The potential impact of Machine Learning and Artificial Intelligence on Clinical Trials in the Irish Pharmaceuticals sector

Research dissertation presented in partial fulfilment of the requirements

for the degree of

MSc in Pharmaceutical Business & Technology

Griffith College Dublin

Dissertation Supervisor: Mark Campbell

Sisira chellurkupadan

2023

Candidate Declaration

Candidate Name: Sisira Chellurkupadan

I certify that the dissertation entitled: The potential impact of Machine Learning and Artificial Intelligence on Clinical Trials in the Irish Pharmaceuticals sector submitted for the degree of: **MSc in Pharmaceutical Business & Technology** is the result of my own work and that where reference is made to the work of others, due acknowledgment is given.

Candidate signature:

Date: 16/05/2023

Supervisor Name: **Mark Campbell**

Supervisor signature:

Date:

Acknowledgements

This endeavor would not have been possible without the constant support from my supervisor Mr. Mark Campbell. I would like to express my deepest appreciation to Mr. Mark Campbell for his constant support and guidance throughout the dissertation process.

I am also grateful to my family for their immense support, love, and encouragement throughout the completion of this study.

I would like to extend my sincere thanks to my friends in Ireland and India for their support during the dissertation process.

Dedication

I dedicate this dissertation to God and my mother.

Table of contents

List of Tables

List of Figures

List of Abbreviations

ACTT	Adaptive COVID-19 Treatment Trial
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
AI	Artificial Intelligence
ANNs	Artificial Neural Networks
BHMM	Bayesian Hierarchical Mixture Model
BHMs	Bayesian Hierarchical Models
BLRM	Bayesian Logistic Regression Model
CDSS	Clinical Decision Support Systems
ChEMBL	Chemical database of bioactive molecules
COVID-19	Coronavirus Disease 2019
CSO	Contract Sales Organization
CT	Clinical Trial
DL	Deep Learning
DNN	Deep Neural Network
EC50	Half maximal Effective Concentration
ECGs	Electrocardiograms
EEA	European Economic Area
EHRs	Electronic Health Records
EMA	European Medicines Agency
EMR	Electronic Medical Records

EU	European Union
EUR	The Euro
EXNEX	Exchangeability-Non-Exchangeability
FDA	Food and Drug Administration
HCPs	Health Care Practitioners
HTS	High-Throughput Sequencing
kNN	k-Nearest Neighbors algorithm
LBVS	Ligand-based virtual screening Ligand-based VS
LO	Lead optimization
MCMC	Markov Chain Monte Carlo
ML	Machine Learning
mTPI	Modified toxicity probability interval
NLP	Natural Language Processing
QSAR	Quantitative Structure-Activity Relationship
RAR	Rapid Assessment Response
RNA-seq	Ribonucleic Acid Sequence
RNN	Recurrent Neural Network
RQ	Research Question
SAR	Structure-Activity Relationship
SARS	Severe Acute Respiratory Syndrome
Sosa-D	Selective optimization of side activities of drug molecules
SVM	Support Vector Machines

UN	United Nations
US	United States
USD	United States dollar
VS	Virtual Screening
WHO	World Health Organization

Abstract

The potential impact of Artificial Intelligence and Machine Learning on Clinical Trials in the Irish Pharmaceuticals sector

Sisira Chellurkupadan

Concept: Artificial Intelligence (AI) has been widely purported to be used throughout stages of the drug development process to identify novel targets, enhance understanding of disease mechanisms, and develop new biomarkers. AI, coupled with Machine Learning (ML), is reported to have the potential to transform pharmaceutical research and development over the coming decade, particularly in the area of Clinical Trials (CT).

Aim: The aim of this study is to examine the use of Artificial Intelligence and Machine Learning in Clinical Trials in the Irish Pharmaceuticals sector. Each new drug brought to the market by the Irish pharmaceutical sector typically costs millions of Euros and requires more than ten years of development. The expense and length are caused by the processes of identifying and testing of chemical entities that could be therapeutic. There may also be the fact that it has not been specifically studied or explored in the Irish context before.

Methods: The research method planned was to conduct online, semi-structured interviews with 20 professionals in the Irish pharma sector. The methodology intended to collect mainly qualitative data to answer the research questions. Due to a poor response rate within the planned an available time-frame, the method was changed to data collection via an online, written, structured questionnaire. This change had a significant effect on the quantity and type of data collected and the time available for analysis reduced the reliability and validity of the data analysis planned, and hence findings and conclusions reached.

Findings: The study found that the use of AI/ML in clinical trials has the potential to reduce the duration of the drug development process, from the identification of novel targets to the development of new biomarkers. By automating certain aspects of the clinical trial process, AI/ML can also improve the accuracy and efficiency of data collection, analysis, and interpretation. This can lead to more effective and targeted drug development, ultimately resulting in faster delivery of end-drugs to patients at lower costs.

Conclusions: The study that investigates the impact of artificial intelligence and machine learning on the clinical trials of Irish pharmaceutical companies found that the use of AI/ML reduces the duration of the stages of the drug development process to identify novel targets, enhance understanding of disease mechanisms, and develop new biomarkers and enables delivery of end-drugs to patients more quickly and at lower costs.

1. Introduction

1.1 Overview

This introduction chapter acts as the entry point for this research work. The concise summary of background, aims, objectives, research questions, and overall organisation of the dissertation are all described in this chapter. The study's goal is to examine how artificial intelligence (AI) and machine learning (ML) are being used in the Irish pharmaceutical industry and how they are used in drug development, particularly in clinical trials. The importance of this study issue and its prospective effects on the pharmaceutical business are also highlighted in this chapter. In the final section of the introduction chapter, a detailed description of the dissertation is offered, giving readers a clear roadmap of what to expect in the following chapters.

1.2 Background

The pharmaceutical business is an important sector that contributes significantly to global health by giving patients all around the world access to vital medications and treatments (Saxena *et al*., 2022). Clinical trials are an integral element of the medication development process because they show a treatment's efficacy and safety before regulatory authorities can authorise it for usage. However, conducting clinical trials can be costly, time-consuming, and difficult due to the fact that many of them fall short of their goals or fail to receive regulatory approval (Fogel, 2018).

Technologies based on AI and ML are showing promise as potential remedies for some of the difficulties clinical trials encounter (Davenport and Kalakota, 2019). Clinical trial design can be improved by utilising AI and ML, including patient cohort selection, recruiting, monitoring, and data analysis. This will increase the effectiveness, speed, and efficiency of clinical trials (Weissler *et al.*, 2021). According to the authors, preclinical investigation and planning are crucial for designing successful clinical trials, during which candidate molecules

and targets are identified, and a strategy for achieving regulatory approval is determined (Mohs and Greig, 2017). ML can be employed to identify, recruit, and retain participants more efficiently and fairly. ML and AI have significant potential applications in clinical research, from preclinical drug discovery to data analysis and reporting (Paul *et al.*, 2021; Weissler *et al.*, 2021). However, there is a shortage of skilled workers, and there are few prospective studies evaluating ML's effectiveness in comparison to conventional methods (Harrer *et al.*, 2019). Moreover, the findings of traditional clinical trials are challenging as these trials may not easily transfer into making personalized treatment decisions at the usual point-of-care (Weissler *et al.*, 2021). To address this issue, the authors recommend utilizing new digital clinical endpoints and treatment response biomarkers that can be closely monitored, improving safety and efficacy, and gaining more insight into the patient journey through sensors and low-cost imaging (Shah *et al.*, 2019). Overall, these studies emphasize the potential of AI and ML in clinical research and suggests several ways to overcome the challenges associated with their implementation.

This research explores the impact of AI and ML in the clinical trials of the Irish pharmaceutical sector. Irish pharmaceutical businesses have significantly increased the investments in their research and development (R&D) initiatives (O'Dwyer *et al.*, 2017). The clinical research and drug development process is exceedingly expensive and protracted process. Many ventures just failing after investing a considerable sum of money.

1.3 Research purpose

This study concentrates the impact of computational methods such as artificial intelligence and machine learning on the Drug research and development. The purpose of this study is to evaluate the significance and impact of AI and ML in the field of drug development in the Irish pharmaceutical industry. Drug research and development involves the

recognition of drug targets, the authentication of the targets, hit-to-lead fructification, lead refinement, preclinical molecule determination, preclinical evaluation, and clinical testing. In order to bring a new prescription drug to market, approximately USD 2.6 billion is required, which takes approximately 10 to 15 years. The development of computer-enabled drug design technology has been praised as the most inventive way to change this tragic situation, but it depends on proper planning during the development process (Sarkar *et al.*, 2023). The use of AI and DL in this discipline is strengthened by historical evidence. It is important to analyse the professional's perspectives on artificial intelligence and machine learning in the success of drug development, clinical trial data management and clinical trial data analysis. The research method employed originally intended to interview professionals in the Irish pharma sector and obtain data from a verbal discussion, but this later changed to collecting a more limited amount and type of data. Due to the limited timeframe and unavailability of participation in the research, only seven professionals were used and two zoom interview methods were used. The interview method was later changed into structured questionnaire. Online, structured questionnaire was conducted with seven professionals in the Irish pharma sector. The participants were selected based on their expertise and experience in clinical trials, drug development, and the use of AI/ML. It will be possible to determine the impact of AI and ML in the Irish pharmaceutical sector by combining interviews, questionnaires, and literature study.

1.4 Aims and objectives

1. To determine, through a secondary data review of relevant literature and research on the topics of AI and ML, the potential of utilizing these techniques to increase the success of drug development, including but not limited to drug discovery, translational research, and the pre-clinical phase.
2. To review the reported impact of AI/ML specifically and critically in clinical trial operations and clinical trial data analysis reported in

the literature.

3. To examine and explore attitudes, opinions, and behaviours towards AI/ML in CT among a cross-section of professionals in the Irish pharma sector.
4. To understand the application of Nonparametric Bayesian learning (a tool in the ML framework) in CT design

1.5 Research questions

1. What effects do artificial intelligence and machine learning have on the pre-clinical stage, translational research, and drug discovery in particular?
2. what are the distinctive and important impacts of AI/ML on clinical trial operations and clinical trial data analysis?
3. What are the thoughts, opinions, and actions of Irish pharma industry experts have about AI/ML in CT?
4. How is nonparametric Bayesian learning—one of the tools in the ML framework—applied in CT design?

1.6 Outline of the study

The Literature review gives the detailed review of the impact of AI and ML in clinical trials in the Irish pharmaceutical industry. The literature review states the effects of AI/Ml on the pre-clinical phase, enrollment of study population, clinical data operations and clinical data analysis. The Research methodology consists of research design, method for gathering the primary research data, description of target participants and the conceptual framework. A late change to the form of the Conceptual Framework was necessary due to a low response rate from those invited to interview, requiring a move to collecting data which saw interviews replaced by process a structured, online, written questionnaire. Qualitative research is carried out for the data collection. The interviews and online, written, structured questionnaires are conducted with the professionals in the Irish pharmaceutical industry through zoom meetings and recorded for the data analysis. The change in data collection method mentioned previously also had an effect on the type of subsequent data analysis open to the researcher. The findings and conclusion describe the findings and analysis from

primary research. Diagrams, figures, and tables are used to represent the data. Recommendations for further research, limitations of the study are provided. The advantages of AI/ML in the pre-clinical trials, patient enrollment, clinical trial operation and data analysis are described. The pre-clinical trials, clinical trials, patient enrollment, clinical trial data management become easier, shorter, and accurate with the exploitation of the AI/ML.

2 Literature Review

2.1 Introduction

This chapter reviews the conceptual and empirical (research) literature related to the potential impact of Artificial Intelligence and Machine Learning in Clinical Trials. It includes a critical evaluation of their use globally and examines their historic and current use in the Irish Pharmaceuticals sector. The concept of artificial intelligence (AI) was first introduced in 1956. It referred to machines that could learn as they gathered and processed information, giving them the capacity to "think" like humans. The area of artificial intelligence encompasses a variety of disciplines, such as pattern recognition, clustering, similarity-based methods, statistical analysis, machine learning. Machine learning and artificial intelligence (AI) techniques have shown remarkable potential in various sectors of the pharmaceutical industry. Clinical trials, the crucial step in drug development, have not been an exception in this regard. The increasing complexity of clinical trials and the need for faster, more efficient, and cost-effective drug development has led to a growing interest in AI and ML techniques. This literature review aims to explore the potential impact of AI/ML in clinical trials in the Irish Pharmaceuticals sector, including drug discovery, translational research, and the pre-clinical phase. Additionally, it will critically review the impact of AI/ML in clinical trial operations and clinical trial data analysis. Lastly, this review aims to understand the application of Nonparametric Bayesian learning, a tool in the ML framework, in clinical trial design.

2.2 Transformation in healthcare

The world is under several changes and lot of transformation in the field of technology and even in the lives of people. Keeping the life expectancy as a point of concern, a main change has been anticipated to occur despite it has been developing in the previous decades. Nearly 1.5 billion people are anticipated to attain 65 years + of age by 2050, i.e., more than twice of what was in the year 2019. As per the estimation

of the World Health Organization (WHO) and the United Nations (UN), most of the diseases are associated with comorbid, chronic conditions. These kinds of conditions generally pave way for an increased and persistent focus from the healthcare professionals. As a result of ageing population, the healthcare cost is as well anticipated to have a radical growth and this is an assured fact.

Therefore, there is a necessity of getting adapted to these changes and the healthcare processes should as well make sure to cope with the rising demand in terms of healthcare service (Bohr and Memarzadeh, 2020a). Contrary to other industries, public companies like healthcare are extremely ahead in using technology at the best of their need and ability. Since the technologies innovations are keep on increasing, there are breakthroughs experienced in both data collection as well as treatments. Making use of this massive data for discovering and examining the correlations amid them is clearly lagging behind. Healthcare is open for disruptive changes which shall happen with the advancement of technology. As per the study of (Davenport and Kalakota, 2019; Bohr and Memarzadeh, 2020b), the transformation is focused at proactive healthcare, whereas in the present times it is vitally reactionary with either few or zero focus on prevention.

2.2.1 Trends in Irish pharmaceutical sector and Healthcare

The tiny country that generates 5% and more of the pharmaceuticals in the world had witnessed a shocking 25% growth in terms of value-added output in the year 2021. The persistent vaccine production along with the international uptick in non-Covid-related healthcare expenses are positives for the intensely export-based pharmaceutical sectors of Ireland.

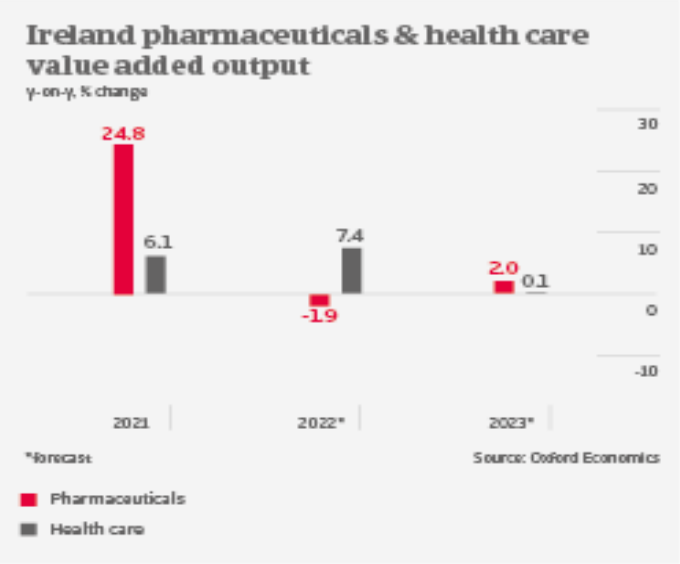

Figure 1 Ireland pharmaceuticals & healthcare value added output (Atradius, 2022)

Because of bigger profitability in the drug field, greater turnover has led to greater net earnings for distributors as well as wholesalers. When there is a decrease in the sales of greater margin products (perfumes, cosmetics, etc), drugstores and pharmacies were facing a major suffering in the year 2020 and 2021 owing to lesser food traffic. There is yet a sizable backlog in the industry of domestic healthcare for non-urgent techniques and consultant appointments due to pandemic outbreak. The Irish pharmaceutical industry, which consists of research, development, manufacturing, marketing, and distribution, is heavily dependent on exports, and in 2021, value-added output increased by 25% as a result of continued vaccine production and an increase in non-Covid-related healthcare costs. There will be an increase in the corporate tax from 12. 5%-15% as in the year 2023 for global pharmaceutical firms which conducts its operations in the region of Ireland along with EUR 750 million (USD 80 million) and more revenues. Furthermore, the increasing lapse of well skilled and talented employees has resulted in wage inflation (GlobalData, 2021). Ireland is a crucial centre for drug production; multinational firms will get profited from increasing global demand in the upcoming times (CSO, 2022).

