formulation is then tested for its fragility/hardness, dissolution, disintegration, dosage uniformity, and stability which are the prerequisite pharmacopoeia tests. The entire process is repeated when there is a change in the release rate, the route of administration, the strength of the active ingredient and the pharmaceutical form, or there is a new active ingredient/new variation of a biologic or excipient. Taking into consideration that only the dissolution test takes 10 hours for each formulation and on average 30 formulations are tested with different combinations of excipients in various pH and manufacturing conditions the process is considered laborious, costly, and time-consuming.

Research results indicate that the time to introduce a new medicine in the market can be reduced by 30% if there is an indicative formulation to start the process and potential results of the prerequisite tests. This 30% can save tremendous amounts of money from the reduction of research salaries, reduction of business operations costs, faster product delivery to the market, the time gained for the development of the next project, and other related benefits. Furthermore, reducing the tests that must be done can be considered as an environmental activity that decreases the environmental footprint. As a result, most pharmaceutical companies are amenable to investing in technology that will offer a clear advantage and accelerate the process compared to the other companies (Leigh Ann Anderson, 2022).

The above is verified with data collected from industry experts. The ELPEN pharmaceuticals R&D Director emphasized on having a possible formulation, to begin with, can remarkably drop the time needed for developing a formulation from two years to about one year and four months. The eight months gained can be used to market the product and rise the company's profitability. Likewise, the Verisfield R&D Formulation director indicates that manual lab work on generics' decisions, without any technological help, takes about 8–10 months and requires numerous and different costly tests done by expensive experts.

Governments worldwide are supporting generic pharmaceutical companies to continue and expand their operations (Forecast, 2021). Precisely, the US health system saved the last decade $2.6+ trillion due to generics with nearly 89% of more than 3.9 billion prescriptions in the last decade being filled with generic drugs (Figure 1) (Association for Accessible medicines, 2022). In Europe each year, more than $110 billion are saved because of the generics (Thepharmaletter, 2016). As a result, Governments are pleased to support the quicker progress of generic drugs as the long-term benefits for them are significant.

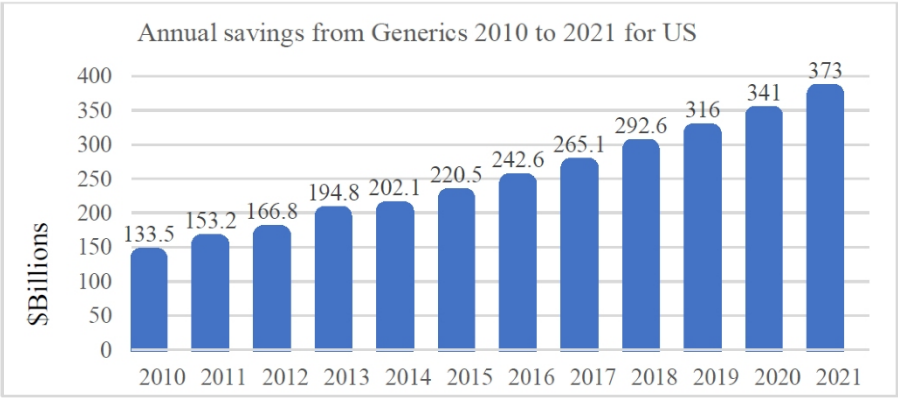

Figure 1: The annual savings for the US government due to generics from 2010 to 2021 (Association for Accessible medicines, 2022).

THE AI HEALTH MARKET

Despite the excipients market growth, almost all the Pharmaceutical companies and especially the generic pharmaceutical companies which must develop a new formulation similar to the original in less than two years, face challenges with the huge amount of data they must deal with (Lubis and Kartiwi, 2011). Nevertheless, there is nothing in the market to help them organize, categorize, and compare the data quickly and efficiently for the development of new formulations. AI can solve those needs and propose potential formulations and results on the prerequisite pharmacopoeia tests with several comparisons and analysis on the existing data.

