# Regulatory Submission Intelligence and Benefit-Risk Signal Extraction: AI-Driven Frameworks for Optimising Drug Approval Pathway Navigation

*Dr. Anibal Traça, Professor of Informatics, Instituto Superior Técnico (IST), Portugal*

## 1. Introduction

Current approaches to drug approval processes have been the subject of criticism and are in need of innovation due to the prevalence of inefficiencies that have resulted in the requirement of years and billions of dollars to bring new treatments to the market. Clinical approval processes involve a cascade of largely sequential examination steps with high significance and expediency, in which the decision at each step ultimately leads to the final approval of the drug. Thus, improvement in the decision-making process may significantly advance the evaluation of such new drugs and ultimately reduce the duration and cost of the entire drug approval process. Recently, the potential of AI-based technologies in the medical field has been profoundly demonstrated, leading to their worldwide adaptation in various medical fields to achieve faster and more accurate diagnoses and treatment plans. Similarly, there is great potential in research and development in the use of AI technologies. AI-based innovative technologies have been suggested to achieve further improvement in effectiveness and efficiency. For instance, the potential use of AI to help optimize the design of clinical research trials may eventually lead to significant advances in drug approval processes.

In the current scenario, developing a new drug in a major pharmaceutical company costs approximately $1.3 billion to $3 billion and requires 10 years on average. Moreover, the net rate of regulatory approval has shown a downward trend over the past decades, associated with increasing requirements for the diseases to be evaluated, achieving statistical power, and the superiority of the investigational product over its comparator. This complexity has resulted in an increasing proportion of terminated or abandoned studies. Hence, the development of innovative technologies could be amplified as drug development approaches new challenges, such as rare diseases, next-

generation precision health, and public health. To thoroughly achieve the effect of the system, it is essential to consider the end-to-end research and development approach, leading up to the drug approval process, the use of drugs, and their subsequent associated health-related outcomes. Therefore, the development and evaluation of innovative drugs should be redefined with attention to the candidate patients for clinical trials and the regulatory hurdles to be overcome. We provide the current landscape of the research and development process in pharmaceuticals and the current trends in drug approval processes. Additionally, we introduce the main components in machine learning and how these can potentially accelerate the decision-making process in the drugs' approval by regulatory agencies.

### 1.1. Background and Significance

Regulatory approval of drugs typically involves safety and efficacy assessments, the requirements of which have evolved over time as a response to industry, society, and technological developments. Long ago, regulations emerged to prevent harm from toxic substances, and the Federal Food, Drug, and Cosmetic Act would later require companies to test for safety. Efficacy evaluation took greater importance following amendments that necessitated formal scientific proof of drug efficacy. Today's approval requirements are guided by interpretations of legal standards, and the current view is that approval must ensure that the benefits of drugs outweigh risks. Although this is an attractive safety precaution, high standards and narrow indications increase the cost of clinical studies and constrict the drug's revenue opportunities. The biopharmaceutical industry finances clinical programs, but when approval is delayed or disallowed, the time and money invested cannot be recouped at the original speed. It costs over $2.5 billion and about 14 years to bring one new drug to market; small-molecule drugs have just a 9.6% chance of approval. The time taken is detrimental to public health, and the requirements and delays have a chilling effect on the development of new and urgently needed products; an exception that has occurred throughout history is in times of global unrest and contagious outbreaks. In the late 1950s, a call for an anti-cancer project led to significant investment in top biomedical research; years later, regulatory flexibility would allow the rapid development of drugs symptomatic of the move towards a 'war' on cancer and infection. Another bottleneck revealed by crisis was the discovery and development of vaccines; this technology is booming, and so is the disease load driving

the demand for quick approvals. To continue the influx of pharmaceutical innovation, regulatory efficiency should become a top national priority.

The global response to COVID-19 shows how politics and human rationality can cause dramatic shifts in economic activity. The biopharmaceutical industry is currently trying to produce a vaccine for COVID-19 in less than 18 months; this rapid pace has never been seen and requires technology and regulatory advancements. Especially in challenging times, regulators seek to provide emergency use and compassionate-use designations to assuage suffering and fear and allow adequate drug access. Given the increased public discontent with long time-to-market and emergency use authorizations, stakeholders are accelerating access to treatments through potential means—advanced therapy-specific parallel scientific assistance and other positive measures. The above considerations suggest that the regulation of drugs is currently far less slow and expensive than it could be under the existence of some common sense; numerous changes could be tested for the introduction of a global trial phase, algorithm-aided approvals, and the introduction of affiliated-revenue regulatory flexibility mechanisms. Limited in scope, this paper highlights an unappreciated, direct approach to bring about reform: optimization. In particular, AI embeds efficient and thus more affordable pathways for drug evaluation and approval.