The pharmaceutical business is not an exception to the growing trend in the EU and Ireland of using AI and ML in healthcare. These innovations could lead to better patient outcomes, more effectiveness, and lower costs. But there are also worries about security, privacy, and moral matters (Atradius, 2022).

The foreign direct investment from the parent companies is considered to be the fundamental source of funding for the manufacturers of pharmaceutical products (Ho *et al*., 2017). Smaller companies can acquire bank loans and there is one more effective option called private equity finance. The average for payment duration in the industry is two months and over the past two years, there has been extra-ordinary payment behaviour (Atradius, 2022). The range of insolvency is extremely low than expected over the business in Ireland. The pharmaceutical firms find the underwriting approach to be highly flexible due to the minimum credit risk and common economic stability of companies. It is the same for distributor as well wholesaler sector, as there has been substantial reduction in amount of market participants due to the recent mergers and takeovers. Owing to certain restrictions for pharmacy stores and pharmacies the sales is yet to return to its pre-pandemic range.

2.3 Transformation in pharmaceutical sector

Over the past few decades, the pharmaceutical sector has undergone a considerable shift, with technical advancements and changes in rules and legislation having a substantial impact on drug distribution, drug discovery, and the field's general operations. Clinical trial automation is currently increasing in Ireland's pharmaceutical industry, and businesses like Nuritas and APC are creating and manufacturing personalised medications. Ireland is also experiencing expansion in the field of digital health, and Irish pharmaceutical firms are exploiting the country's fledgling healthcare industry as a proving ground for their latest innovations. Irish regulatory frameworks are connected to US

and European FDA procedures, and supply chain optimisation in Europe and Ireland specifically benefits from AI and ML.

AI and ML are the two important technical breakthroughs, have arisen in recent years (Saeed *et al.*, 2022). By advancing drug research, development, and distribution, these trends are altering the pharmaceutical business and have the potential to improve patient care by enabling personalised medicine. The growing automation of clinical trials in drug development is another trend that has caught on in the market. Clinical trials have become more effective and efficient thanks to automation in clinical trials, which has sped up access to life-saving drugs. Real-time data collection in clinical trials is made possible by the use of automation and digital technologies, which can speed up the drug development process.

The pharmaceutical industry has gone through a crucial transformation over the last few decades because of different technological developments and changes in the law and regulations. This transformation has had a crucial effect on the drug distribution and drug development along with the entire operations of pharmaceutical field (Liang *et al.*, 2020). Here are few of the vital areas where transformation has been witnessed in the pharmaceutical industry.

2.3.1 Recent changes in Drug Development

The process of drug development is noted to be expensive and long process which involves years of development and analysis accompanied by regulatory approvals and clinical trials (Badria, 2020). In the recent technological developments like artificial intelligence and machine learning there has been crucial reduction in the timeline of drug development through predicting the toxicity and efficiency of drug in the early developmental stage (Liang *et al.*, 2020). This has resulted in more productive as well cost-efficient drug development.

Manufacturing is quite consistent in the present times instead of batch-based and testing is preferably conducted in order to reinforce the range

of efficacy during manufacturing process with the help of tools such as sensors. Through AI-enabled technologies, investigators can carry out their analysis about the present and past trials, examine big data and modify the upcoming trails in accordance with the circumstance. Modern digital technologies like next-generation sequencing can be used to acquire bigger knowledge and understanding about the disease mechanism in huger set of patients and simultaneously create personalized options of treatment (Weissler *et al.*, 2021). A pharmaceutical product can be effectively manufactured from the bench to bedside using Artificial Intelligence. This is due to the fact that AI is useful in decision making, making rational drug design and also to handle clinical data for the drug development in the upcoming period of time. Through artificial intelligence, the compounds can be hit and led, the drug target can be validated more rapidly and the drug structure design can be optimized in a more productive way (Paul *et al.*, 2021).

2.3.2 Growth in personalised medicine

The rise of personalized medicine has transformed the pharmaceutical sector by allowing doctors to tailor treatment plans based on a patient's genetic makeup (Bhuskute *et al.*, 2021). This has led to more targeted and effective treatments, resulting in improved patient outcomes.

2.3.3 Key trends in digital health

The increasing use of digital health technologies, such as wearables and mobile health apps, has transformed the way healthcare is delivered (Lin *et al.*, 2020). These technologies have enabled patients to take a more active role in managing their health and have facilitated remote patient monitoring, leading to more efficient and cost-effective healthcare.

2.3.4 Regulatory Changes

Regulatory changes in the pharmaceutical sector have had a significant impact on drug development and distribution (Dubois and Sæthre,

2020). For example, the introduction of the FDA's Breakthrough Therapy Designation has led to faster approval of drugs for serious or life-threatening conditions, allowing patients to access treatments more quickly. The European prescriptions administrative framework depends on an organization of around 50 administrative specialists from the31 EEA nations (28 EU Part States in addition to Iceland, Liechtenstein and Norway), the European Commission and EMA. This organization makes the EU administrative framework novel. The organization is upheld by a pool of thousands of specialists drawn from across Europe, permitting it to source the best conceivable logical ability for the guideline of medications in the EU and to give logical guidance of the greatest quality.

2.3.5 Supply Chain Optimization

The pharmaceutical supply chain has become increasingly complex over the years, with drugs often passing through multiple intermediaries before reaching the end consumer (Shashi and Gossett, 2022). However, recent technological advancements, such as blockchain, have made the supply chain more transparent and efficient, reducing the risk of counterfeiting and improving drug safety (Saeed *et al.*, 2022). Overall, the transformation of the pharmaceutical sector has had a significant impact on the way drugs are developed, distributed, and consumed. As technology continues to advance, it is likely that the sector will continue to evolve, leading to further improvements in healthcare delivery and patient outcomes.

2.4 Artificial Intelligence

There is no commonly accepted definition for AI, and this is mainly due to the fact that intelligence does not have a common definition (Russell and Norvig, 2016). AI is the term used to describe computers that are capable of performing tasks that are generally performed by intelligent beings, including as interpreting natural language,

recognising images, making decisions, and learning from experience. It is stated that the term is generally implemented in the developing systems which possess human qualities to have reasoning capability, generalizing, finding meaning or learning from the previous experience. This is done with the help of massive data (Luke *et al*., 2020). As human capabilities are rooted in an extensive sense, it is as well possible to define AI like represented in the Figure 2 (Wu, 2019).

The application of AI in the pharmaceutical sector has greatly increased recently, with numerous businesses looking into its potential in areas including drug research, clinical trials, and supply chain management (Walch, 2020). For instance, Pfizer has employed AI to forecast clinical trial outcomes by examining data from prior studies (Sharma *et al.*, 2022). In order to better understand the efficacy and safety of its medications, Sanofi has employed AI to examine real-world patient data (Sanofi, 2022). AI is also utilised by the pharmaceutical sector in Ireland. In order to find trends and enhance patient outcomes, the Health Research Board, for instance, is employing AI to analyse patient data (HRB, 2021). In addition, Deciphex, an Irish start-up, won €2.3 million in financing in 2020 to create AI-based tools for examining medical photos (Taylor, 2019).

However, using AI in the pharmaceutical sector also has drawbacks, including ethical issues, legal issues, and privacy concerns (Naik *et al.*, 2022). Despite this, the pharmaceutical industry is likely to continue investigating AI's usage in many parts of medication research and delivery given the potential benefits it may provide.

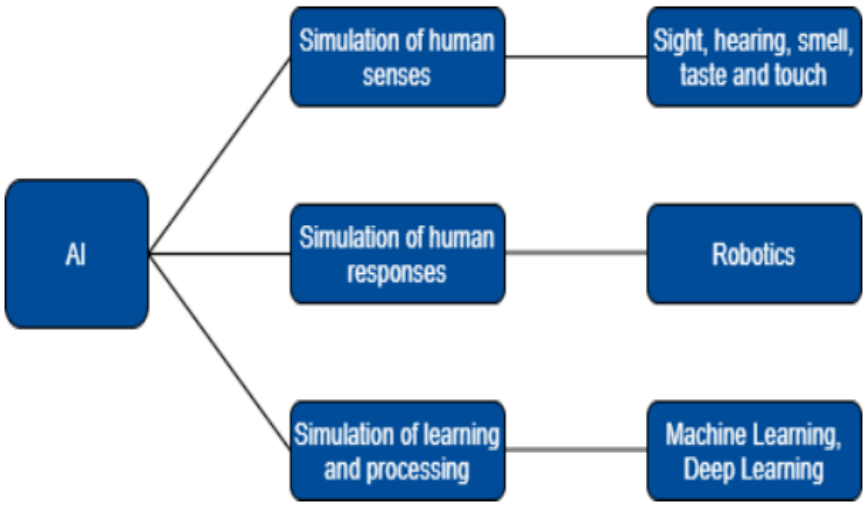

Figure 2: AI definition, (Wu, 2019)

2.5 Machine learning

Machine Learning is nothing but a part of AI. It is utilized for getting a knowledge or thorough understanding about the data and discovering patterns. These patterns can be possibly utilized in understanding various health contexts, for example: detect a diagnosis on the basis of input data. ML is a technique where mathematical models are used in the program on the basis of statistics and possibility to learn and train data without using traditional programming. Learning is acquired by using learning algorithms, where the potential of learning is described by ML through discovering patterns in huge datasets. A dataset contains many data points and all of these points tend to show a unit that need analysis. This learning process of these algorithms can be categorized in various kinds.

Making use of the input variables and linking them with a predetermined output is one type. This is a time-consuming method as the data is labelled in a manual way. Supervised learning is quite appropriate to detect associations between the defined result (as output) and patient's traits (as input) in terms of healthcare. When the labelled data is preferred to be utilized as a foundation, the learning program/algorithm can in turn be injected with unlabelled data so that

assumptions can be made. There are several learning models or classification among ML, however natural language processing (NLP), deep learning (DL) and artificial neural networks (ANNs) are considered as the most familiarly used ones among those (Bohr and Memarzadeh, 2020a).

2.6 Potential of AI/ML Techniques to Increase Success in Drug Development

According to the study of Lipinski *et al.* (2019), from the last twenty years, a huge challenge has been imposed amid the biological and chemical scientists due to the development of advanced and effective systems for the targeted offering of therapeutic agents with minimal risks and maximum effectiveness. In addition to that, the time consumption and advancement in developing novel therapeutic agents was considered as new setback in the development process and drug design (Hamet and Tremblay, 2017). Investigators around all parts of the world have preferably shifted towards computational methods like molecular docking and virtual screening (VS) that are as well called as traditional approaches. These methods as well tend to impose challenges like ineffectiveness and inaccuracy (Hassanzadeh *et al.*, 2019). Therefore, there is rush in applying novel techniques that are quite self-adequate for removing the challenges witnessed in traditional computational methods. AI that includes ML and deep learning (DL) algorithms has come out as a potential solution that is capable of overcoming the issues and obstacles in the discovery process of drug and drug design (Duch *et al.*, 2007). In addition to that, designing and drug discovery contains complicated and long steps like validation and target selection, lead compound optimization and screening, clinical trials and pre-clinical trials and finally the practices of manufacturing. These steps tend to impose one more huge challenge in the recognition of proficient medication against any kind illness. Hence everyone in the pharmaceutical firms were raised with a huge question about how

the manage the speed and cost of the process (Zhang *et al.*, 2017). AI has answer for all those questions in the most scientific and simple way possible. This subsequently decreased the consumption of time as well as the cost of process. In addition to that (Jordan, 2018), stated that the rise of data digitalization in pharmaceutical firms and healthcare industry drove towards the AI implementation in order to overcome the issues of scrutinizing complex data.

The primary goal of clinical AI is to collect relevant data from high-dimensional, complicated data so that decision making can be possible. Clinicians are added rarely in data science crews that shall result in the reduction or limitation of explainability, quality enhancement, and clinical relevance of workflow compatibility of AI solutions. This extends the communication gap amid the developers and physicians (Koçak *et al.*, 2023). During the last three decades, clinical drug development has remained unmodified in most of the times.

This is indeed an outcome of different category of factors such as absence of appropriate actionable biomedical data sources and modern analytics to generate hypotheses which is potential of inspiring the advancement of novel treatments or diagnostics, risk aversion, scepticism and unpredictability's of regulatory needs about the rapidly growing yet hugely unproven technologies (like wireless health monitoring devices, sensors and machine learning) and uncertainties related to regulatory needs (Shah *et al.*, 2019). Because of the expansion, adaptation and advancements of excess amount of data accessible to create suitable insights, AI-enabled technology has been actively employed to meet tiny but significant troubling problems related to the development and drug departments.

Owing to the invention of microarray, high-throughput sequencing (HTS) and RNA-seq technologies, an excess range of biomedical data is generated every single day because of which modern drug discovery has resulted in the transition of huge data. The drug development process can be split into four phases such as clinical, pre-clinical, post-

marketing and drug discovery. ML/AI techniques have exhibited capability in every single phase but the main attention is placed in the initial three phases.

2.6.1 Use of AI/ML in drug discovery

AI innovation is predicted to permit several components of drug development and drug discovery and it turned into an idea for computer supported medication plans with persistent gathering of clinical data and enhanced AI calculations. The combined advancement of technological growth and mechanization must lead to the enhancement of analysis of complex and huge datasets. This will be needed for the acceleration of drug development, increase of success rates and lesser costs.

These technologies not alone are responsible in enhancing the effectiveness of the process but in certain cases, it even removes or decreases the necessity for clinical trials through conducting simulations in their area. They even pave way for the investigators to deeply examine molecules without trails that lessens ethical concerns and costs. However, integration of machine learning and AI is assumed to revolutionize the analysis of drug in the upcoming times. Yet there are some barriers which may occur in the midway like the necessity of clearing different datasets and unstructured datasets and infrequent incompetence of computer hardware. A new period for the pharmaceutical trade will start once after the increase of these obstacles that permits the growth of machine learning and AI which needs extensive improvement and usage (Patel and Shah, 2022).

Drug discovery is nothing but an initial stage of drug development, where new drug candidates are recognized. ML/AI techniques are used to examine extensive range of data from different sources such as genomics, metabolomics, chemical data, genetics, chemical data and proteomics. For example, Wang *et al.* (2021) preferred to use data learning models for predicting the protein-ligand binding affinities in most accurate form. This could significantly result in the reduction of

time and cost need for drug discovery. Due to the developments in automated drug discovery techniques involving ML and AI, it is considerably simple to differentiate amid the novel chemical structures and present drugs. For instance, He *et al.* (2019a) preferably implemented a computational method for screening the hepatotoxic ingredients in case of traditional Chinese medicines, on the other hand, He *et al.* (2019b) exhibited the structure-toxicity relationship, herb-ingredient network and phylogenetic relationship via computational method. In the recent times, Zhang *et al.* (2020) attempted to apply computational analysis against the corona virus, where various biologically active compounds have been screened by the authors to fight against severe acute respiratory syndrome (SARS). Subsequently the compounds were subjected to docking analysis and ADME. As per the study of Zhang *et al.* (2020), the outcomes came up with a conclusion that present Chinese traditional medicines were seemingly efficient again novel corona virus. Therefore, the concepts of conventional chemistry-based drug discovery and development coupled with computational drug designing offers an extensive range of research platform in the upcoming period of time. Furthermore, chemical scientists and system biologists from different parts of the world coordinated with the computational scientist to create contemporary ML principles and algorithms to improvise the drug development and discovery.