The global market of AI was estimated at $119.78 billion in 2022 and is predicted to reach $1,597.1 billion by 2030 with a 38.1 % CAGR (PR, 2023). On the other hand, the AI Health market revenues are projected to grow from $6.9 billion in 2022 to $67.4 billion in 2027 reaching $188 billion by 2030 with a CAGR of 45.3% (CISION, 2021; GVR, 2021; Statista, 2022a). However, when there are many companies active in this field, the more competitive the market becomes. Nevertheless, research literature shows that over 90% of pharmaceutical companies are enthusiastic to invest in AI to rise the success rates of new drugs while reducing operational costs. As of today about $54 billion are saved from pharma companies due to AI (Canterbury, 2022; ITIF, 2022). It is estimated that more than 60% of the pharma companies globally plan to adopt AI technologies by 2030.

RESEARCH AND MARKET GAP

Despite the advancement of Artificial Intelligence, research indicates a gap in AI applications in the formulation development process (Figure 2) (Paul et al., 2021).

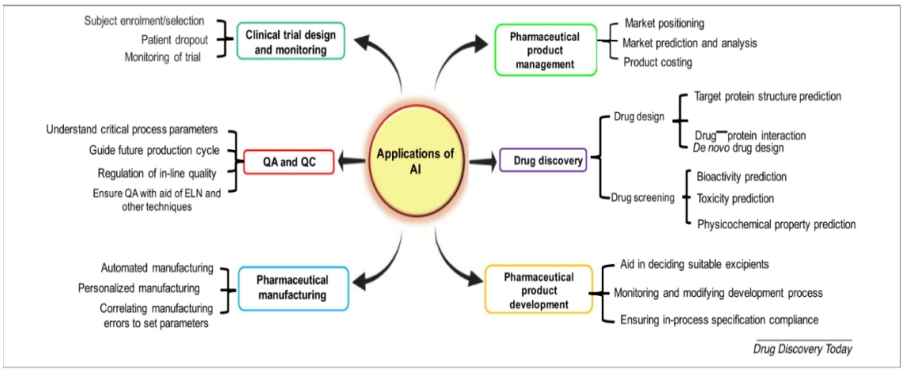

Figure 2: Absence of AI excipient selection applications from the contribution of AI in the drug development process (Paul et al., 2021).

Diving further into the AI application in the pharmaceutical sector and specifically, in the drug discovery sector, Table 1 indicates several AI tools used in drug discovery (Paul *et al.*, 2021). However, none of them is focused

Table 1. Examples of AI tools used in drug discovery (Paul et al., 2021).

Tools	Details	Website URL
DeepChem	MLP model that uses a python-based AI system to find a suitable candidate in drug discovery	https://github.com/deepchem/deepchem
DeepTox	Software that predicts the toxicity of total of 12 000 drugs	www.bioinf.jku.at/research/DeepTox
DeepNeuralNetQSAR	Python-based system driven by computational tools that aid detection of the molecular activity of compounds	https://github.com/Merck/DeepNeuralNet-QSAR
ORGANIC	A molecular generation tool that helps to create molecules with desired properties	https://github.com/aspuru-guzik-group/ORGANIC
PotentialNet	Uses NNs to predict binding affinity of ligands	https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507
Hit Dexter	ML technique to predict molecules that might respond to biochemical assays	http://hitdexter2.zbh.uni-hamburg.de
DeltaVina	A scoring function for rescoring drug–ligand binding affinity	https://github.com/chengwang88/deltavina
Neural graph fingerprint	Helps to predict properties of novel molecules	https://github.com/HIPS/neural-fingerprint
AlphaFold	Predicts 3D structures of proteins	https://deepmind.com/blog/alphafold
Chemputer	Helps to report procedure for chemical synthesis in standardized format	https://zenodo.org/record/1481731

on formulation development by excipient selection based on the physicochemical attributes of the drug or in the proposition of the results of the pharmacopoeia tests.