### 1.2. Research Objectives

As indicated above, we are focused on the drug approval process, and the objective of our research is to investigate how AI-driven technologies can help address the inherent inefficiencies of drug approval processes and tools. In particular, this research will explore the possible focus areas for AI technologies and assess the extent of their effectiveness in delivering supplements or replacements for existing tools and processes. The primary goal is to observe the impact of AI in drug approval timelines and decision-making processes and study the key criteria for decision support. We aim to trace and expand upon the role of AI technologies used by global regulatory agencies to support the objectives mentioned above. Additionally, we are interested in identifying any obstacles or challenges in deploying such tools effectively. It is necessary to stress the need to empirically test the effectiveness of these tools in real-world settings in order to inform decisions and advance the field of AI in regulatory science. Therefore, we also consider the need to delve into specific case studies that will benefit from an AI-focused

tool. We want to study a range of stakeholder attitudes on the successful deployment of AI technologies in the regulatory review process and propose evaluative criteria that can be used by policymakers and top management at pharmaceutical companies to gauge the AI proposal's efficacy.

**2. Current Challenges in Drug Approval Processes**

According to current procedures, approval of drugs is subject to the assessment of submitted clinical trials by regulatory agencies. The submission of dossiers and responses to requests for further information, as well as the review process, typically consume years. High direct and incidental costs promote risk aversion in the pharmaceutical industry. Unclear regulatory guidelines and poor reproducibility studies are major obstacles to pharmaceutical innovation, along with difficult access for small and medium-sized enterprises and stakeholders from emerging countries. Innovative study design with ready-made procedures, as well as digital drug development strategies, promise optimization. High hurdles for the approval of new technologies are particularly large, and regulatory authorities are very cautious in cases of doubt, as is currently the case for medical artificial intelligence. Each country has its own rules on clinical trials, quality, and consumer protection. The pooling of data from unrelated clinical trials also remains a niche offering. First and foremost, the reduction of the time required for clinical studies would promote competition and thus the economic price of drug development internationally. This would require the recognition and approval of patient groups and surrogate parameters such as biomarkers. Two key steps in the development of pharmaceuticals present until disagreement on the epitope level is the worldwide approval of a phase III design by all health authorities. Regulatory strategies differ fundamentally between individual countries, with completely different designs of the approval strategy depending on country-specific approval requirements. For the technology or biotechnology product for the treatment of unmet medical needs, innovative phase III designs are desirable. With heavy antibiotic resistance uncontrolled for commercial reasons, these must be initiated in all concerned countries. Aligning the so-called health authority assays is not possible on a case-by-case basis due to the socio-economic needs and epidemiological situation in each country. Predominant elimination pathways, such as metabolism and primary non-renal elimination, and evaluation of immunogenicity can also be checked in a country-specific approval strategy earlier and evaluated on the basis of a topical study with 50 or 100 subjects.

### 2.1. Regulatory Hurdles

From discovery to approval, the drug development process comprises several regulatory hurdles. While all regulatory interventions aim to ensure patients' protection, bottlenecks or shortcomings in various stages of this process primarily contribute to delays. The formal decision-making procedures require large multinational drug discovery companies an average of 12 years to personalize their respective investment in R&D of the candidate, after which it takes a further 1 to 4 years for the Document of Compliance and corresponding marketing assets to be approved. In addition, the costs of achieving acceptance may be greater than USD 10,000, with the probability of success widely uncertain. Costs related to developing new drugs are estimated to be between $1 and $2.6 billion per NME.

In countries with more straightforward bureaucratic processes, central healthcare organizations typically require market entry documentation, which ends up in ground-based assets. The deal, marketing, and price policies of the low to high pharmaceutical manufacturers must be carefully weighed to satisfy bureaucratic requirements and regulatory compliance rather than put patients into care. This part also requires a large sum of installment money, such as €9,000 per floor privacy, and an expensive clinical development program, €10,000 and upwards. Tools are now evolving to identify ways that can help pharmaceutical companies make the documentation process faster and more effective to address these dilemmas. This would remove bottlenecks, reflecting the regulatory requirements of the questioned pharmaceutical companies' documentation, and their influence is useful. Given these initiatives, healthcare agencies and drug agencies must aim to succeed in protecting the health of clinical trial patients and consumers while eliminating obstacles to market entry, thus increasing patients' access to improved therapy.