2.6.2 The role of AI/ML in the pre-clinical stage

Drugs are taken for testing in pre-clinical phase in vivo and vitro for examining their level of toxicity and effectiveness. ML/AI techniques can be employed for predicting effectiveness and toxicity on the basis of preclinical data. For instance, the OpenTox project prefers to apply ML/AI models for predicting the chemicals' toxicity depending on the molecular structure (Hastings et al., 2011). In addition to that, Julkunen *et al.* (2020) created a novel ML-driven tool named comboFM that

ascertain the dosage and drug combinations in pre-clinical researches such as cancer cell lines. comboFM detects proper drug combinations along with the dosage through the help of factorization machines- an ML framework for greater dimensional data analysis.

As per the research of Kolluri *et al.* (2022), natural language processing (NLP) is applied in pre-clinical space to grab the scientific insights from unstructured electronic medical records (EMR), biomedical literature and insurance claims which in turn helps to recognize novel targets. Following the same predictive modelling is applied for predicting protein structures and provide molecular compound design and optimization so that drug candidates with greater possibility of success can be chosen productively.

2.6.3 The role of AI/ML in the Clinical stage

ML/AI techniques can as well be helpful in clinical phase of drug development through assuming the effectiveness and safety of drugs. For example, clinical decision support systems (CDSS) can crucially examine patient data like electronic health record (EHRs) for predicting drug efficacy and drug safety (Sutton *et al.*, 2020). In addition to that ML/AI techniques can be useful in optimizing clinical trial design through recognizing the patient populations which has more possibilities of showing positive responses to the drug (Weissler *et al.*, 2021). Few studies have exhibited that ML/AI are preferably utilized for observing clinical trials. Sertkaya *et al.* (2016) asserted that trial site monitoring is amid the leading 3 cost drivers of clinical trial expenditures (9 to 14% of overall cost) and the process was conducted through the help of ML/AI tool.

2.7 Impact of AI/ML in Multiple Areas of R&D

AI/ML techniques have a significant impact on multiple areas of R&D in the pharmaceutical industry. This section will focus on three areas: drug discovery, drug repurposing, and personalized medicine.

2.7.1 Drug Discovery

Using AI/ML techniques, the search for new medications could be sped up. For instance, deep learning models can be used to predict the affinity of tiny molecules for binding to proteins, saving time and money on experimental validation (Dhakal *et al.*, 2022). Massive volumes of data from many sources, such as genetic data, clinical data, and chemical data, can also be evaluated using AI/ML models in order to uncover potential treatment targets (Vamathevan *et al.*, 2019). Machine learning (ML) methods build models that learn from problem-specific training data using sophisticated pattern recognition and implicit programming. They are frequently used in preclinical drug discovery because of these features (Schuhmacher *et al.*, 2021). According to (Ekins *et al.*, 2019), they have also been utilised to precisely predict bioactivity, ADMET-related outcomes, and physicochemical characteristics.

They are commonly utilised in preclinical drug discovery for these reasons and have proved effective in improving the accuracy of predictions of bioactivity, ADMET-related endpoints, and physicochemical qualities.

2.7.2 Drug Repurposing

AI/ML techniques can also be used for drug repurposing, which is the process of identifying new uses for existing drugs. This can be achieved by analyzing large amounts of data from various sources, including EHRs, scientific literature, and clinical trial data. For instance, Lv *et al.* (2021) used natural language processing (NLP) and machine learning algorithms to identify potential drug candidates for the treatment of COVID-19. This approach can potentially speed up the drug repurposing process and lead to the development of new therapies for various diseases.

2.7.3 Personalized Medicine

Personalising a patient's medical treatment in accordance with their genetic and clinical characteristics is the aim of personalised medicine. Using AI/ML algorithms, large patient data sets can be evaluated to identify biomarkers and predict the effectiveness of treatments. For instance, Benzekry *et al.* (2021) used machine learning models to predict how lung cancer patients would respond clinically to immunotherapy. This approach might result in less expensive medical care and better treatment outcomes.

2.8 Impact of AI/ML in Clinical Trial Operations and Data Analysis:

AI/ML techniques can have a significant impact on clinical trial operations and data analysis. This section will focus on two areas: patient recruitment and clinical trial data analysis.

2.8.1 Patient Recruitment

Patient recruitment is a critical aspect of clinical trials, and delays in recruitment can significantly impact the timeline and cost of clinical trials. AI/ML techniques can be used to identify potential trial participants based on their electronic health records and other data sources. For instance, Zheng *et al.* (2023) used NLP and machine learning algorithms to identify patients with atrial fibrillation who were eligible for a clinical trial. This approach can potentially improve patient recruitment and reduce the time and cost required for clinical trials.

2.8.2 Clinical Trial Data Analysis

AI/ML techniques can also be used to analyse clinical trial data and identify patterns and trends. This can potentially improve the accuracy of clinical trial results and reduce the risk of adverse events. For instance, Lui *et al.* (2022) used machine learning models to predict the risk of adverse events in clinical trials of immunotherapy for cancer.

This approach can potentially improve patient safety and lead to the development of more effective therapies.

2.8.3 Augmenting virtual screening using AI

In comparison to experimental screening approaches like high-throughput screening, VS is a computational methodology that offers a complementary and economical way for hit detection (Ripphausen *et al.*, 2010; Zhu *et al.*, 2013; Damm-Ganamet *et al.*, 2019; Gorgulla *et al.*, 2020). Instead of physically screening every chemical in the screening collection, VS prioritises a subset of compounds for assessment in a primary test using computational methodologies.

A substantial challenge for traditional VS approaches is posed by the growing amount of "make-on-demand" screening libraries and the rising number of high-value, difficult pharmacological targets discovered through functional genomics screening. In order to effectively explore the chemical space for hit identification, AI methods that complement VS methodologies have attracted a lot of attention in the drug discovery process.

2.8.4 Ligand-based virtual screening

Based on the idea of molecular similarity, ligand-based VS (LBVS) approaches seek to find active substances from a chemical library. They consist of form matching, similarity searching, pharmacophore screening, and predictive modelling.

The traditional QSAR modelling paradigm is extended by predictive modelling for VS. Traditional QSAR builds explanatory models that retroactively quantify SAR trends using statistical data-modeling techniques on a congeneric series. The ChEMBL database (Mendez *et al.*, 2019) and PubChem's BioAssay (Wang *et al.*, 2017) as well as improvements in ML and DL algorithms that can handle massive data sets, have opened up new possibilities for QSAR modelling as a VS approach. As a result, many successful QSAR-based VS process applications for hit recognition have been documented. An ML-based

QSAR approach for VS was successfully implemented, according to Zhang *et al.* (2013), and this resulted in the identification of brand-new antimalarial drugs.

The scientists constructed a binary classifier model (active or inactive) utilising 3133 chemicals with known antimalarial activity using two machine learning methods (SVM and kNN). The ChemBridge database was screened using the QSAR models, and 174 compounds were chosen for further testing in cellular and growth-inhibitory experiments for Plasmodium falciparum. 25 of the chosen compounds were shown to be active during experimental validation, producing a success rate of 14.2%; the most potent hit had an EC50 value of 95.6 nM. The use of ML and DL-based QSAR procedures as prospective VS tools has since been described in numerous publications (Anantpadma *et al.*, 2019; Chen *et al.*, 2020; Donlin *et al.*, 2021).

Web-based cheminformatics workbenches that expedite and automate ML- and DL-based QSAR operations for VS have gained popularity over the past ten years. DeepScreening, an intuitive open-source web service created by (Liu *et al.*, 2019), enables users to construct and verify RNN models using either user-provided data sets for VS or ChEMBL bioactivity data. Based on the bioactivity data gathered from ChEMBL 24, DeepScreening additionally offers prebuilt DNN models for 1251 targets. Both QSAR specialists and nonexperts can execute VS against a particular target of interest because to the user-friendly interface and the availability of prebuilt QSAR models. Another open-source web server, DpubChem (Soufan *et al.*, 2018) generates categorical QSAR models from PubChem data using ML techniques.

Although chemogenomic databases like PubChem and ChEMBL offer enough bioactivity information to develop models, there are still substantial drawbacks to using these tools. The presence of bioactivity data from disparate sources and an unbalanced ratio of active to inactive molecules for a specific target are the two main problems. In comparison to other virtual screening techniques, this makes the

generalised usage of QSAR-based procedures for VS more challenging to deploy utilising public data sets (Vijayan *et al.*, 2022).

2.8.5 *In silico* ADMET prediction

A paradigm shift occurred in the pharmaceutical industry as a result of the discovery that poor pharmacokinetics of drug candidates was a significant contributor to clinical attrition in the late 1990s (Kola and Landis, 2004). It saw the creation of numerous developability measures to manage compound attributes during LO (Martin, 2005; Young *et al.*, 2011; Wager *et al.*, 2016) as well as several property-based drug-likeness rules, such as Lipinski's Ro5 (Lipinski *et al.*, 2001). Additionally, the development of miniature, high-throughput in vitro ADMET profiling assays led to the concurrent assessment of effectiveness and ADMET in the early stages of drug discovery. The goal of in silico ADMET modelling is to aid project teams in the design and selection of novel compounds with superior ADMET properties and in the allocation of experimental resources to the most advantageous compounds, thereby minimising the total number of compounds that must be synthesised and profiled (van de Waterbeemd and Gifford, 2003; Lombardo *et al.*, 2017). Pharmaceutical companies have implemented numerous global in silico ADMET models in their discovery pipelines throughout the years.

2.9 Application of Nonparametric Bayesian Learning in Clinical Trial Design

A potent technique in the ML framework for clinical trial design is nonparametric Bayesian learning. The risk of overfitting is decreased and model accuracy is increased with nonparametric Bayesian models because they can automatically assess the complexity of the model depending on the data. This strategy may enhance the planning and analysis of clinical studies. For instance, Spanbauer and Sparapani (2021) utilised a nonparametric Bayesian model to examine clinical trial data and pinpoint patient groupings that had a higher chance of responding favourably to a treatment. By identifying patient

demographics who are most likely to benefit from a medicine, this strategy may be able to enhance patient outcomes while also lowering healthcare expenditures.

Many of the current machine learning techniques, also known as model selection, are centred on learning a set of parameters within a class of models using the proper training data. The possibility for model over- or under-fitting, as well as the identification of an underlying data structure and associated reasons, are significant problems that arise in practise (Webb *et al.*, 2011). Examples could be choosing how many clusters to use in a clustering problem, how many hidden states to use in a hidden Markov model, how many latent variables to use in a latent variable model, or how complex to make the features in a nonlinear regression model. As a result, it is critical to properly train ML techniques to function dependably and provide reliable predictions in the actual world. Cross-validation is a popular and effective technique for assessing how well ML algorithms work when choosing tuning parameters.

Nonparametric Due to its adaptability, Bayesian learning has become a potent tool in contemporary machine learning frameworks, giving a Bayesian framework for model selection utilising a nonparametric method. A Bayesian nonparametric model, in particular, enables us to employ an infinite-dimensional parameter space and only use a small portion of the available parameters on the given sample set. A popular Bayesian nonparametric model among them is the Dirichlet process, particularly in Dirichlet process mixture models (also known as infinite mixture models). Without specifying the number of components in advance, Dirichlet process mixtures offer a nonparametric method for modelling densities and locating latent clusters within the observed variables. Using finite truncations or directly sampling from infinite mixtures, Markov Chain Monte Carlo (MCMC) approaches have advanced.

Such Bayesian nonparametric models have various uses in clinical trial design. For instance, nonparametric Bayesian learning can provide effective and efficient dose selection in oncology dose-finding clinical trials. Patients with different forms of cancer are frequently enrolled in oncology clinical trials, which results in heterogeneity. These problems might be more pronounced in cell and immunotherapies. Designs that fail to take into account the diversity of safety or efficacy characteristics among different tumour types may result in erroneous dose selection and ineffective target population identification. Nonparametric Bayesian learning-based approaches for adaptive dose finding with different populations were proposed by (Li *et al.*, 2020). These Bayesian logistic regression model (BLRM)-based designs enable data-driven information borrowing across several populations while accounting for variability, enhancing the effectiveness of the dose search as well as the precision of calculation of the ideal dose level. By utilising Bayesian nonparametric learning across many indications, Liu (2020) extended another frequently used dose-finding design, modified toxicity probability interval (mTPI) designs, to BNP-mTPI and fBNP-mTPI. The Dirichlet process, which is more adaptable in prior approximation, is used in these designs, and it can automatically cluster patients with similar characteristics depending on what is learned from the emerging data.

Nonparametric Bayesian learning can also be used in master protocols, such as basket, umbrella, and platform trials, which enable the examination of numerous treatments, numerous diseases, or both, within a single trial (Saville and Berry, 2016; Woodcock and LaVange, 2017). These trials offer a greater potential to generate efficacy and safety data faster through adaptive decision-making thanks to the application of nonparametric Bayesian learning. This may result in a shorter medication development timeframe in a medical field with a high unmet medical need. Adaptive platform trials, for instance, have quickly become an important tool in the evaluation of potential COVID-19 therapies. The clinical benefits of remdesivir and

dexamethasone have been shown using such methods in the Adaptive COVID-19 Treatment Trial (ACTT) and the Recovery trial.

Whether it is appropriate to borrow from other treatments or indications is one of the main questions in master protocols. If there is a naive pooling of subgroup information, there may be bias due to the limited sample size, variability, and inflated type I error. To get over the limitations of utilising either independent testing or naive pooling procedures, various Bayesian hierarchical models (BHMs) have been developed, such as the exchangeability-non-exchangeability (EXNEX) model and the Bayesian hierarchical mixture model (BHMM).

The misspecification of parameters might be an issue when there is little prior knowledge of the heterogeneity among various illness subtypes. Bayesian nonparametric learning is developing as an effective tool to enable flexible shrinkage modelling for heterogeneity across particular subgroups and for automatically capturing the additional clustering, overcoming the possible limitations of conventional parametric borrowing approaches. According to Bunn *et al.* (2020), these models allow for more trustworthy data-driven decision-making in basket trials and call for fewer assumptions than other, more widely utilised approaches. These adaptable Bayesian borrowing techniques are expanded upon by Hupf *et al.* (2021) to take into account real-world or historical data.