THE PHARMESDD TECHNOLOGY

Based on the excipient market growth and the costly and time-consuming tests that must be done for the medicine to enter the market there is an imperative need to develop an Expert system (ES) that would access and utilize drug-excipient relationship data and indicative results of the prerequisite tests scattered in scientific literature.

This research introduces PHARMESDD (PHARMaceutical Expert System Drug Development). An AI technology that integrates Machine Reading Comprehension (MRC) and Natural Language Processing (NLP) Expert System (ES) to interpret existing drug formulations using Machine Learning (ML), but also Deep Learning (DL), Cheminformatics, and Big Data to recommend results based on the physicochemical characteristics of the active ingredient and the route of administration. PHARMESDD is designed to suggest indicative proportions of excipients, potential results of the pharmacopoeia tests, ways of manufacturing the medicine, adverse interactions between drug-excipient and excipient-excipient, and identify if a patent covers the proposed formulation.

PHARMESDD utilizes open-access online databases to develop and improve its logic and thinking process. There are however indirect threats such as the Formulation diary, the FDA (Food and Drug Administration), the EMA (European Medicinal Agency), and the EMC (Electronic Medicines Compendium), which are free online databases that afford information for active ingredients and excipients, but data without intelligent technologies like PHARMESDD cannot be utilized efficiently. Such a technology, able to process thousands of accessible formulations can instantly propose potential formulations by reducing the cost, time, and effort needed for such output. Research conducted indicates that technology such as PHARMESDD can achieve tremendous time reductions in the drug discovery process, some of which are listed in Table 2.

Table 2. Pharmacopoeia tests time reduction that can be achieved with the use of related software.

Pharmacopeia tests	Time reduction for each test
Dissolution	10h to 2s
Disintegration	45min to 2s
Hardness/friability	15min to 2s
Uniformity of dosage units	15min to 2s

SYSTEM ARCHITECTURE OF THE PHARMESDD

PHARMESDD uses Expert Systems as the core artificial technology for its operations. The user opens a dynamic dialogue with the technology where formulation requirements are passed from the expert system to two other artificial intelligence technologies (the Machine Learning Comprehension and the Natural Language Processing subsystem) for the development of the output. In this process, data requests are made to accessible databases located in medical centers, open-access databases, universities, and on PHARMESDD data that has been collected from past experiences. The triangulated operations between the Expert System, the Machine Learning Comprehension and Natural Language Processing, and the databases are continuous until the expert system rules are completed enough to deliver a result with a high certainty factor (above 80%). During this process, any data imported from the databases are checked for physicochemical characteristics similarity, validated, and verified for completion, correction, and coherence.

Figure 3 presents a high-level system architecture of PHARMESDD. It must be noted that this is XAI technology (Explanatory AI) as the results given to the user are supported by a detailed explanation of the logic, data, actions, and reasoning followed to deliver them.