## 3. Machine Learning Applications in Healthcare

Machine learning is a type of computational method that allows systems to automatically learn from previous data patterns, characteristics, or experiences and provide a basis for future decision-making using algorithms without explicit instructions. One of the breakthroughs of machine learning is recognizing patterns in large data sets. The difference between traditional computational models and machine learning models is that the latter allows for complex models and identifies patterns too

complex for human cognition. AI can help with many of the barriers and data analysis while clinical trials are being completed to get the drug through the regulatory process more quickly and used in the clinic more efficiently. Right now, after clinical trials, a drug that gets through the FDA in terms of approval or new indication takes 1-2 years to get a guideline, then an additional 5 to 6 years to get that drug approved through NICE. With machine learning, we can substantially reduce the time spent in the guideline approval process.

The rise in the precision and uniqueness of diagnosis and treatment, with the requirement of personalization, pushes the unattainable boundaries of traditional medicine to one that could be solvable only in the exponential rise of digital data beyond human cognition. Machine learning can be explained as a problem-solving mechanism utilizing actions or patterns as data with a database of sufficiently large cases that the machine can learn from. In healthcare, it primarily aims to replicate the manageable digital world of data within human cognitive limits. Many algorithms today have surpassed the capacity of experts to trust human cognition alone, like self-driving cars, and the use of machine learning is extending more and more, such as in social media, airport security, and harvesting pictures. However, the ruling system in healthcare is one that is called "explainable AI." This is one step ahead of machine learning; it could provide insights that cannot be comprehended by our cognitive levels through the large masses of data and computing techniques. Right now, we're working to bring those deep learning tools into the clinic, but the path forward for most machine learning in general and deep learning specifically is the smoothing of the path to getting regulatory acceptance of using AI/ML as an automated tool in healthcare.

### 3.1. Overview of Machine Learning

Machine learning is a subfield of artificial intelligence. It is an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. There are three types of ML that have a wide variety of applications in different data-driven problem solving, especially in the healthcare domain, including drug development: 1. Supervised learning 2. Unsupervised learning 3. Reinforcement learning Using ML and specifically supervised ML, methods and algorithms have created systems that can learn from data. Data and corresponding outcomes are often used as input to the algorithms for these models. The

outcomes may have different types of information such as categorization, regression, or even ranking. As models ingest more data and outcomes, they continuously improve over time and can make better decisions. The most important aspect of these algorithms and models is their continuous learning ability—meaning the more data fed into these models, the better results produced. This implies the significance of data or analytics and its quality in training these algorithms.

Machine learning technologies continue to evolve and become better, and can easily adapt or change their own model when new data is presented. In other words, the AI system of algorithms can discover the unknowns, and often, that leads to deeper insights. It has the ability to solve problems that are not in the original program or new challenges, often called robustness. ML is a highly dynamic area with great potential for revolutionizing the healthcare industry. The area of drug development, in particular, will benefit greatly from such models. The more data these algorithms ingest, the better the predictions and decisions. These machine learning predictions have become as good as leading clinicians in predicting the best treatment for patients. These models can now cover clinical trials, healthcare providers, and ultimately payers and patients, to provide insight beyond diagnostics on how the disease should be treated.

**4. AI Solutions for Drug Approval Optimization**

In response to drug innovations, several AI-driven solutions have been developed with the precise goal of optimizing and transforming the drug approval processes to better manage resource allocation and speed up the regulatory process in order to increase the drugs' approval with new mechanisms of action to benefit the patients. The AI techniques used in the proposed solutions are different actual state-of-the-art machine learning frameworks, integrated with the latest in regulatory science and advanced analytics. The applications of the AI-based solutions include, among others, the analysis of regulatory submissions to model the profiling of those drugs that are more likely to obtain an expedited review and approval. The model can identify on a calendar time the ones that can generate queries that prevent rapid approval. Therefore, AI can support the future approval of candidate drugs, ensuring that applicants provide effective pro-regulatory platforms.