2.10 Conceptual framework

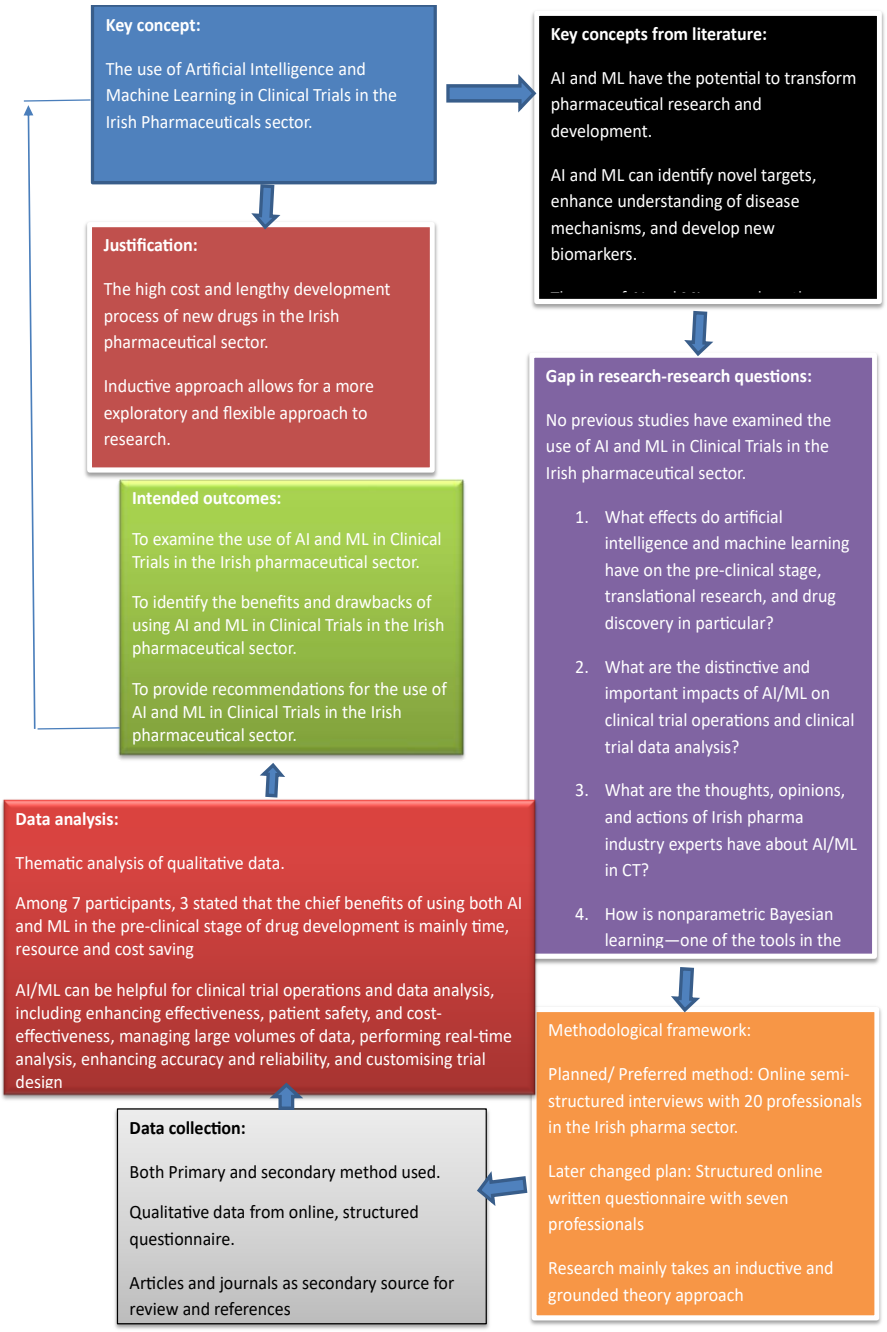

Figure 3 Conceptual framework

The impact of computational techniques like artificial intelligence and machine learning on drug research and development is the focus of this study. Examining the relevance and effects of AI and ML in the area of drug discovery in the Irish pharmaceutical sector is the purpose of this study. It costs around USD 2.6 billion and takes ten to fifteen years to

bring a new prescription medicine to market. The most creative solution to this unfortunate predicament has been hailed as the development of computer-enabled drug design technology, but it depends on careful planning throughout the development process.

AI/ML techniques have shown significant potential in various aspects of drug development, including drug discovery, translational research, and the pre-clinical and clinical phases. Additionally, AI/ML techniques have a significant impact on multiple areas of R&D in the pharmaceutical industry, including drug discovery, drug repurposing, and personalized medicine. Moreover, AI/ML techniques can improve clinical trial operations, patient recruitment, and clinical trial data analysis.

The intended study technique was conducting semi-structured online interviews with 20 industry professionals in Ireland's pharmaceutical industry. The methodology's primary goal was to gather qualitative information to address the study objectives. The approach was modified to data collection using an online, written, structured questionnaire due to a low response rate within the anticipated and available timeframe. The intended conceptual framework was significantly impacted by this alteration. Actually, the number, kind, and amount of time that were available for analysis decreased the validity and reliability of the intended data analysis, and therefore the results and conclusions.

3 Research Methodology

3.1 Overview

This study aims to explore the potential impact of machine learning and artificial intelligence (AI/ML) on clinical trials in the Irish pharmaceutical sector. The primary objective is to determine the potential of utilizing these techniques to increase the success of drug development, including but not limited to drug discovery, translational research, and the pre-clinical phase. The research will also examine and explore attitudes, opinions, and behaviours towards AI/ML in clinical trials among a cross-section of professionals in the Irish pharma sector. Finally, the study aims to understand the application of Nonparametric Bayesian learning (a tool in the ML framework) in CT design. To attain these objectives, researcher aims to conduct qualitative research because it allows for in-depth exploration and understanding of complex phenomena. It also provides flexibility and adaptability in data collection and analysis, which is particularly important when exploring a relatively new and rapidly evolving field such as AI/ML in clinical trials. The research was planned to carry out through interviews, but due to difficulties experienced in setting these up, it was changed into an additional and alternative method of data collection – a written questionnaire. After a period of three weeks attempting to recruit interviewees, none responded positively, leading to the deployment of an alternative method, an online, written, structured questionnaire. This research belongs to the naturalistic paradigm because the goal of the study is to comprehend the experts' subjective and context-dependent attitudes, views, and behaviors about AI/ML on the clinical trials in Ireland. Interpretivist philosophy is used in this study. Inductive is the research approach that the researcher took into account. There was a mix of grounded theory/survey strategies and most of the information has been obtained via structured, online, written questionnaire.

3.2 Research design

According to (Bryman and Bell, 2011), research design is essential for carry out the research since it offers a structure for methodically gathering, examining, and interpreting data. It aids researchers in organizing their research questions, choosing the proper methodology, and selecting the most effective data collection methods in general. The multilayered approach also referred to as the onion model is a framework that illustrates the complexity of research design. In this aspect, this research consider the research onion model which was suggested by Saunders *et al*. (2015).

According to this model, there are five different layers that are pertinent to any research studies include research philosophy, approach, strategy, choice, time horizon, and techniques. The following figure provide the detailed view of onion model that is suitable for this research:

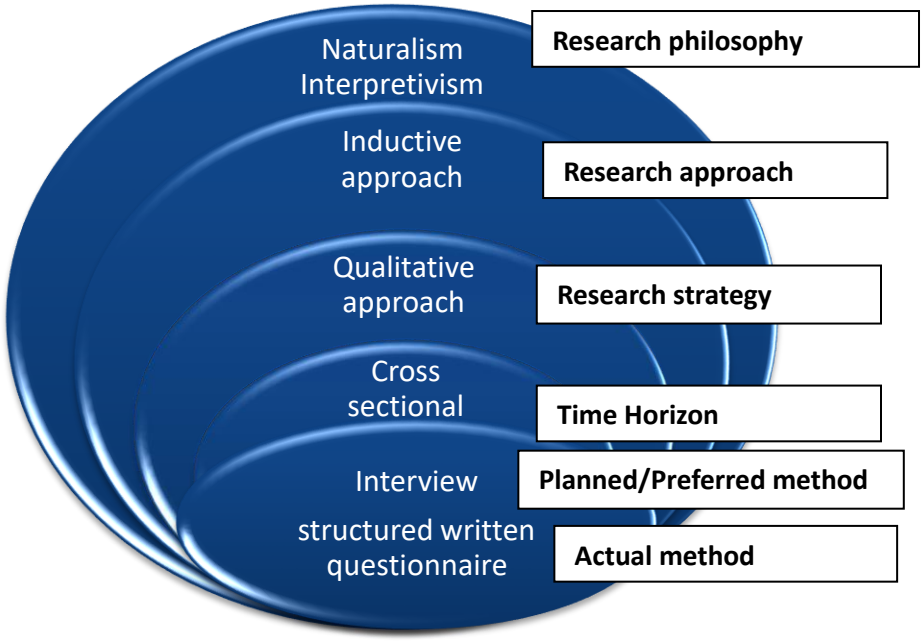

Figure 4 Research Onion model

Adopted from Saunders et al. (2003)

It was originally intended for the research to be conducted through interviews, but due to problems in setting them up, the researcher used an online, written, structured questionnaire to gather data due to a low

response rate within the anticipated and reachable period. The suitability, reliability, and validity of the data obtained are impacted by the change in data gathering method.

3.3 Research Paradigm

A research paradigm is a broad framework that directs how the researcher views the research problem and shapes the research approach. The research paradigm a person chooses has an impact on their worldview, how they approach the research question, and the techniques and methods they employ to gather and analyse data. It also affects how the researcher interprets and extrapolates the study's findings (Kivunja and Kuyini, 2017). There are different types of paradigms, positivist, interpretive, critical, pragmatic and constructivist paradigm (Armstrong, 2010).

The overall paradigm followed in this study is naturalistic. The naturalistic paradigm emphasizes the importance of studying the subjective experiences and meanings that individuals attach to events, rather than solely focusing on objective facts or measurements (Armstrong, 2010). This paradigm acknowledges the complexity and diversity of human experiences and the need to understand them in their natural contexts. In this study, the focus is on understanding the attitudes, opinions, and behaviours of professionals towards AI/ML in clinical trials, which are subjective and context-dependent. By adopting a naturalistic paradigm, this study acknowledges the importance of interpreting the subjective experiences and meanings of the participants and how they relate to the research questions, rather than seeking objective facts or measurements. Therefore, a naturalistic paradigm is an appropriate approach for this study, as it allows for a nuanced understanding of the complex phenomenon being investigated. The change of data collection method has not impacted the paradigm.

3.4 Research Philosophy

The main philosophy followed in this study is interpretivism. Interpretivism is a philosophical approach that emphasizes the importance of understanding human experiences through the interpretation of subjective meanings that individuals attach to events (Chowdhury, 2014). This approach recognizes that the meaning of events and experiences is constructed through social interactions and is context-dependent. In this study, the focus is on understanding the attitudes, opinions, and behaviours of professionals towards the use of AI/ML in clinical trials, which are subjective and context-dependent. Interpretivism is an appropriate approach because it emphasizes the importance of interpreting the subjective experiences and meanings of the participants and how they relate to the research questions. It allows for a more nuanced understanding of the complex phenomenon being investigated and acknowledges that reality is constructed through the interpretations of individuals rather than being an objective, external reality (Neubauer *et al*., 2019).

3.5 Research approach

The association between theory and research can be approached in a different way as described by (Ridder, 2017). This research adopts inductive approach since researcher aimed to show the potential impact of AI and ML on Irish clinical trials. The major reason for choosing the inductive approach perhaps related to the qualitative research while deductive approach to the quantitative choice of methodology (Kaushik and Walsh, 2019). This study's methodology uses an inductive technique, which is frequently used in qualitative research to understand human experiences, perceptions, and behaviours more deeply. The inductive approach enables a more flexible and experimental strategy to accomplish this goal, which is to investigate the possible influence of AI and ML on Irish clinical trials. Due to a low response rate within the anticipated and accessible timeframe, the researcher collected data using an online, written, structured

questionnaire. In order to develop new ideas or concepts pertinent to the study topics, patterns and themes in the acquired data were analysed.

It is a bottom-up approach where the researcher collects and analyzes data, identifies patterns and themes, and then generates new theories or concepts. The justification for using an inductive approach is that it allows for a more exploratory and flexible approach to research. It is particularly useful when the research area is complex and there is limited existing knowledge or theory. By collecting data and analyzing it without preconceived theories or hypotheses, it allows for the emergence of new and unexpected insights. Additionally, the inductive approach is often used in qualitative research where the goal is to gain a deeper understanding of human experiences, perceptions, and behaviors. Due to a low response rate within the anticipated and accessible timeframe, the researcher collected data using an online, written, structured questionnaire. In order to develop new ideas or concepts pertinent to the study topics, patterns and themes in the acquired data were analysed.

3.6 Research strategy

A research strategy can be thought of as a plan for carrying out the research to address the research question (Antwi and Hamza, 2015). It involves several types like experimental, survey, archival, action research, grounded theory, ethnography, and case study (Saunders *et al*., 2012). In this research, researcher consider grounded theory. The major reason for considering this strategy is the researcher aims to carry out qualitative research to understand the potential impact of ML and AI on performance of Irish clinical trials. The grounded theory is used for qualitative research (Sosa-Díaz and Valverde-Berrocoso, 2022). Hence, researcher consider the grounded theory as the suitable one and this theory also helps the researcher to know about the complete data collection and analysis process. The data collected from interview can be analyzed using grounded theory to identify key

concepts and themes that emerge from the data. These concepts and themes can then be used to develop a theoretical framework that explains how AI/ML is impacting clinical trials in the Irish pharmaceutical sector. Using grounded theory in this context could provide a deeper understanding of the challenges and opportunities associated with the use of AI/ML in clinical trials, and could inform the development of policies and guidelines to ensure that these technologies are used in an ethical and effective manner. Grounded theory is also used to examine how these emerging technologies are changing the landscape of clinical trials and how they are affecting the various stakeholders involved, including patients, researcher, regulators, and industry professionals.

The initial goal of the study strategy was to use grounded theory to conduct qualitative research and comprehend the possible effects of ML and AI on the effectiveness of Irish clinical trials. However, the methodology was changed to collecting data via an online, written, structured questionnaire due to a low response rate within the anticipated and accessible timeframe. With this approach, the researcher is still able to gather qualitative information and analyse it to pinpoint important ideas and themes. Open-ended questions can be used in the questionnaire's design to elicit thorough responses from respondents regarding their experiences with AI/ML in clinical trials. To find patterns and themes, the data can subsequently be analysed using qualitative research techniques like thematic analysis. The advantages and disadvantages of using AI/ML in clinical trials may be better understood using this method, which could also help with the creation of rules and regulations that will ensure their ethical and efficient application. To give a more thorough knowledge of the influence of AI/ML on clinical trials in the Irish pharmaceutical industry, future study might take into account employing a mixed-methods approach to gather both quantitative and qualitative data.

3.7 Research Method

The research mainly takes an inductive and grounded theory approach, which requires appropriate methods for collecting qualitative data. These methods include reviewing documents and literature, making observations of practice, interviewing practitioners, and perhaps reviewing audio visual material. The inductive approach allow for the generation of new theories and concepts based on the data collected, while the grounded theory approach allow for the development of theories that are rooted in the data and context of the study (Tie *et al*., 2019). These approaches are well-suited to exploring the experiences and perceptions of individuals and gaining a deep understanding of complex phenomena.

Additionally, using a single research method provides a more straightforward and streamlined research process, therefore researcher adopted this method in this research. Further, the qualitative method enables the researcher to fully comprehend the experiences and viewpoints of the participants. According to Merriam (2009), qualitative research is a technique that involves a variety of participants and non-numerical data. Each participant's research perceptions are given a unique meaning based on their personal experiences. It is beneficial to delve deeply into their thoughts, feelings, and opinions and to enquire about topics that may be challenging to cover in closed-ended survey questions. Additionally, rather than measuring variables, the researcher is more interested in people's experiences and perceptions, so a qualitative approach is more appropriate than a quantitative one. Initially, researcher planned to conduct online, semi-structured interviews with 20 professionals in the Irish pharma sector. The methodology intended to collect mainly qualitative data to answer the research questions as mentioned earlier. Due to a poor response rate within the planned an available time-frame, the method was changed to data collection via an online, written, structured questionnaire.

3.8 Target population, sampling and sample size

In this research, professionals in the Irish pharmaceutical industry with knowledge of clinical trials or AI/ML in drug development was selected. Their relevant professional expertise in the pharmaceutical industry, particularly in clinical trials or they used AI/ML in drug development, was the basis for the selection criteria that is purposive sampling method was used to choose the respondents.

The researcher used online directories of professionals in the pharmaceutical sector, professional networks, and contacts in the industry to find suitable volunteers. The researcher then approached potential volunteers and invited them to take part in the study while detailing the eligibility requirements and study goals. Evaluations of their availability, candidness, and desire to engage were all part of the selection process. The researcher used initial email invites and subsequent phone calls to recruit individuals. However, the researcher was able to enroll 7 people who matched the study's eligibility requirements. By taking into account elements like their degree of expertise, the researcher made sure that the people chose were representative of the general population.

3.9 Data collection process and procedure

Primary and secondary data are the two main categories of data collection process (Creswell, 2009). In this study, researcher used both the methods in necessary time. Under the primary method, researcher used semi-structured interview questions as per the plan of the initial work. However, the researcher used an online, written, structured questionnaire to gather data because there was a low response rate within the expected and reachable timeframe. Potential volunteers who matched the study's inclusion requirements were sent the questionnaire through email. The purpose of the questionnaire was to obtain data on the participants' experiences, viewpoints, and views regarding the subject matter of the study. A sequence of predetermined questions provided in a systematic fashion made up the questionnaire.

Participants had to choose an answer from a list of choices or respond in writing to open-ended questions. The questionnaire was made to be finished in about 15-20 minutes, which is an acceptable amount of time. A structured questionnaire has the advantage of allowing for consistency in data collection and being easily quantifiable. The answers to the questions were noted and examined for themes and patterns pertinent to the research topics. This methodology made it possible for fresh notions or ideas related to the study to emerge. Thematic analysis was used to analyse the data that had been collected. This keeps the interview's structure and emphasis while allowing for a deeper study of the opinions and experiences of the participants. The questions were constructed in such a way that they can be expanded upon dependent on the respondents' degree of interest and the dialogue that is had throughout the interview. After data collection was complete, the researcher downloaded the responses from the online survey platform and imported the data into a spreadsheet program for analysis. The data were coded and analysed for patterns and themes relevant to the research questions.