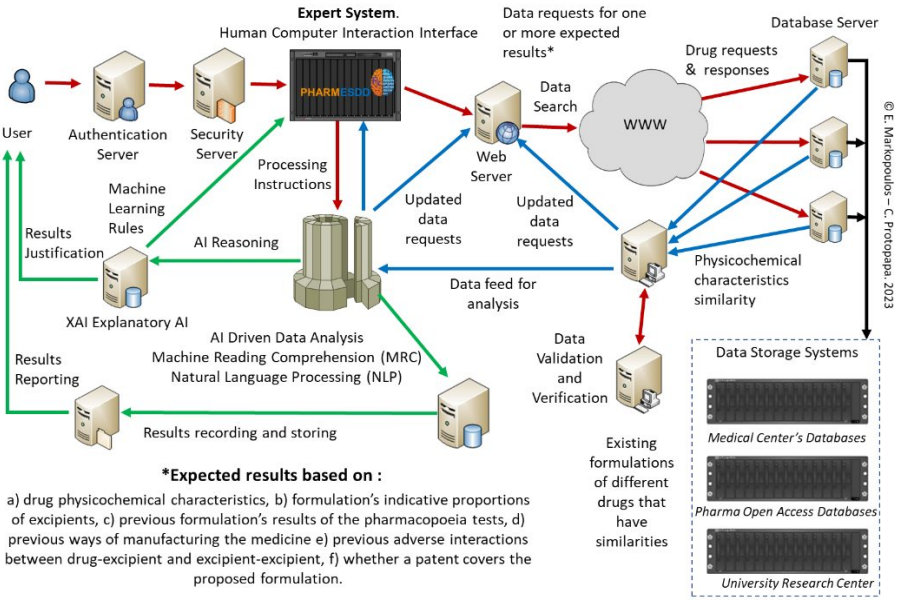

Figure 3: PHARMESDD system architecture.

SUSTAINABILITY AND TARGET MARKET

PHARMESDD's innovative design and the market gap in HealthTech AI lead to a blue ocean in terms of market innovation (Kim & Mauborgne, 2004) but also into a Green Ocean in terms of sustainable innovation (Markopoulos *et al.*, 2020a) and environmental protection. Several tests for the excipients' selection use organic solvents. The pharmaceutical industry spends hundreds of millions of dollars each year to manage millions of Kg of organic solvents. Therefore, reducing the needed tests becomes an environmentally friendly act and decreases the costs for pharmaceutical companies.

From a business suitability perspective, PHARMESDD can be primarily adapted by AI-powered organizations to identify the physicochemical properties of the excipients, organize the data into smaller groups, and connect inputs and outputs. These are usually large pharmaceutical companies that develop brand-name drugs, with a significant budget to spend and investments to do for the optimization of their operations. As of today, none of the technologies used offers possible formulations, potential methods to develop the formulations, potential interactions between excipient-excipient and drug-excipient, or results of the pharmacopoeia tests.

On the other hand, PHARMESDD can also be used by companies targeting the generics market as a Pink Ocean Strategy for social innovation that can democratize the low-cost drug development process (Markopoulos *et al.*, 2020b).

In general, PHARMESDD targets the global pharmaceutical industry and in particular the small to medium size companies that cannot afford to invest much in drug development. The global pharmaceutical market was estimated to be $1.42 trillion in 2021 (Statista, 2022b). Research indicates that Germany has more than 500 pharmaceutical companies leading the EU with 44.2 billion euros in revenue in 2020 (GVR, 2020). On the same note, there are 3,000 pharmaceutical companies and 10.500 pharmaceutical units in India (IBEF, 2022). Overall the global pharmaceutical market value in 2021 was $1.4 trillion (Statista, 2022c).

Even though the global number of pharmaceutical companies is not known, the numbers gathered in this research indicate that the critical mass needed for an innovation such as PHARMESDD to be successful is larger than expected for the potential success of the PHARMESDD technology.

LIMITATIONS AND AREAS OF FURTHER RESEARCH

The software architecture's design and identifying the market for PHARMESDD has been the first step in this research initiative. PHARMESDD as a concept and system architecture initially derived from two research projects, one of which studied the prediction of the excipients for biologic formulations using Natural Language Processing, and the other the market gap and commercialization of such technologies. These projects formed the base of this research, supported by primary and secondary research to identify an AI gap in pharmaceutical applications. Therefore, further research is needed to validate PHARMESDD algorithms and develop a working prototype that can

be tested within the industry to measure the effectiveness of the technology and the accuracy of the results.