The relationship between advanced analytics, natural language processing, modeling, predictive analytics, and area of interest monitoring has deepened as there are several

AI solutions that keep evolving. One of these applications is predictive analytics: using advanced analytics, a model could, for example, predict the survival advantage based on a posteriori data by the regulatory agency, originating the probability of generating a classifier developed for different stakeholders. AI could enhance automatic interpretation that can improve each R&D profitability by facilitating decision-making, advocacy, and audit in the innovation phase. By using advanced analytics, the interpretation of the outcomes is still a critical step in the entire process. The traceability option of the result includes the conversion of the narrative into objective terms that a computer can understand. New systems can be easily implemented only if AI is recognized as a tool to aid the current available workflow by integrating AI in the regulatory guideline as an option. One advantage of involving the need for an additional regulatory perspective is the increase in collaboration between all stakeholders in the approval of a new drug in order to identify the right studies for the best and most rapid form that a new molecule can take to make it to market. At the same time, it is necessary to increase the number of case studies with real data where the application of artificial intelligence was decisive in approving the high unmet medical need. One case of a successful use of AI was the analysis of previously available data in order to optimize and put in place acceptable clinical development agreed upon with the regulatory agency: this application was directed toward supporting scientific advice, and the most expedited result is more powerful because it is based on the largest available majority of scientific information. In conclusion, AI needs content, better when linked to all the written and non-written knowledge of an area in order to be useful for further reducing the risk of non-innovative development. In this perspective, natural language processing can help applicants fine-tune research projects in order to fast-track market results because it may allow for quantification and qualification without additional clinical studies data, which can normally reduce the development needed for the part already known of the product.

### 4.1. Data Analytics and Predictive Modeling

A cornerstone of AI-driven drug approval optimization is the power of data, analytics, and predictive modeling. From the generation of data from real-world patients that are suitable for use in drug approval processes to decisions and insights based on the analysis of data, analytics play a crucial role. AI-driven data analytics can transform raw data into a stream of meaningful information about patients, treatment effects, and

disease patterns. Apart from 'descriptive' analytics, modeling and prediction provide insights needed to make decisions and act optimally in advance. These range from statistical methods, such as adjusted comparisons between 'treated' versus 'non-treated', to machine learning tools. These are more complex and can be used for automatically connecting high-dimensional patient and treatment spaces and to build models to provide outcome predictions.

Predictive modeling tools are particularly useful to estimate unknown quantities, treatment effects, and optimal parameters for trial designs. AI-driven predictive models have already been acknowledged for their ability to predict clinical trial outcomes. They can be used to combine multiple different trial endpoints into a single multivariate composite outcome, to account for the coprimary nature of signal endpoints. A strength of these models is that they may predict outcomes far in advance, giving sufficient time to modify the original drug development plan and study protocol. Data-driven optimization for drug approval purposes has shown superiority over traditional RCTs because it enables data integration across multiple institutions and test centers while fully maintaining data quality, ethical, and safety standards. Leveraging cross-institutional real-world evidence can significantly reduce regulatory decision-making times, ensuring approved drugs are released faster into the market. Nonetheless, researchers and AI should be vigilant about the ethical and social consequences of such tools.

**5. Case Studies and Success Stories**

AI-driven analytics can speed up the clinical trial approval procedure. There are real-world examples of such a solution, where innovative AI algorithms work hand in hand with organizations to reduce the approval timeframes. The resulting success stories confirm that the R&D phase of new drug development could be efficiently accelerated by implementing automatically analyzed data into the previously manual regulatory approval procedures for investigating new drugs. The key players in optimizing the approval process for new drugs in clinical trials and protocols for new indications were pharmaceutical companies and regulatory bodies. Those wishing to carry out clinical trials offered potential candidates for cases. A single criterion was developed and mutually agreed upon at the beginning of the procedure to be satisfied by the AI algorithms; their output should be directly included in the assessment process.

In this way, an effective solution that complements the assessment of the experts was assured. There are case studies that confirm that initially emerging AI-based algorithms, trained and adapted to very specific needs, have the potential to assist regulators in speeding up specific steps. However, there are some common lessons and conclusions. The practical exercise to investigate existing technical and organizational solutions based on AI, natural language processing, and machine learning data analytics to support certain regulatory procedures included use cases that are technically and legally limited to supporting the financed medicine area of regulatory procedures. The cases demonstrated the feasibility of such solutions to support the shortening of regulatory processes. A legal entity engages in clinical trial submissions and reviews or in human medicinal product marketing authorization common assessment procedures; the legal frameworks need to explicitly include such systems and services as beneficiaries.