3.10 Data analysis

After the data collection was completed through an online, written, structured questionnaire, the data was organized and analyzed using a thematic approach. The textual data was reviewed and categorized into themes and sub-themes based on the research questions and the relevant literature. The data was then coded and analyzed multiple times to ensure a coherent and logical order of the themes and sub-themes (Vaismoradi *et al.*, 2016). This approach was adopted to gain a comprehensive understanding of the potential impact of AI/ML on Irish clinical trials based on the data collected through the questionnaire. In addition, thematic analysis was used to analyze the data and identify any patterns or trends that emerged. This approach was chosen as it enabled a more detailed and nuanced understanding of the data collected and allowed for a more in-depth exploration of the research questions.

3.11 Ethics and access

The University and faculty members provide formal ethical permission before the researcher can start the study. At every level of the study process, the researcher adheres to the ideals of respect for human dignity, beneficence, and humanity. Instead, then using their names or asking direct inquiries, pseudonyms and code words were employed. The secrecy would be protected by this procedure (Creswell, 2007). The transcripts were altered to remove all identifying information and replace it with a pseudonym.

3.12 Summary

A thorough summary of the study's methodology is provided in this chapter. The Onion Research Model and qualitative research techniques were used in the research design. Semi-structured interviews were the main method of data collection used in the study, and they were conducted in person. However, an online, written, structured questionnaire was also employed to gather data because there was a poor response rate within the anticipated time range. Data for the study were gathered from primary and secondary sources. The data were examined repeatedly, and codes were allocated based on the topics of the study, in a process known as thematic analysis. Single major statements that provided a variety of themes served as the basis for coding units.

4 Data analysis and findings

4.1 Introduction

The analysis and conclusions of the qualitative data that were gathered for the research study are the main topics of this chapter. The observations from the chapter's literature review will be discussed in more detail later on. The first two research questions are based on a review of the literature; thus, the researcher will conduct a content analysis before moving on to the next two topics through a thematic analysis. The researcher's original intention was to do qualitative analysis through conducting interviews. The amount and type of data obtained were reduced as a result of a late modification in the data gathering procedure, which was mentioned in the preceding chapter. In turn, this had an impact on the amount of time that could be used for analysis and decreased the validity and dependability of the planned data analysis. A more simplistic kind of analysis was feasible due to the limitations of the data gathering technique, which might have limited the breadth of the results, conclusions, and consequences. This chapter makes use of both sources they are verbal non-numerical data and interview that were gathered from various individuals by the researcher. It meant a more a basic form of analysis was possible. Two zoom interviews and seven questionnaire responses were received.

In this first section, the results of the interview that were done will be discussed. Each component of the findings will be followed by an analysis and discussion of how they relate to the literature. The purpose is to investigate the respondents' opinions and experiences about the

use of attitudes, opinions, and behaviours towards AI/ML in clinical trials among experts in the Irish pharmaceutical industry. After a quick introduction of the respondents, the conversation will focus on their personal information and experiences in general. The primary effects of AI/ML specifically and critically on the administration of clinical trials and the analysis of their data will next be covered. A discussion of the respondent's views on the use of nonparametric Bayesian learning (a tool in the ML framework) in clinical trial design concludes the chapter.

4.1.1 Research objective

To examine and explore attitudes, opinions, and behaviours towards AI/ML in CT among a cross-section of professionals in the Irish pharma sector.

4.1.2 Research questions

1. What effects do artificial intelligence and machine learning have on the pre-clinical stage, translational research, and drug discovery in particular? (This links to question 1,2,3,4 and 7)
2. what are the distinctive and important impacts of AI/ML on clinical trial operations and clinical trial data analysis? (This links to question 5 and 6)
3. What are the thoughts, opinions, and actions of Irish pharma industry experts have about AI/ML in CT? (This links to question 8 and 9)
4. How is nonparametric Bayesian learning—one of the tools in the ML framework—applied in CT design? (This links to question 10,11,12,13 and14)

4.1.3 Related questions on questionnaire

- What do you see as the main benefits of using AI and ML in the pre-clinical stage of drug development?

- How can AI and ML improve the efficiency of translational research?
- In what ways can AI and ML help accelerate drug discovery?
- Are there any potential drawbacks to using AI and ML in the pre-clinical, translational and drug delivery stages?
- In your experience, how has the use of AI and ML impacted the efficiency of clinical trial operations?
- What are the key advantages of using AI and ML for clinical trial data analysis?
- How familiar are you with the use of AI and ML in clinical trials?
- What are some of the challenges associated with implementing AI and ML in clinical trials?
- What steps do you think need to be taken to ensure that the use of AI and ML in clinical trials is ethical and responsible?
- Can you describe how nonparametric Bayesian learning works in the context of clinical trial design?
- What are some of the benefits of using this nonparametric Bayesian learning approach in clinical trial design?
- Are there any limitations or drawbacks to using nonparametric Bayesian learning in CT design?
- How does the use of nonparametric Bayesian learning affect the role of statisticians in clinical trial design?
- In your opinion, what are some other applications of nonparametric Bayesian learning in the pharmaceutical industry beyond clinical trial design?

4.2 Respondents characteristics

To comprehend the respondent's backgrounds, a quick presentation of their qualities will be made. The majority of this data was gathered through interviews; however, some data was also gathered from research on the people who were questioned. Interviews were

conducted with a total of seven people, four of whom were men and three of whom were women. Due to the respondents' absence from the interview, the researcher had planned to speak with 20 persons. Only seven respondents could be gathered by the researcher. A copy of the questionnaire included in the Appendix A.

Respondents code	**Age**	**Experience**	**Current Position**
A	34+	4 years	Research Scientist
B	36	9 years	Clinical Trial Coordinator
C	40+	7 years	Pharmaceutical Executive
D	Not specified	10+ years	Data Scientist
E	35	5 years	Research Scientist
F	45	12 years	Biotech Entrepreneur
G	42	10 years	Clinical Research Associate

Table 1 Respondents characteristics

It has been noticed that the respondents have a variety of backgrounds and levels of expertise in the pharmaceutical industry, per the respondent information received. Most of them have more than five years of experience, and some have more than ten. There is a wide age range, with some people being in their mid-thirties to mid-forties. The Respondents offer a solid representation of people employed in various positions within the pharmaceutical sector. Research Scientist, Clinical Trial Coordinator, Pharmaceutical Executive, Data Scientist, Biotech Entrepreneur, and Clinical Research Associate are just a few of the roles represented by the participants in this study.

The respondents' varied origins and degrees of expertise offer a wide range of viewpoints on the subject under study, especially in the context of the pharmaceutical sector. These various positions and experiences offer distinctive viewpoints on the advantages, disadvantages, and difficulties of applying AI and ML at various stages of drug development and clinical trials. The participants' varied degrees of expertise show that they have a thorough awareness of the sector and its inner workings, which can lend depth to the study's results and conclusions. A greater grasp of the implications and prospects of AI and ML in the pharmaceutical sector can result from the different Respondents and viewpoints of the participants, who collectively offer a well-rounded view of the topic under consideration. The themes listed below were created using the study objectives, and each topic provides a comprehensive response to the research questions.

4.3 Qualitative -Thematic analysis

4.3.1 Main benefits of using AI and ML in the pre-clinical stage of drug development

This theme was developed to obtain the data about the main benefits of using AI and ML in the pre-clinical stage of drug development. This theme will give answer to the research question 1 of "What effects do AI and ML have on the pre-clinical stage, translational research, and drug discovery in particular". Preclinical research and planning are essential for creating effective clinical trials because they help identify potential compounds and targets and develop a strategy for obtaining regulatory permission. The researcher planned to obtain data about the situation in the Irish pharma sector specifically.

Out of seven respondents, three respondents stated that the chief benefits of using both AI and ML in the pre-clinical stage of drug development is mainly time, resource and cost saving. For example, Respondent A suggests that

"AI and ML can help to identify promising drug candidates more quickly and accurately, which can save time and resources by reducing the number of candidates that need to be tested in vivo"

In line with Respondent A statement, Respondent C suggests that AI and ML can help to make better decisions about which drug candidates to pursue and how to optimize the drug discovery process, which can reduce the cost and time required to bring new treatments to market and benefit the broader healthcare system. Respondent F also suggests that AI and ML can help to identify new drug targets and develop new therapies more quickly and efficiently, by reducing the number of candidates that need to be tested in vitro and in vivo, ultimately saving time and resources.

Other four respondents like Respondent B, D, E and F stated that the major benefit of using AI and ML in pre-clinical research is improving the patient outcome. For instance,

Respondent B argues that:

"Using AI and ML to optimize pre-clinical research can lead to better predictions of safety and efficacy, which can ultimately improve patient outcomes by designing trials that are more likely to succeed…"

Likewise, respondent D suggests that AI and ML can help to better understand the relationships between different biological pathways and identify potential drug targets that might be overlooked using traditional methods. This can lead to the design of more effective and targeted drugs, ultimately improving patient outcomes. Respondent E highlights the potential for AI and ML to better understand the underlying biological mechanisms of diseases, which can inform the development of new treatments and therapies. By analyzing complex data and identifying new targets, individuals can design drugs that are more effective and targeted, improving patient outcomes.

All these respondents suggest that using AI and ML in the pre-clinical stage of drug development can have a range of potential benefits, including reducing the time and resources needed to identify promising drug candidates, designing trials that are more likely to succeed, and developing more effective and targeted drugs. By using predictive models to analyze large amounts of complex data, researcher can gain insights that might not be possible using traditional methods, ultimately leading to improved patient outcomes and a more efficient drug discovery process. In line with our finding, Zheng *et al.* (2023) used NLP and ML algorithms to identify patients with atrial fibrillation who were eligible for a clinical trial. This approach can potentially improve patient recruitment and reduce the time and cost required for clinical trials.

4.3.2 Role of AI and ML in increasing the efficiency of translational research

This theme the role of AI and ML in increasing efficiency of translational research was developed to answer the research question 1

of “What effects do AI and ML have on the pre-clinical stage, translational research, and drug discovery in particular”.

Out of seven respondents, two respondents like Respondent A and Respondent D both highlight the importance of automating data analysis using AI and ML to improve the efficiency of translational research. Though, these two respondents shared different opinions and their view about the role of AI and ML in increasing efficiency of translational research is similar. According to the Respondent received, Respondent A focuses on identifying patterns to save time and resources, while Respondent D emphasizes the ability to derive insights from large datasets. However, other two respondents like Respondent B and G highlights the potential for AI and ML to improve patient outcomes and optimize clinical trial designs. Respondent B specifically mentions using real-time data analysis to adjust protocols, while Respondent G emphasizes the importance of accurate patient selection and stratification. These two respondents’ statement are important as they suggest the specific ways in which AI and ML can improve patient outcomes and optimize clinical trials.

Respondent C and Respondent F both highlight the potential for AI and ML to transform drug development processes and bring new treatments and therapies to market faster. According to Respondent C statement, the ability to identify new targets and develop new therapies faster and with greater precision, while Respondent F emphasizes the ability to reduce costs and time required for drug development. Even though, Respondent F does not directly mention the feasibility of AI and ML in translational research but stated its importance in terms of cost and time reduction. Finally, only one respondent, Respondent E emphasizes the

“... importance of using AI and ML to identify patterns in complex datasets to better understand disease mechanisms and develop new treatments and therapies….”

This perspective is unique in its emphasis on understanding disease mechanisms, which is a critical step in drug development. However, it

could benefit from further elaboration on the specific ways in which AI and ML can help identify patterns and how this can lead to the development of new treatments and therapies.

Each respondents provides valuable insights into the potential benefits of using AI and ML in translational research, they could all benefit from further elaboration and specific examples to support their claims. Empirical literature supports the notion that AI and ML have significant potential benefits in translational research. For example, a study by Rafique *et al.* (2021) found that AI and ML algorithms could improve the accuracy and efficiency of identifying potential drug targets for cancer therapy. Another study by Isaksson *et al.* (2020) demonstrated that AI and ML algorithms could be used to predict the efficacy and toxicity of different cancer treatments. Similarly, a study by Badwan *et al.* (2023) showed that AI and ML algorithms could be used to analyze large-scale genomic and transcriptomic data to identify potential biomarkers for cancer diagnosis and treatment.

4.3.3 Role of AI and ML in accelerating the drug discovery

In terms of understanding the AI and ML role on accelerating the drug discovery process, seven respondents were asked to answer the question in the context and this question is appropriate to answer the RQ1.

For this question, all the respondents shared their different views. Overall, the respondents highlight the potential benefits of using AI and ML in drug discovery, including identifying drug targets (Respondent A and B), predicting drug efficacy and reducing the number of candidates to be tested (Respondent C and F), better understanding the biology of diseases (Respondent D and E), and only one respondent stated that the potential benefits of using AI and ML in drug discovery is optimizing clinical trial design. However, it's important to note that these predictions are only as good as the underlying data, and the drug discovery process is complex and unpredictable. Despite the potential benefits, caution and skepticism are needed when interpreting AI and

ML predictions. According to Respondent A and B, the common statement is;

"If the underlying data is flawed or incomplete, then the predictions will be less accurate. Additionally, AI and ML models can be complex and difficult to interpret, which can make it challenging to determine why a particular drug candidate was selected or rejected…"

Respondent C and F also stated that there is always the risk in applying AI and ML that a promising drug candidate will fail in clinical trials. Even though, AI and ML can help them to make more informed decisions about which candidates to pursue, it is important to maintain a certain degree of caution and skepticism when interpreting their predictions. Similar to these perspective, Respondents D and E both emphasize the role of AI and ML in better understanding the biology of diseases and identifying new drug targets. While this is certainly an important area of research, it's worth noting that the translation of basic research findings into new drugs is often fraught with challenges. Even if a promising drug target is identified, it can be difficult to design a drug that targets that target in a safe and effective way. Additionally, there is often a significant gap between basic research findings and clinical applications, and many promising findings fail to translate into effective treatments.

Respondent G focuses on the potential of AI and ML to optimize clinical trial design. While this is certainly an important area of research, it is value observing that the design of clinical trials is highly regulated and complex, and there are often many factors beyond drug efficacy that can impact the outcome of a trial. Additionally, there is always the risk that a promising drug candidate will fail in clinical trials, regardless of how well the trial is designed.

There is empirical support for the findings that caution and skepticism are needed when interpreting AI and ML predictions in drug discovery. A study by Paul *et al.* (2021) found that some of the AI-based predictions on drug efficacy had limited accuracy due to limitations in

data quality and quantity, as well as the complexity of the drug discovery process. Another study by (Shah et al., 2019), similarly highlighted the limitations of AI and ML in predicting drug efficacy, noting that the predictions are highly dependent on the quality and quantity of the data used for training the models. Therefore, while AI and ML hold promise in drug discovery, careful consideration and evaluation of the underlying data and methods are crucial for ensuring reliable predictions.

4.3.4 Drawbacks to using AI and ML in the pre-clinical, translational and drug delivery stages

The above-mentioned theme was developed to identify the drawbacks of using AI and ML in the three different states of clinical trials include pre-clinical, translational and drug delivery stages. To attain this, the question was asked to respondents are "*Are there any potential drawbacks to using AI and ML in the pre-clinical, translational and drug delivery stages*?".