In terms of testing PHARMESDD data will be used from the lab of Associate professor Dr. Marilena Vlachou at the Department of Pharmaceutical Technology of the University of Athens. The lab holds formulations for various drugs with many excipients combinations throughout its 15 years of operations. Specifically, the lab holds over 300 formulations for the modified release of a specific drug named Melatonin (MLT). This results in 300 combinations of over 30 excipients to be used by PHARMESDD in the pilot test phase. MLT has hypnotic properties in both animals and humans and is released at night. It is used extensively in aging jet lag and shift work where the circadian rhythm has been deregulated. Melatonin showed that can be used to improve the symptoms caused by SARS-CoV-2 when used at an early stage as it is well known for its immunomodulatory, antioxidant, and anti-inflammatory (Vlachou *et al.*, 2022).

CONCLUSION

The evolution of technology offers a wide range of applications to be developed across industries and societies. The health sector hosts many advanced innovations, ambitious plans, and state-of-the-art technologies and initiatives. Artificial Intelligence has always been closely related to the health sector since its early years with the development of the MYCIN and DENDRAL expert systems (Lindsay *et al.*, 1993). This research builds on the AI innovation heritage in the health sector with a contribution that can revolutionize the drug development industry.

PHARMESDD attempts to democratize drug development for those in need, help organizations that aim for compliance and alignment of their operations with the ESG criteria and the United Nations 2030 Sustainable agenda (Markopoulos *et al.*, 2021a), and contribute to health crisis resolution strategies with knowledge sharing practices (Markopoulos *et al.*, 2021b). PHARMESDD can be considered a sustainable and social innovation research. It combines environmentally friendly operations with social business impact in a circular economy that can benefit all who decide to utilize it.

REFERENCES

Association for Accessible medicines (2022) The U. S. Generic & Biosimilar Medicines Savings Report. Accessible Meds website: https://accessiblemeds.org/resources/reports/2022-savings-report

Generic-Biosimilar-Medicines-Savings-Report.pdf.

Canterbury (2022) Growth of AI in pharma. Canterbuty Website at: https://canterbury.ai/growth-of-ai-in-pharma/

CISION (2021) Global AI in Healthcare Market (2021 to 2026) - by Sections, Diagnosis, End-user, and Geography. PR News website: https://www.prnewswire.com/news-releases/global-ai-in-healthcare-market-2021-to-2026-\/-\/-by-sections-diagnosis-end-user-and-geography-301314928.html

Fathima, N. et al. (2011) Drug-excipient interaction and its importance in dosage form development, Journal of Applied Pharmaceutical Science, 1(6), pp. 66–71.

Gallagher, A. (2022) Global Generic Drug Market Forecast to Hit $531.8 Billion by 2028, Pharmacy Times. MJH Life Sciences (August 2022), 88. Pharmacy Times website: https://www.pharmacytimes.com/view/global-generic-drug-market-forecast-to-hit-531-8-billion-by-2028.

GVR (2020) Germany Pharmaceuticals Market Size, Share & Trends Analysis Report By Drug Class (Anti-cancer, Anti-viral, Anti-diabetics, Anti-rheumatics), By Type, By Formulation, By Application, And Segment Forecasts, 2020 - 2027. Grandview research website: https://www.grandviewresearch.com/industry-analysis/germany-pharmaceuticals-market.

GVR (2021) Artificial Intelligence In Healthcare Market Worth $120.2 Billion By 2028: Grand View Research, Inc. PR News wire website at: https://www.prnewswire.com/news-releases/artificial-intelligence-in-healthcare-market-worth-120-2-billion-by-2028-grand-view-research-inc-301302563.html#: ~: text=In-Language News-, Artificial Intelligence In Healthcare Market Worth %24120.2 Billion, %3A Gran.

Haywood, A. and Glass, B. D. (2011) 'Pharmaceutical excipients - where do we begin?', Australian Prescriber, 34(4), pp. 112–114. doi: 10.18773/austprescr.2011.060.

IBEF (2022) Indian pharmaceutical industry. IBEF website: https://www.ibef.org/industry/pharmaceutical-india.