### 5.1. Real-world Implementation

AI is taking off in various drug approval settings, both in the public and private sectors. Companies working in these spaces are increasingly utilizing AI to optimize the drug R&D process, based on data generated by regulatory agencies in an effort to decrease attrition rates in early drug development. Major organizations experimenting with AI in drug development span various regulatory bodies. In addition, a multi-stakeholder consortium launched an initiative to modernize health technology assessment to capture real-world evidence to support the use and payment of drugs across several jurisdictions. AI that is based on a new paradigm of interdisciplinary collaboration between scientists, regulators, and technologists is propelling this real-world implementation effort. Some of the technologies used are unsupervised machine learning, natural language processing for text mining, and statistical and agent-based models for simulating clinical trial design and recruitment. A cognitive computing platform is using a powerful AI engine that brings together all publicly available medical and scientific information to help oncologists learn about the latest medical research and clinical trial information, personalize a treatment for a particular patient, and even increase the likelihood that a particular dose of medicine will be approved. To ensure quality and accurate decision-making, this AI is continually learning from scientists and from every interaction it has with a pharmaceutical company or a healthcare professional. Another intelligence system constitutes an example of a system that can use machine learning to understand associative implications of genes, proteins,

metabolites, and other biological entities, apart from mining images, charts, and graphs. Hence, this will not only help in finding new links—and therefore new opportunities in research and drug development—but also to avoid unnoticed off-target effects. All these efforts demonstrate the hunger for smarter drug development processes and a growing need for new regulatory methodologies to support fast-emerging trends of AI-driven drug development and patient care.

**6. Future Direction**

Many emerging trends and technologies offer potential opportunities to significantly enhance the efficiency of regulatory processes, although the investment in generating these approaches must be weighed against the benefits. One such opportunity lies in the use of meta-learning to model an expected time gain of proposed drugs or therapies when applying advanced predictive models to information available during regulatory procedures. In the future, AI might be used for making inferences on drug efficacy and safety across conditions by using advanced machine learning models and big data. It should be noted that adaptations or updates to the regulatory frameworks are required to ensure responsible introduction of the technologies as they evolve, since the AI technologies will evolve significantly. Finally, the drug regulatory silos approach will have to be revisited, and new interdisciplinary frameworks between regulatory authorities, other healthcare markets, societies, clinicians, and patients to encompass benefits, risks, and effective therapies must be considered.

Novel opportunities for using combined clinical trial, real-world data, and advanced machine learning for the assessment of drugs with the objective to fast-track approvals could reduce ethical concerns regarding the incorporation of benefits to patients at the time of approval while they are reversible. Yet, this approach may also pose additional ethical challenges, especially related to transparency, accountability, and challenges to decision-makers, such as how to meaningfully take on board multi-faceted predictions from advanced machine learning systems in a transparent way into the decision-making process. Keen attention will need to be paid to the extent to which emerging AI technologies need to stay within the regulatory bounds of static algorithms and explanations, and how they can be adapted to rapidly changing AI and machine learning technologies. Further collaboration between all stakeholders and cross-border

would be valuable in ensuring responsible and beneficial adoption of AI for drug approval process optimization.

**7. Conclusion**

In conclusion, the optimization of drug approval processes achieved through artificial intelligence could present a solution for many traditional challenges of regulatory science. Consequently, patients and society would finally benefit from this technology through more tailored treatment options, and scarce resources could be allocated in a more sustainable way. The two case studies demonstrate just how AI-driven changes in regulatory requirements can lead to specific advancements in healthcare: the precision drug development package using the example of the treatment of cystic fibrosis presented as a condensed way of evidence generation tailored to the specific degree of unmet need; the innovative personalized medicine platform that could be functional for this project for the coordinating joint HTA and regulatory advice. Our findings indicate the need for continuously addressing such topics via an informed discussion between policymakers, healthcare providers, the pharmaceutical industry, and other actors in the public health interest to ensure that regulatory advances can actually be translated into better healthcare for patients. This work identifies some next steps on our way to more research regarding the optimal timing of health technology assessments.

The two case studies presented suggest a number of promising and also specifically quantifiable effects when using AI-driven optimization of drug approval processes. Their empirical implementation in the new age of cohorts will be a challenging but unequivocally rewarding venture, and should therefore not have to be approached reactively, but rather proactively.