All of the respondents listed several and different potential drawbacks to using AI and ML in the pre-clinical, translational, and drug delivery stages including the risk of biased or inaccurate results, the need for balance with other methods, the importance of clear regulatory guidelines, the need for high-quality data, concerns around replacing human intuition and creativity, the lack of transparency in AI/ML decision-making, and ethical considerations around privacy and equity. These issues require careful attention to ensure that AI/ML is used effectively and ethically in the drug development process. The respondent's view is shared here as an example,

Respondents	**Opinions**
Respondent A	The risk of biased or inaccurate results if AI/ML algorithms are not properly calibrated or trained on biased data, which can lead to incorrect predictions and potentially harm

	patients. This is an important issue that requires careful attention to ensure that the algorithms are trained on diverse and representative data.
Respondent B	A valid concern about over-reliance on AI/ML and the potential for neglecting other important aspects of drug development, such as patient safety and regulatory compliance. It is important to strike a balance between the use of AI/ML and other methods to ensure that all necessary factors are considered in the drug development process.
Respondent C	Underscores the importance of clear regulatory guidelines for the use of AI/ML in drug development to avoid uncertainty and variability in how AI/ML is used across different drug development programs. This will be especially important as the use of AI/ML becomes more widespread in drug development.
Respondent D	Need for large amounts of high-quality data to effectively use AI/ML in the pre-clinical and translational stages, which may not be available for certain diseases or patient populations. This may limit the effectiveness of AI/ML models in these cases.
Respondent E	Raises the concern that AI/ML could replace human intuition and creativity in the drug discovery process. While AI/ML can analyze large amounts of data and identify patterns, it may not be able to replicate the creativity and

	problem-solving skills of human researchers. It is important to strike a balance between the use of AI/ML and human expertise in the drug discovery process.
Respondent F	The lack of transparency in how AI/ML algorithms make decisions, which can make it difficult to validate and interpret results and potentially slow down the drug discovery process. It is important to develop methods for interpreting and explaining AI/ML results to ensure that they can be effectively utilized in drug development.
Respondent G	Raises ethical concerns around the use of AI/ML in drug development, particularly around issues of privacy, data ownership, and access to healthcare. These issues will need to be carefully considered to ensure that the use of AI/ML in drug development is ethical and equitable.

Table 2 Drawbacks to using AI and ML in the pre-clinical, translational and drug delivery stages

The respondents concluded that AI/ML has the potential to significantly improve drug development, it is important to carefully consider and address these potential drawbacks to ensure that its use is safe, effective, and ethical.

4.3.5 Use of AI and ML impacted the efficiency of overall clinical trial operations

It is observed from the literature, there is an increasing focus of AI and ML in clinical trials, as a result the efficiency of the clinical process is improved (Weissler *et al.*, 2021). Considering this fact, the research

question 2 was framed in chapter 1 as “what are the distinctive and important impacts of AI/ML on clinical trial operations and clinical trial data analysis?”. Hence, the question was asked to the respondents regarding this RQ2 is “In your experience, how has the use of AI and ML impacted the efficiency of clinical trial operations”. For this, two of the respondents like respondent A and F agree that AI/ML can improve clinical trial efficiency by optimizing trial design, patient stratification, and automating tasks such as patient recruitment and data analysis. According to Respondent A focuses more on the benefits of reducing failed trials and bringing new treatments to market more quickly. Respondent F highlights the potential for AI/ML to free up researchers' time for more complex tasks.

Another two respondents like B and E highlight the potential for AI/ML to improve patient safety in clinical trials by providing regulators with more comprehensive and reliable data and identifying potential safety issues earlier in the trial process, respectively.

But other respondents shared different views as, Respondent C emphasizes the cost savings potential which is not mentioned in the other respondents and Respondent D emphasizes the ability to analyze large volumes of data in real-time enabling researchers to make data-driven decisions more quickly. This point is not explicitly mentioned in the others. Respondent G stresses the importance of ethical considerations in the use of AI/ML in clinical trial operations. Overall, the respondents highlight the potential benefits of using AI/ML in clinical trial operations, such as improving efficiency, patient safety, and cost-effectiveness. However, there are also ethical and practical considerations that must be taken into account to ensure that AI/ML is implemented responsibly and effectively.

4.3.6 Advantages of using AI and ML for clinical trial data analysis

Given evidence to the RQ2, the questions were asked to discuss about the advantages of using AI and ML for clinical trial data analysis. All the respondents reveal the various advantages of using AI/ML for clinical trial data analysis, including its ability to handle large volumes of data, perform real-time analysis, improve accuracy and reliability, reduce costs and time, and personalize clinical trial design. However, there is a need to balance these advantages with ethical considerations around patient privacy and data security.

Out of seven respondents, three respondents reveals both AI and ML helps the clinical experts to handle large volumes of data and perform real-time analysis with greater accuracy and reliability. For example, Respondent A highlights the advantage of AI/ML in handling large volumes of complex data. This is echoed in Respondent D, which emphasizes the ability of AI/ML to perform real-time analysis and help researchers make informed decisions. Another similar respondent is Respondent F who also shared that AI/ML can reduce the risk of human error and improve the accuracy of trial results.

However, respondent B notes that the use of AI/ML can help ensure that trial data meets regulatory requirements, which can speed up the drug approval process. But respondent C also reveals that AI/ML can reduce the time and costs associated with data analysis, enabling pharmaceutical companies to bring new treatments to market more quickly and at a lower cost. Notably, respondent E focuses on the ability of AI/ML to identify patient subpopulations that may respond differently to a treatment, enabling more personalized approaches to clinical trial design and ultimately improving patient outcomes. Respondent G also highlight the importance of protecting patient privacy and ensuring data security when using AI/ML for clinical trial data analysis.

4.3.7 Challenges associated with implementing AI and ML in clinical trials

The respondents were asked to know about their perception regarding challenges, hence the below question was framed as;

"What are some of the challenges associated with implementing AI and ML in clinical trials?"

All the respondents shared different challenges and there is no similarity in responses received. For example, respondent A focuses on the challenge of data quality and accuracy, which is a critical issue that can affect the validity and reliability of the machine learning models. Further, respondent B expands on the regulatory challenges associated with AI/ML in clinical trials, which can be a barrier to adoption and may require additional resources for validation and documentation. Respondent C also highlights the challenge of investment in technology and infrastructure, which may be more difficult for smaller pharmaceutical companies with limited resources. Respondent D also touches on the challenge of data volume, but focuses more specifically on the need for large volumes of data to train the machine learning models. Respondent E addresses the challenge of interpreting machine learning results, which can be difficult due to the complexity of the models and the lack of transparency in how they arrive at their conclusions. Respondent F focuses on the challenge of integrating AI/ML into existing clinical trial workflows and infrastructure, which requires careful planning and execution.

Finally, Respondent G raises ethical concerns around data privacy, patient consent, and potential biases in the data used to train the models. This is an important consideration in any AI/ML implementation, particularly in clinical trials where patient safety and well-being are at stake.

Overall, these respondents cover a range of challenges associated with implementing AI and ML in clinical trials, from technical and

regulatory issues to ethical considerations. It is important for researchers and stakeholders to carefully consider these challenges and work to address them in order to realize the full potential of AI and ML in improving clinical trial operations and patient outcomes.

4.3.8 Ethics of AI and ML in clinical trials

Significantly, the important question among the others was asked to respondents are “What steps do you think need to be taken to ensure that the use of AI and ML in clinical trials is ethical and responsible?” As per the data obtained, it is noted that the respondents provide valuable insights into the steps necessary to ensure ethical and responsible use of AI/ML in clinical trials. They emphasize the importance of monitoring and evaluating the performance of models (Respondent A), regulatory guidelines (Respondent B), transparency and accountability from pharmaceutical companies (Respondent C), representative data (Respondent D), ethical adaptation of clinical trial designs (Respondent E), patient involvement (Respondent F), and informed consent (Respondent G). From the above responses, it is understood that there are several important considerations when it comes to the ethical and responsible use of AI/ML in clinical trials. These considerations include ensuring the accuracy and reliability of models, following clear regulatory guidelines, maintaining transparency and accountability, avoiding biases in the data, involving patients in the development and implementation of technologies, and obtaining informed consent from patients. Overall, these insights highlight the need for careful planning and ethical considerations throughout the entire process of using AI/ML in clinical trials.

4.3.9 Nonparametric Bayesian Learning in Clinical Trial Design

This theme was created based on the final RQ, “How is nonparametric Bayesian learning—one of the tools in the ML framework—applied in CT design?”. All the respondents are generally agreeing that nonparametric Bayesian learning is a powerful tool for optimizing clinical trial design. However, some respondents are more specific and

detailed than others. Among the seven respondents, two of the respondents A and B are very similar in emphasizing the potential of nonparametric Bayesian learning to increase the chances of success in clinical trials. Likewise, two respondents like C and D are also similar, both highlighting the efficiency and adaptability of nonparametric Bayesian learning in trial design. Respondent E and F are similar in that they are both from the perspective of a research scientist and emphasize the responsiveness of nonparametric Bayesian learning to the data. They also stress the importance of maintaining scientific rigor. Even though, each two of the six respondents shared the similar views. Contradictorily, respondent G shared different view than others and the person is somewhat more cautious, acknowledging the potential benefits of nonparametric Bayesian learning but also emphasizing the importance of ensuring that the trial remains scientifically rigorous. Overall, the respondents are generally positive about nonparametric Bayesian learning in the context of clinical trial design, while also recognizing the need for scientific rigor and caution.

4.3.10 Benefits of using Nonparametric Bayesian Learning in Clinical Trial Design

This theme was also created based on the final RQ, “What are some of the benefits of using this nonparametric Bayesian learning approach in clinical trial design?”. Four respondents A, B, E, and G all highlight the benefits of nonparametric Bayesian learning in clinical trial design, including increased flexibility and adaptability, better use of data, improved efficiency, and more accurate predictions. These respondents also demonstrate the value of the approach from different perspectives, including healthcare research, AI/ML, and general scientific expertise.

But respondents C and D specifically mention the adaptability of the model as a benefit, with respondent C emphasizing the potential for optimization and success in the trial. Respondent F also mentions improved predictions and adaptation to the data, but does not elaborate on the specific benefits of these capabilities. Overall, the respondents

provide a variety of perspectives on the benefits of nonparametric Bayesian learning in clinical trial design, with common themes being flexibility, adaptability, and improved use of data. There are few studies like (Saville and Berry, 2016; Woodcock and LaVange, 2017) has discussed the benefits of nonparametric Bayesian Learning in clinical trial design.

4.3.11 Challenges in Nonparametric Bayesian Learning in Clinical Trial Design

This theme discusses the challenges in nonparametric Bayesian Learning in Clinical Trial Design; hence respondents were asked to share their views about the same and the details found in below table.

Respondents code	Respondents' response
A	Computational complexity, interpretability, data requirements, and model validation
B	Less interpretable than traditional statistical methods, making it difficult to explain results to stakeholders
C	Computational complexity, interpretability, data requirements, and model validation
D	Computational complexity, interpretability, data requirements, and model validation
E	Less well-suited for all types of clinical trials
F	Selection of appropriate prior distribution can impact accuracy of results
G	Requires a large amount of data to be effective, which may not be available in all cases

Table 3 Challenges in Nonparametric Bayesian Learning in Clinical Trial Design

Three of the respondents in the research highlight several potential limitations and drawbacks of nonparametric Bayesian learning in clinical trial design, such as computational complexity, interpretability, data requirements, and model validation (Respondent A, C and D). Two of the respondents B and E shared that the potential drawback of nonparametric Bayesian learning in that it may be less interpretable than traditional statistical methods, making it difficult to explain results to stakeholders. According to Respondent E, the nonparametric Bayesian learning may not be well-suited for all types of clinical trials, which is another potential limitation. However, Respondent F discusses the importance of selecting an appropriate prior distribution, which is a limitation that can impact the accuracy of results. Lastly, Respondent G mentions a potential limitation in that nonparametric Bayesian learning may require a large amount of data to be effective, which may not be available in all cases.

4.3.12 Use of nonparametric Bayesian learning affect the role of statisticians in clinical trial design

The responses suggest that nonparametric Bayesian learning can have significant benefits for clinical trial design. It is particularly useful in identifying patterns and relationships in the data that may not be apparent with traditional statistical methods. This approach also allows statisticians to optimize study designs, adjust them as needed, and account for uncertainty in the data. However, some of the responses highlight potential limitations and drawbacks to this approach. Respondents A, C, D, and E all suggest that nonparametric Bayesian learning can be computationally complex, less interpretable, and may require a large amount of data to be effective. Respondent F mentions the importance of selecting an appropriate prior distribution, which can impact the accuracy of results. It is important to recognize these limitations and ensure that the use of nonparametric Bayesian learning is carefully considered and justified for each clinical trial.

Moreover, it is worth considering the role of statisticians in clinical trial design beyond the use of nonparametric Bayesian learning. The design of clinical trials involves numerous complex decisions, including determining the sample size, defining endpoints, and selecting appropriate statistical methods. While nonparametric Bayesian learning can provide valuable insights, it is only one tool in the statistician's toolkit. Inclusively, the use of nonparametric Bayesian learning in clinical trial design has great potential but must be used carefully and in conjunction with other statistical methods. Statisticians must carefully consider the specific circumstances of each trial and weigh the potential benefits and limitations of this approach before implementing it. Similarly, an important issue encountered in practice is the potential model over-fitting or under-fitting, as well as the discovery of an underlying data structure and related causes in nonparametric Bayesian learning (Webb *et al.*, 2011).

4.3.13 Applications of nonparametric Bayesian learning in the pharmaceutical industry beyond clinical trial design

The potential applications of nonparametric Bayesian learning in the pharmaceutical industry mentioned by the respondents are promising, but it is important to note that these applications are still in their early stages and may face challenges in implementation. For instance, while Respondent A and C mention the use of nonparametric Bayesian learning in drug safety and toxicity, the development of predictive models may require large amounts of data and computationally intensive methods, which may not always be feasible in practice. Additionally, the interpretation of the models may be difficult, which may limit their utility in decision-making.

Similarly, the use of nonparametric Bayesian learning for personalized medicine, drug discovery, and pharmacokinetics, as mentioned by Respondent B, D, E, and G, may also require large amounts of data and complex models. It is also important to ensure that these models are

robust and accurate, and that they do not perpetuate biases or discrimination in healthcare.

Finally, while Respondent F mentions the use of nonparametric Bayesian learning for patient recruitment and retention, it is important to recognize that these methods may not always address underlying systemic issues that may contribute to low recruitment and retention rates in clinical trials. On the whole, while the potential applications of nonparametric Bayesian learning in the pharmaceutical industry are promising, their implementation may face challenges related to data requirements, computational complexity, model interpretation, and bias. Therefore, it is important to carefully evaluate these methods and their potential benefits and limitations before adopting them in practice. In line with our finding, there are few studies in earlier showed the nonparametric Bayesian learning for patient recruitment and retention like ML methods have established their supremacy in disease prediction (Ma *et al.*, 2018), disease classification (Fakoor *et al.*, 2013), imaging diagnosis (Liu *et al.*, 2014), drug manufacturing (Pantuck *et al.*, 2018), medication assignment (Liang *et al.*, 2014), and genomic feature identification tasks (Zhou and Troyanskaya, 2015).

4.4 Summary of the findings

Insights into the advantages and drawbacks of applying AI and ML to drug development, translational research, clinical trial operations, and clinical trial design were shared by the respondents. They emphasised the potential advantages of applying AI and ML in enhancing drug discovery, clinical trial design, and clinical trial operations, such as saving time and money, enhancing efficacy and safety, providing individualized care, and optimizing dosing. They did, however, also highlight several possible negatives, such as the necessity for high-quality data, ethical and legal considerations, the possibility of bias and mistakes, and the danger of human intuition and creativity being replaced. The responders also talked about nonparametric Bayesian learning's possible advantages and disadvantages in clinical trial

design, highlighting the importance of careful thought and scientific rigour. In general, the prospective uses of AI, ML, and nonparametric Bayesian learning in the pharmaceutical sector are exciting, careful analysis of the unique circumstances, potential advantages, and potential drawbacks is needed before their adoption.

5 Conclusion

5.1 Conclusion

RQ1: What effects do artificial intelligence and machine learning have on the pre-clinical stage, translational research, and drug discovery in particular?

The study's findings point to a number of possible advantages for AI/ML in the drug discovery process, including saving time and money, creating more successful clinical trials, and creating more effective and tailored medications. However, because to the complexity and unpredictability of the drug discovery process, care must be taken when interpreting AI/ML predictions. Other potential limitations include the requirement for a balance of different approaches, the need for clear regulatory guidelines, the need for high-quality data, and ethical problems.