ITIF (2022) Fact of the Week: Artificial Intelligence Can Save Pharmaceutical Companies Almost $54 Billion in R&D Costs Each Year. ITIF: https://itif.org/publications/2020/12/07/fact-week-artificial-intelligence-can-save-pharmaceutical-companies-almost/.

Leigh Ann Anderson (2022) Generic Drug FAQs. Drugs.com website: https://www.drugs.com/article/generic_drugs.html.

Lindsay R. K., Buchanan B., Feigenbaum E., and Lederberg J. (1993). DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation. Artificial Intelligence 61, 2: 209–261.

Lubis, M. and Kartiwi, M. (2011). 12. Data Management Challenges in Pharmaceutical Industry, (July 2012). PR (2023) Artificial Intelligence (AI) market. Precedence research website: https://www.precedenceresearch.com/artificial-intelligence-market#: ~: text=The global artificial intelligence (AI, USD 147.58 billion in 2021.

Kim W. C. and Mauborgne R. (2004). 'The Blue Ocean Strategy'. Harvard Business Review October 2004. Pages 76-84

Market Data Forecast (2021) Generic drugs market. Market Data Forecast website: https://www.marketdataforecast.com/market-reports/global-generic-drugs-market

Markopoulos E., Staggl A., Gann E. L., Vanharanta H. (2021a) Beyond Corporate Social Responsibility (CSR): Democratizing CSR Towards Environmental, Social and Governance Compliance. Advances in Creativity, Innovation, Entrepreneurship, and Communication of Design. AHFE 2021. Lecture Notes in Networks and Systems, vol 276, pp. 94–103. Springer. https://doi.org/10.1007/978-3-030-80094-9_12

Markopoulos E., Kirane I. S., Balaj D., Vanharanta H. (2021b) Organizing Global Democratic Collaboration in Crisis Contexts: The International Triangulation System. Advances in Creativity, Innovation, Entrepreneurship, and Communication of Design. AHFE 2021. Lecture Notes in Networks and Systems, vol 276, pp 206–213. Springer. https://doi.org/10.1007/978-3-030-80094-9_25

Markopoulos E., Kirane I. S., Piper C., Vanharanta H. (2020a) Green Ocean Strategy: Democratizing Business Knowledge for Sustainable Growth. In: Ahram T., Karwowski W., Pickl S., Taiar R. (eds) Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing, chapter 20, pp.115–125. vol 1026. Springer, Cham. https://doi.org/10.1007/978-3-030-27928-8_19

Markopoulos E., Ramonda M. B., Winter L. M. C., Al Katheeri H., Vanharanta H. (2020b) Pink Ocean Strategy: Democratizing Business Knowledge for Social Growth and Innovation. In: Markopoulos E., Goonetilleke R., Ho A., Luximon Y. (eds) Advances in Creativity, Innovation, Entrepreneurship and Communication of Design. AHFE 2020. Advances in Intelligent Systems and Computing, pp. 39-51, vol. 1218. Springer, Cham. https://doi.org/10.1007/978-3-030-51626-0_5

Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discovery Today. 2021; 26(1): 80–93. Available from: https://doi.org/10.1016/j.drudis.2020.10.010

Statista (2022a) Artificial intelligence (AI) in the healthcare market size worldwide from 2021 to 2030. Statista website: https://www.statista.com/statistics/1334826/ai-in-healthcare-market-size-worldwide/.

Statista (2022b) Global pharmaceutical industry - statistics & facts.

Statista (2022c) Revenue of the worldwide pharmaceutical market from 2001 to 2021.: https://www.statista.com/statistics/263102/pharmaceutical-market-worldwide-revenue-since-2001/#:~:text=As of end-2021%2C the, what people pay for medication.

Thepharmaletter (2016) Generic drugs "save European 100 billion euros per year". Available at: https://www.thepharmaletter.com/article/generic-drugs-save-european-100-billion-euros-per-year.

Vlachou, M. et al. (2022) 'Modified Release of the Pineal Hormone Melatonin from', Polymers for Advanced Technologies.