RQ2: What are the distinctive and important impacts of AI/ML on clinical trial operations and clinical trial data analysis?

The results imply that AI/ML can be helpful for clinical trial operations and data analysis, including enhancing effectiveness, patient safety, and cost-effectiveness, managing large volumes of data, performing real-time analysis, enhancing accuracy and reliability, and customising trial design. However, ethical issues of patient confidentiality and data security must also be taken into account.

RQ3: What are the attitudes, opinions, and behaviours of Irish pharma industry experts have about AI/ML in CT?

According to the study's findings, Irish pharma sector professionals are largely supportive of the usage of AI/ML in CT. Although they are aware of the potential advantages, such as increased effectiveness, accuracy, and patient outcomes, they also understand the need for prudence and cautious preparation owing to moral and practical reasons.

RQ4: How is nonparametric Bayesian learning—one of the tools in the ML framework—applied in CT design?

According to the study's findings, nonparametric Bayesian learning has a lot of potential for CT design, including flexibility, adaptability, and better data usage. Before using this strategy, statisticians must carefully assess the unique circumstances of each trial and weigh the advantages and disadvantages of it. Furthermore, difficulties with data needs, computational complexity, model interpretation, and bias must be taken into account.

5.1.1 Experiences, behaviours, attitudes and opinions on AI and ML

"Two key questions for this research are closely linked. Following a review, the reported impact of AI/ML specifically and critically in clinical trial operations and clinical trial data analysis reported in the literature (research question 2) and to examine and explore attitudes, opinions and behaviours towards AI/ML in CT among a cross-section of professionals in the Irish pharma sector (research question 3). A review of recent literature shows many case-studies, vignettes and examples that purport evidence of the first – there are many clinical trials that have occurred over the past five years in which AI and ML have been successfully used. Moreover, based on the answers given by the small number of respondents to the questionnaire used in this study, Irish pharma professionals claimed and demonstrated that they had high levels of awareness of the potential benefits and advantages of using AI and ML, specifically in preclinical trials.

While the evidence of 7 people is not representative of the entire Irish pharma sector it is sufficient to give us indications that might be pursued in later, larger, and better-resourced studies. Respondents did not predict wider usage of AI an ML in preclinical trials in Irish pharma companies, but as they have a high level of awareness of them, and the potential benefits they could bring to professionals in the sector, the

likelihood is that AI models will be deployed on trials over the next decade, perhaps starting with more basic applications in the areas of identification and counting, but progressing to what Aifiola (2022) described as “the potential to significantly enhance the currently tedious and costly preclinical drug discovery workflow providing fast and accurate results”.

5.2 Recommendations

Based on the findings and conclusions, the following recommendations can be made:

- Promote the use of AI and ML in drug discovery and clinical trial operations: These processes stand to gain significantly from AI and ML given their potential. But it's crucial to approach these technologies with prudence, objectivity, and morality. In order to secure the ethical use of these technologies, stakeholders, including researchers, regulators, and industry leaders, should promote responsible use of these technologies.

- Enhance data openness and quality: For AI and ML to produce accurate and trustworthy predictions, high-quality data is a requirement. Stakeholders should therefore concentrate on enhancing data transparency and quality in clinical trial and medication development procedures. This entails making sure that data is gathered consistently and ethically and that data exchange is transparent.

- Balance AI and ML with other approaches: While AI and ML offer a lot of potential benefits, it's important to balance them with other strategies to guarantee a thorough and correct strategy. Therefore, to ensure that the advantages of new technologies are realised while minimising the hazards, academics and stakeholders should think about merging AI and ML with conventional approaches.

- Address ethical and regulatory issues: The application of AI and ML to drug research and clinical trials poses a number of ethical and legal issues, including bias and data privacy. In order to ensure the proper use of new technologies, stakeholders should give top priority to addressing these concerns and creating precise ethical and legal rules.

- Create training and educational initiatives: With the use of AI and ML in drug discovery and clinical trial operations increasing, there is a need for specialised training and educational initiatives. In order to ensure

that researchers, regulators, and industry leaders have the abilities and information necessary to use new technologies ethically and successfully, stakeholders should implement training and education programmes.

- More analysis and research are required: Although AI and ML have the potential to be very useful in drug discovery and clinical trials, more analysis and research is required to fully comprehend their limitations and hazards. To guarantee that these technologies are used successfully and morally, stakeholders should prioritise more study and evaluation.

- Repeat the study with a larger target sample and a greater quantity of respondents. carry out a positivist and quantitative study to see if the issues explored and identified here holds true when they are discussed with a larger number of professionals in the pharma sector.

5.3 Implications

In different phases of drug discovery and clinical trials, artificial intelligence and machine learning have potential advantages and drawbacks that are highlighted in this paper. For the pharmaceutical sector and researchers, the findings have a number of consequences, including:

- The necessity of cooperation: Collaboration amongst various stakeholders, including researchers, doctors, and data scientists, is necessary for the use of AI and ML in drug discovery and clinical trials. The deployment of practical and morally sound methods depends on this teamwork.

- Requirement for high-quality data: The quality of the data used determines how well AI and ML models perform. Therefore, gathering and curating high-quality data that can be utilised to train these algorithms is crucial.

- Importance of ethical considerations: Patient privacy and data security are key ethical issues that are brought up by the use of AI and ML in drug research and clinical trials. Therefore, it is essential to give ethical issues top priority when creating and using these technologies.

- Clear regulatory guidance is required for the application of AI and ML to drug research and clinical trials. This advice can aid in ensuring the efficient and moral use of these technologies.

- Possibility to enhance patient outcomes: By facilitating the development of more effective and targeted medications, enhancing trial design and analysis, and lowering costs, the application of AI and ML in drug discovery and clinical trials has the potential to improve patient outcomes.

- Why Precaution is necessary: Although AI and ML have the potential to revolutionize drug research and clinical trials, caution is required when evaluating predictions and findings. Due to the complexity and

unpredictability of the drug development process, these strategies are only as effective as the data they are based on.

- Nonparametric Bayesian learning has the potential to be a useful tool for designing clinical trials, but it should be used in concert with other statistical techniques and thoroughly assessed for each study.

The use of AI and ML in drug discovery and clinical trials has the potential to enhance patient outcomes and speed up the drug development process, according to the study's findings overall, but caution and careful planning are required to ensure that these technologies are applied effectively and morally.

References

Ames, H., Glenton, C. and Lewin, S. (2019) 'Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication', *BMC Medical Research Methodology*, 19(1), p. 26. doi: 10.1186/s12874-019-0665-4.

Anantpadma, M. *et al.* (2019) 'Ebola Virus Bayesian Machine Learning Models Enable New in Vitro Leads', *ACS Omega*, 4(1), pp. 2353–2361. doi: 10.1021/acsomega.8b02948.

Antwi, S. K. and Hamza, K. (2015) 'Qualitative and Quantitative Research Paradigms in Business Research: A Philosophical Reflection', *European Journal of Business and Management*, 7(3), pp. 217–225. Available at: https://core.ac.uk/download/pdf/234626233.pdf.

Armstrong, J. (2010) 'Naturalistic Inquiry', in Salkind, N. (ed.) *Encyclopedia of Research Design*. Thousand Oaks, California, United States: SAGE Publications, Inc. doi: 10.4135/9781412961288.

Atradius (2022) *Pharmaceuticals Industry Trends Ireland - 2022*, *Atradius*. Available at: https://group.atradius.com/publications/industry-trends/pharmaceuticals-industry-trends-ireland-2022.html (Accessed: 7 May 2023).

Badria, F. A. (2020) *Drug Repurposing: Hypothesis, Molecular Aspects and Therapeutic Applications*. London, United Kingdom: Books on Demand. Available at: https://books.google.co.in/books?hl=en&lr=&id=_JQtEAAAQBAJ.

Badwan, B. A. *et al.* (2023) 'Machine learning approaches to predict drug efficacy and toxicity in oncology', *Cell Reports Methods*, 3(2), p. 100413. doi: 10.1016/j.crmeth.2023.100413.

Benzekry, S. *et al.* (2021) 'Machine Learning for Prediction of Immunotherapy Efficacy in Non-Small Cell Lung Cancer from Simple Clinical and Biological Data', *Cancers*, 13(24), p. 6210. doi: 10.3390/cancers13246210.

Bhuskute, H., Shende, P. and Prabhakar, B. (2021) '3D Printed Personalized Medicine for Cancer: Applications for Betterment of Diagnosis, Prognosis and Treatment', *AAPS PharmSciTech*, 23(1), p. 8. doi: 10.1208/s12249-021-02153-0.

Bohr, A. and Memarzadeh, K. (eds) (2020a) *Artificial Intelligence in Healthcare*. New York City, New York, United States: Elsevier Science. Available at: https://www.google.co.in/books/edition/Artificial_Intelligence_in_Healthcare/FDLXDwAAQBAJ.

Bohr, A. and Memarzadeh, K. (2020b) 'The rise of artificial intelligence in healthcare applications', in *Artificial Intelligence in Healthcare*. Elsevier, pp. 25–60. doi: 10.1016/B978-0-12-818438-7.00002-2.

Bryman, A. and Bell, E. (2011) *Business Research Methods*. 3rd edn. Oxford, England: OUP Oxford. Available at: https://books.google.co.in/books/about/Business_Research_Methods_3e.html?id=YnCcAQAAQBAJ.

Bunn, V. *et al.* (2020) 'Flexible Bayesian subgroup analysis in early and confirmatory trials', *Contemporary Clinical Trials*, 98, p. 106149. doi: 10.1016/j.cct.2020.106149.

Chen, X. *et al.* (2020) 'Discovery of Dual FGFR4 and EGFR Inhibitors by Machine Learning and Biological Evaluation', *Journal of Chemical Information and Modeling*, 60(10), pp. 4640–4652. doi: 10.1021/acs.jcim.0c00652.

Creswell, J. W. (2007) *Qualitative Inquiry and Research Design: Choosing among Five Approaches*. 2nd edn. Thousand Oaks, CA: Sage.

Creswell, J. W. (2009) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. London, United Kingdom: SAGE Publications. Available at: https://books.google.co.in/books/about/Research_Design.html?id=bttwENORfhgC.

CSO (2022) *Irish Industrial Production by Sector 2021*, *Central Statistics Office*. Available at: https://www.cso.ie/en/releasesandpublications/ep/p-iips/irishindustrialproductionbysector2021/ (Accessed: 7 May 2023).

Damm-Ganamet, K. L. *et al.* (2019) 'Accelerating Lead Identification by High Throughput Virtual Screening: Prospective Case Studies from the Pharmaceutical Industry', *Journal of Chemical Information and Modeling*, 59(5), pp. 2046–2062. doi: 10.1021/acs.jcim.8b00941.

Davenport, T. and Kalakota, R. (2019) 'The potential for artificial intelligence in healthcare', *Future Healthcare Journal*, 6(2), pp. 94–98. doi: 10.7861/futurehosp.6-2-94.

Dhakal, A. *et al.* (2022) 'Artificial intelligence in the prediction of protein–ligand interactions: recent advances and future directions', *Briefings in Bioinformatics*, 23(1). doi: 10.1093/bib/bbab476.

Donlin, M. J. *et al.* (2021) 'Discovery of 5-Nitro-6-thiocyanatopyrimidines as Inhibitors of Cryptococcus neoformans and Cryptococcus gattii', *ACS Medicinal Chemistry Letters*, 12(5), pp. 774–781. doi: 10.1021/acsmedchemlett.1c00038.

Dubois, P. and Sæthre, M. (2020) ‘On the Effect of Parallel Trade on Manufacturers’ and Retailers’ Profits in the Pharmaceutical Sector’, *Econometrica*, 88(6), pp. 2503–2545. doi: 10.3982/ECTA16039.

Duch, W., Swaminathan, K. and Meller, J. (2007) ‘Artificial Intelligence Approaches for Rational Drug Design and Discovery’, *Current Pharmaceutical Design*, 13(14), pp. 1497–1508. doi: 10.2174/138161207780765954.

Ekins, S. *et al.* (2019) ‘Exploiting machine learning for end-to-end drug discovery and development’, *Nature Materials*, 18(5), pp. 435–441. doi: 10.1038/s41563-019-0338-z.

Fakoor, R., Ladhak, F. and Nazi, A. (2013) ‘Using deep learning to enhance cancer diagnosis and classification’, in *30 th International Conference on Machine Learning*. Georgia, USA: JMLR: W&CP, p. 8. Available at: https://www.cs.columbia.edu/~faisal/publications/whealth.pdf.

Fogel, D. B. (2018) ‘Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review’, *Contemporary Clinical Trials Communications*, 11, pp. 156–164. doi: 10.1016/j.conctc.2018.08.001.

GlobalData (2021) *The end of the inversion: will pharma leave Ireland after tax rises?*, *Pharmaceutical Technology*. Available at: https://www.pharmaceutical-technology.com/comment/pharma-ireland-tax-rises/ (Accessed: 7 May 2023).

Gorgulla, C. *et al.* (2020) ‘An open-source drug discovery platform enables ultra-large virtual screens’, *Nature*, 580(7805), pp. 663–668. doi: 10.1038/s41586-020-2117-z.

Hamet, P. and Tremblay, J. (2017) ‘Artificial intelligence in medicine’, *Metabolism*, 69, pp. S36–S40. doi: 10.1016/j.metabol.2017.01.011.

Harrer, S. *et al.* (2019) ‘Artificial Intelligence for Clinical Trial Design’, *Trends in Pharmacological Sciences*, 40(8), pp. 577–591. doi: 10.1016/j.tips.2019.05.005.

Hassanzadeh, P., Atyabi, F. and Dinarvand, R. (2019) ‘The significance of artificial intelligence in drug delivery system design’, *Advanced Drug Delivery Reviews*, 151–152, pp. 169–190. doi: 10.1016/j.addr.2019.05.001.

He, S., Zhang, X., *et al.* (2019) ‘A Computational Toxicology Approach to Screen the Hepatotoxic Ingredients in Traditional Chinese Medicines: Polygonum multiflorum Thunb as a Case Study’, *Biomolecules*, 9(10), p. 577. doi: 10.3390/biom9100577.

He, S., Zhang, C., *et al.* (2019) ‘Herb-Induced Liver Injury: Phylogenetic Relationship, Structure-Toxicity Relationship, and Herb-

Ingredient Network Analysis', *International Journal of Molecular Sciences*, 20(15), p. 3633. doi: 10.3390/ijms20153633.

Ho, C.-Y., Li, X. and Zhou, W. (2017) *Foreign Direct Investment Spillovers And Pharmaceutical Innovation: The Role Of Intellectual Property Rights*. Available at: https://www.adb.org/sites/default/files/publication/357656/adbi-wp775.pdf.

Hupf, B. *et al.* (2021) 'Bayesian semiparametric <scp>meta-analytic-predictive</scp> prior for historical control borrowing in clinical trials', *Statistics in Medicine*, 40(14), pp. 3385–3399. doi: 10.1002/sim.8970.

Isaksson, L. J. *et al.* (2020) 'Machine Learning-Based Models for Prediction of Toxicity Outcomes in Radiotherapy', *Frontiers in Oncology*, 10. doi: 10.3389/fonc.2020.00790.

Jordan, A. M. (2018) 'Artificial Intelligence in Drug Design—The Storm Before the Calm?', *ACS Medicinal Chemistry Letters*, 9(12), pp. 1150–1152. doi: 10.1021/acsmedchemlett.8b00500.

Julkunen, H. *et al.* (2020) 'Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects', *Nature Communications*, 11(1), p. 6136. doi: 10.1038/s41467-020-19950-z.

Kaushik, V. and Walsh, C. A. (2019) 'Pragmatism as a Research Paradigm and Its Implications for Social Work Research', *Social Sciences*, 8(9), p. 255. doi: 10.3390/socsci8090255.

Kivunja, C. and Kuyini, A. B. (2017) 'Understanding and Applying Research Paradigms in Educational Contexts', *International Journal of Higher Education*, 6(5), p. 26. doi: 10.5430/ijhe.v6n5p26.

Koçak, B. *et al.* (2023) 'Must-have Qualities of Clinical Research on Artificial Intelligence and Machine Learning', *Balkan Medical Journal*, 40(1), pp. 3–12. doi: 10.4274/balkanmedj.galenos.2022.2022-11-51.

Kola, I. and Landis, J. (2004) 'Can the pharmaceutical industry reduce attrition rates?', *Nature Reviews Drug Discovery*, 3(8), pp. 711–716. doi: 10.1038/nrd1470.

Kolluri, S. *et al.* (2022) 'Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review', *The AAPS Journal*, 24(1), p. 19. doi: 10.1208/s12248-021-00644-3.

Krishnaveni, C., Arvapalli, S. and Sharma, J.V.C. (2019) 'ARTIFICIAL INTELLIGENCE IN PHARMA INDUSTRY- A Review'.

Li, M. *et al.* (2020) 'Bayesian Semi-parametric Design (BSD) for

adaptive dose-finding with multiple strata', *Journal of Biopharmaceutical Statistics*, 30(5), pp. 806–820. doi: 10.1080/10543406.2020.1730870.

Liang, G. *et al.* (2020) 'The emerging roles of artificial intelligence in cancer drug development and precision therapy', *Biomedicine & Pharmacotherapy*, 128, p. 110255. doi: 10.1016/j.biopha.2020.110255.

Liang, Z. *et al.* (2014) 'Deep learning for healthcare decision making with EMRs', in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, pp. 556–559. doi: 10.1109/BIBM.2014.6999219.

Lin, W.-C. *et al.* (2020) 'Applications of Artificial Intelligence to Electronic Health Record Data in Ophthalmology', *Translational Vision Science & Technology*, 9(2), p. 13. doi: 10.1167/tvst.9.2.13.

Lipinski, C. A. *et al.* (2001) 'Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997)', *Advanced Drug Delivery Reviews*, 46(1–3), pp. 3–26. doi: 10.1016/S0169-409X(00)00129-0.

Lipinski, C. F. *et al.* (2019) 'Advances and Perspectives in Applying Deep Learning for Drug Design and Discovery', *Frontiers in Robotics and AI*, 6. doi: 10.3389/frobt.2019.00108.

Liu, R., Lin, J. and Li, P. (2020) 'Design considerations for phase I/II dose finding clinical trials in Immuno-oncology and cell therapy', *Contemporary Clinical Trials*, 96, p. 106083. doi: 10.1016/j.cct.2020.106083.

Liu, Siqi *et al.* (2014) 'Early diagnosis of Alzheimer's disease with deep learning', in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, pp. 1015–1018. doi: 10.1109/ISBI.2014.6868045.

Liu, Z. *et al.* (2019) 'DeepScreening: a deep learning-based screening web server for accelerating drug discovery', *Database*, 2019. doi: 10.1093/database/baz104.

Lombardo, F. *et al.* (2017) 'In Silico Absorption, Distribution, Metabolism, Excretion, and Pharmacokinetics (ADME-PK): Utility and Best Practices. An Industry Perspective from the International Consortium for Innovation through Quality in Pharmaceutical Development', *Journal of Medicinal Chemistry*, 60(22), pp. 9097–9113. doi: 10.1021/acs.jmedchem.7b00487.

Lui, T. K. L., Cheung, K. S. and Leung, W. K. (2022) 'Machine learning

models in the prediction of 1-year mortality in patients with advanced hepatocellular cancer on immunotherapy: a proof-of-concept study', *Hepatology International*, 16(4), pp. 879–891. doi: 10.1007/s12072-022-10370-3.

Luke, J., Porter, D. and Santhanam, P. (2020) *Beyond Algorithms: Delivering AI for Business*. New York City, New York, United States: CRC Press. Available at: https://www.google.co.in/books/edition/Beyond_Algorithms/BUdpEAAAQBAJ.

Lv, H. *et al.* (2021) 'Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design', *Briefings in Bioinformatics*, 22(6). doi: 10.1093/bib/bbab320.

Ma, F. *et al.* (2018) 'Risk Prediction on Electronic Health Records with Prior Medical Knowledge', in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: ACM, pp. 1910–1919. doi: 10.1145/3219819.3220020.

Martin, Y. C. (2005) 'A Bioavailability Score', *Journal of Medicinal Chemistry*, 48(9), pp. 3164–3170. doi: 10.1021/jm0492002.

Mendez, D. *et al.* (2019) 'ChEMBL: towards direct deposition of bioassay data', *Nucleic Acids Research*, 47(D1), pp. D930–D940. doi: 10.1093/nar/gky1075.

Merriam, S. B. (2009) *Qualitative research: a guide to deisgn and implemenation*. San Fransico: John Wiley & Sons. Available at: https://books.google.co.in/books/about/Qualitative_Research.html?id=tvFICrgcuSIC&redir_esc=y.

Mohs, R. C. and Greig, N. H. (2017) 'Drug discovery and development: Role of basic biological research', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(4), pp. 651–657. doi: 10.1016/j.trci.2017.10.005.

Neubauer, B. E., Witkop, C. T. and Varpio, L. (2019) 'How phenomenology can help us learn from the experiences of others', *Perspectives on Medical Education*, 8(2), pp. 90–97. doi: 10.1007/s40037-019-0509-2.

O'Dwyer, L., Nolan, L. and Fisher, C. (2017) 'Supporting Innovation through Regulation and Science: Ireland as an Innovation Hub for Health Products', *Biomedicine Hub*, 2(Suppl. 1), pp. 1–8. doi: 10.1159/000481427.

Pantuck, A. J. *et al.* (2018) 'Modulating BET Bromodomain Inhibitor ZEN-3694 and Enzalutamide Combination Dosing in a Metastatic Prostate Cancer Patient Using CURATE.AI, an Artificial Intelligence

Platform', *Advanced Therapeutics*, 1(6), p. 1800104. doi: 10.1002/adtp.201800104.

Patel, V. and Shah, M. (2022) 'Artificial intelligence and machine learning in drug discovery and development', *Intelligent Medicine*, 2(3), pp. 134–140. doi: 10.1016/j.imed.2021.10.001.

Paul, D. *et al.* (2021) 'Artificial intelligence in drug discovery and development', *Drug Discovery Today*, 26(1), pp. 80–93. doi: 10.1016/j.drudis.2020.10.010.

Rafique, R., Islam, S. M. R. and Kazi, J. U. (2021) 'Machine learning in the prediction of cancer therapy', *Computational and Structural Biotechnology Journal*, 19, pp. 4003–4017. doi: 10.1016/j.csbj.2021.07.003.

Ramesh, A. *et al.* (2004) 'Artificial Intelligence in Medicine'. *Annals of The Royal College of Surgeons of England*, 86(5), pp. 334–338. DOI: 10.1308/147870804290.

Ridder, H.-G. (2017) 'The theory contribution of case study research designs', *Business Research*, 10(2), pp. 281–305. doi: 10.1007/s40685-017-0045-z.

Ripphausen, P. *et al.* (2010) 'Quo Vadis, Virtual Screening? A Comprehensive Survey of Prospective Applications', *Journal of Medicinal Chemistry*, 53(24), pp. 8461–8467. doi: 10.1021/jm101020z.

Russell, S. and Norvig, P. (2016) *Artificial Intelligence: A Modern Approach*. CreateSpace Independent Publishing Platform. Available at: https://www.google.co.in/books/edition/Artificial_Intelligence/PQI7vgAACAAJ.

Saeed, G. *et al.* (2022) 'A systematic review of digital technology and innovation and its potential to address anti-corruption, transparency, and accountability in the pharmaceutical supply chain', *Expert Opinion on Drug Safety*, 21(8), pp. 1061–1088. doi: 10.1080/14740338.2022.2091543.

Sarkar, C. *et al.* (2023) 'Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development'. *International Journal of Molecular Sciences*, 24(3), p. 2026. DOI: 10.3390/ijms24032026.

Saunders, M. N. K., Lewis, P. and Thornhill, A. (2015) *Research Methods for Business Students*. 7th edn. Upper Saddle River, New Jersey, United States: Pearson. Available at: https://www.pearson.com/uk/educators/higher-education-educators/program/Saunders-Research-Methods-for-Business-Students-7th-Edition/PGM1089011.html.

Saunders, M., Lewis, P. and Thornhill, A. J. (2003) *Research Methods for Business Students*. 3rd edn. Harlow: FT Prentice Hall.

Saville, B. R. and Berry, S. M. (2016) 'Efficiencies of platform clinical trials: A vision of the future', *Clinical Trials*, 13(3), pp. 358–366. doi: 10.1177/1740774515626362.

Saxena, K., Balani, S. and Srivastava, P. (2022) 'The role of pharmaceutical industry in building resilient health system', *Frontiers in Public Health*, 10. doi: 10.3389/fpubh.2022.964899.

Schuhmacher, A. *et al.* (2021) 'Big Techs and startups in pharmaceutical R&D – A 2020 perspective on artificial intelligence', *Drug Discovery Today*, 26(10), pp. 2226–2231. doi: 10.1016/j.drudis.2021.04.028.

Sertkaya, A. *et al.* (2016) 'Key cost drivers of pharmaceutical clinical trials in the United States', *Clinical Trials*, 13(2), pp. 117–126. doi: 10.1177/1740774515625964.

Shah, P. *et al.* (2019) 'Artificial intelligence and machine learning in clinical development: a translational perspective', *npj Digital Medicine*, 2(1), p. 69. doi: 10.1038/s41746-019-0148-3.

Shashi, M. and Gossett, K. (2022) 'Exploring Strategies to Leveraging the Blockchain in Pharmaceutical Supply Chains', *International Journal of Research and Analytical Reviews*, 9(3), pp. 284–290. doi: 10.1729/Journal.31224.

Sosa-Díaz, M.-J. and Valverde-Berrocoso, J. (2022) 'Grounded Theory as a Research Methodology in Educational Technology', *International Journal of Qualitative Methods*, 21, p. 160940692211332. doi: 10.1177/16094069221133228.

Soufan, O. *et al.* (2018) 'DPubChem: a web tool for QSAR modeling and high-throughput virtual screening', *Scientific Reports*, 8(1), p. 9110. doi: 10.1038/s41598-018-27495-x.

Spanbauer, C. and Sparapani, R. (2021) 'Nonparametric machine learning for precision medicine with longitudinal clinical trials and Bayesian additive regression trees with mixed models', *Statistics in Medicine*, 40(11), pp. 2665–2691. doi: 10.1002/sim.8924.

Sutton, R. T. *et al.* (2020) 'An overview of clinical decision support systems: benefits, risks, and strategies for success', *npj Digital Medicine*, 3(1), p. 17. doi: 10.1038/s41746-020-0221-y.

Tao, N. *et al.* (2008) 'Chemical composition of essential oil from the peel of Satsuma mandarin', *African Journal of Biotechnology*, 7(9), pp. 1261–1264. doi: 10.5897/AJB08.033.

Thomas, B. (2021) ‘Artificial Intelligence: Review of Current and Future Applications in Medicine’. *Federal Practitioner*, (38 (11)). DOI: 10.12788/fp.0174.

Tie, Y. C., Birks, M. and Francis, K. (2019) ‘Grounded theory research: A design framework for novice researchers’, *SAGE Open Medicine*, 7, p. 205031211882292. doi: 10.1177/2050312118822927.

Vaismoradi, M. *et al.* (2016) ‘Theme development in qualitative content analysis and thematic analysis’, *Journal of Nursing Education and Practice*, 6(5). doi: 10.5430/jnep.v6n5p100.

Vamathevan, J. *et al.* (2019) ‘Applications of machine learning in drug discovery and development’, *Nature Reviews Drug Discovery*, 18(6), pp. 463–477. doi: 10.1038/s41573-019-0024-5.

Vijayan, R. S. K. *et al.* (2022) ‘Enhancing preclinical drug discovery with artificial intelligence’, *Drug Discovery Today*, 27(4), pp. 967–984. doi: 10.1016/j.drudis.2021.11.023.

Wager, T. T. *et al.* (2016) ‘Central Nervous System Multiparameter Optimization Desirability: Application in Drug Discovery’, *ACS Chemical Neuroscience*, 7(6), pp. 767–775. doi: 10.1021/acschemneuro.6b00029.

Wang, K. *et al.* (2021) ‘DeepDTAF: a deep learning method to predict protein–ligand binding affinity’, *Briefings in Bioinformatics*, 22(5). doi: 10.1093/bib/bbab072.

Wang, Y. *et al.* (2017) ‘PubChem BioAssay: 2017 update’, *Nucleic Acids Research*, 45(D1), pp. D955–D963. doi: 10.1093/nar/gkw1118.

van de Waterbeemd, H. and Gifford, E. (2003) ‘ADMET in silico modelling: towards prediction paradise?’, *Nature Reviews Drug Discovery*, 2(3), pp. 192–204. doi: 10.1038/nrd1032.

Webb, G. I. *et al.* (2011) ‘Dirichlet Process’, in *Encyclopedia of Machine Learning*. Boston, MA: Springer US, pp. 280–287. doi: 10.1007/978-0-387-30164-8_219.

Weissler, E. H. *et al.* (2021) ‘The role of machine learning in clinical research: transforming the future of evidence generation’, *Trials*, 22(1), p. 537. doi: 10.1186/s13063-021-05489-x.

Woodcock, J. and LaVange, L. M. (2017) ‘Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both’, *New England Journal of Medicine*. Edited by J. M. Drazen et al., 377(1), pp. 62–70. doi: 10.1056/NEJMra1510062.

Wu, J. (2019) *AI and Cognitive Computing*, *Digital Leaders*. Available at: https://digileaders.com/ai-and-cognitive-computing/ (Accessed: 11

April 2023).

Young, R. J. *et al.* (2011) 'Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity', *Drug Discovery Today*, 16(17–18), pp. 822–830. doi: 10.1016/j.drudis.2011.06.001.

Zhang, D. *et al.* (2020) 'In silico screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus', *Journal of Integrative Medicine*, 18(2), pp. 152–158. doi: 10.1016/j.joim.2020.02.005.

Zhang, L. *et al.* (2013) 'Discovery of Novel Antimalarial Compounds Enabled by QSAR-Based Virtual Screening', *Journal of Chemical Information and Modeling*, 53(2), pp. 475–492. doi: 10.1021/ci300421n.

Zhang, L. *et al.* (2017) 'From machine learning to deep learning: progress in machine intelligence for rational drug discovery', *Drug Discovery Today*, 22(11), pp. 1680–1685. doi: 10.1016/j.drudis.2017.08.010.

Zheng, C. *et al.* (2023) 'Identification of Recurrent Atrial Fibrillation using Natural Language Processing Applied to Electronic Health Records', *European Heart Journal - Quality of Care and Clinical Outcomes*. doi: 10.1093/ehjqcco/qcad021.

Zhou, J. and Troyanskaya, O. G. (2015) 'Predicting effects of noncoding variants with deep learning–based sequence model', *Nature Methods*, 12(10), pp. 931–934. doi: 10.1038/nmeth.3547.

Zhu, T. *et al.* (2013) 'Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis', *Journal of Medicinal Chemistry*, 56(17), pp. 6560–6572. doi: 10.1021/jm301916b.

Appendix -A

Questionnaires

1. What do you see as the main benefits of using AI and ML in the pre-clinical stage of drug development?
2. How can AI and ML improve the efficiency of translational research?
3. In what ways can AI and ML help accelerate drug discovery?
4. Are there any potential drawbacks to using AI and ML in the pre-clinical, translational and drug delivery stages?
5. In your experience, how has the use of AI and ML impacted the efficiency of clinical trial operations?
6. What are the key advantages of using AI and ML for clinical trial data analysis?
7. How familiar are you with the use of AI and ML in clinical trials?
8. What are some of the challenges associated with implementing AI and ML in clinical trials?
9. What steps do you think need to be taken to ensure that the use of AI and ML in clinical trials is ethical and responsible?
10. Can you describe how nonparametric Bayesian learning works in the context of clinical trial design?
11. What are some of the benefits of using this nonparametric Bayesian learning approach in clinical trial design?
12. Are there any limitations or drawbacks to using nonparametric Bayesian learning in CT design?
13. How does the use of nonparametric Bayesian learning affect the role of statisticians in clinical trial design?
14. In your opinion, what are some other applications of nonparametric Bayesian learning in the pharmaceutical industry beyond clinical trial design?