

Review

# Promoting Sustainability through Next-Generation Biologics Drug Development

Katharina Paulick  <sup>1,\*</sup> , Simon Seidel <sup>1</sup>, Christoph Lange <sup>1</sup>, Annina Kemmer <sup>1</sup>, Mariano Nicolas Cruz-Bournazou <sup>1,2</sup> , André Baier <sup>1</sup> and Daniel Haehn <sup>3</sup> 

<sup>1</sup> Chair of Bioprocess Engineering, Faculty III Process Sciences, Institute of Biotechnology, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany; simon.seidel@campus.tu-berlin.de (S.S.); christoph.lange@tu-berlin.de (C.L.); kemmer@tu-berlin.de (A.K.); n.cruz@datahow.ch (M.N.C.-B.); andrebaier@gmail.com (A.B.)

<sup>2</sup> Datahow AG, 8600 Zürich, Switzerland

<sup>3</sup> Department of Computer Science, University of Massachusetts Boston, Boston, MA 02125, USA; haehn@seas.harvard.edu

\* Correspondence: katharina.paulick@tu-berlin.de

**Abstract:** The fourth industrial revolution in 2011 aimed to transform the traditional manufacturing processes. As part of this revolution, disruptive innovations in drug development and data science approaches have the potential to optimize CMC (chemistry, manufacture, and control). The real-time simulation of processes using “digital twins” can maximize efficiency while improving sustainability. As part of this review, we investigate how the World Health Organization’s 17 sustainability goals can apply toward next-generation drug development. We analyze the state-of-the-art laboratory leadership, inclusive personnel recruiting, the latest therapy approaches, and intelligent process automation. We also outline how modern data science techniques and machine tools for CMC help to shorten drug development time, reduce failure rates, and minimize resource usage. Finally, we systematically analyze and compare existing approaches to our experiences with the high-throughput laboratory KIWI-biolab at the TU Berlin. We describe a sustainable business model that accelerates scientific innovations and supports global action toward a sustainable future.

**Keywords:** Industry 4.0; pharmaceutical industry; CMC; automation; sustainability; diversity; bioprocess development; digital twin; machine learning; intelligent systems



**Citation:** Paulick, K.; Seidel, S.; Lange, C.; Kemmer, A.; Cruz-Bournazou, M.N.; Baier, A.; Haehn, D. Promoting Sustainability through Next-Generation Biologics Drug Development. *Sustainability* **2022**, *14*, 4401. <https://doi.org/10.3390/su14084401>

Academic Editors: Baihua Li, Fei Chao and Amir Mosavi

Received: 30 December 2021

Accepted: 14 March 2022

Published: 7 April 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Objective and Significance of Work

The Human Genome Project, advances in personalized medicine, and high throughput drug development processes result in more therapeutic candidates to test and validate. Screening processes are challenging and meaningful, putting increasing pressure on pharmaceutical companies to efficiently and rapidly manage their CMC (chemistry, manufacturing, and control) functions, which accelerate the initial drug development into a therapy.

The mean cost of drug development is estimated to be USD 1336 million [1]. To withstand global competitiveness and fulfill stakeholders’ expectations, future drug development departments, i.e., CMC departments, are experiencing constant change. In fact, the World Health Organization (WHO) has published a guide for advancing health and sustainable development and sees digital transformation, lean leadership, and innovation as tools to promote health and well-being [2,3]. The presented systematic review draws a vision for future drug development departments [4]. The review includes an analysis of data science technologies, diversity studies, and a prediction for a business model that meets the UN (United Nations) sustainability goals (Figure 1).



**Figure 1.** The 17 UN sustainability goals demand to overcome hunger by global action. The goals must be implemented consistently in every project, company, and political decision. (Image from Welthungerhilfe.org [5,6]).

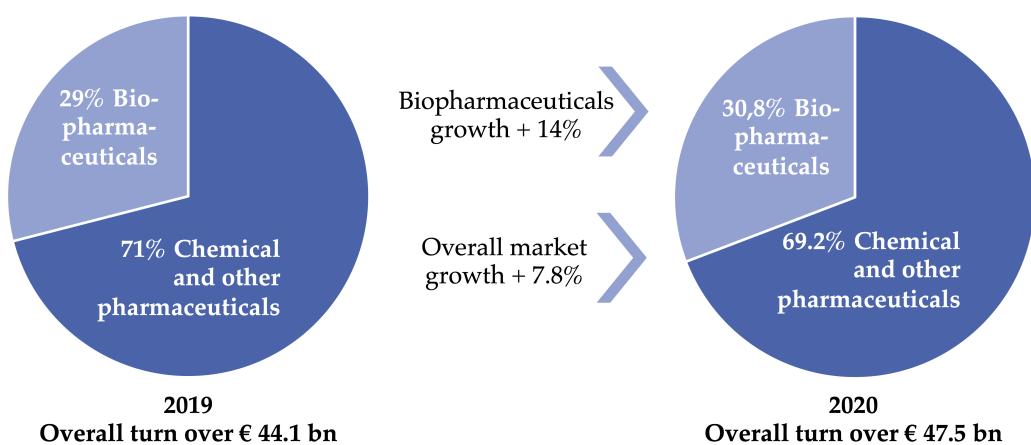
### 1.2. Structure of the Work

This review highlights the latest advancements in life sciences to build a vision for future drug development. To understand the interdisciplinary topic of the work, multiple definitions from the pharmaceutical, biotechnological, and mathematical areas are given. The pharmaceutical industry's current challenges and technological innovations are also shaping its future and will be supported by data science. As the scope of machine learning effects on sustainability is difficult to estimate, an overview of the status in CMC functions is followed by machine learning sustainability use cases, which are then building the basis for what could change in the future.

### 1.3. Economic Significance and Current Challenges in the Pharmaceutical Industry

The pharmaceutical market in Germany reached sales of EUR 44.1 billion in 2019 and EUR 47.5 billion in 2020, resulting in a market growth of 7.8%. The market share of biologicals is permanently increasing—lastly, with a gain of 14% from 2019 to 2020. Now, evolving biologicals make up 30.8% and EUR 14.6 billion of all new therapies in Germany. Biosimilars have a market share of 60% within the biologicals market in Germany [7]. The growth of biologicals comprises sales in immunology and oncology therapies, making up EUR 8.2 billion of the total sales (Figure 2). Pharmaceutical production is mainly based in Europe, the U. S. (United States), and Japan [8].

Regarding the new drug approvals, 56 new therapies were approved by the European Union in 2020, with 25 being biopharmaceuticals [7]. A recent study in the U.S. revealed that median investment for drug development per product is USD 985 million, including failed trials, clinical payments, and cost of capital [1]. The new therapies' shape describes the challenges drug development departments face. Among the newly approved drugs was the COVID-19 mRNA vaccine. BioNTech developed this first mRNA-based vaccine in collaboration with Pfizer. Moreover, the newly approved drugs comprise eight recombinant antibodies, more vaccines, and a recombinant protein. It is important to note that nine orphan drugs were also approved in 2020. The overall variety shows that many different processes had to be established in the past and will be necessary for the future to meet the demands of humankind.



**Figure 2.** Turnover and market growth of biological, chemical, and other pharmaceuticals, adapted from vfa Pharma (*Die forschenden Pharma-Unternehmen*) [7].

Boston Consulting Group and vfa bio (*Die forschenden Pharma-Unternehmen bio*) published the industry report “Medical Biotechnology in Germany 2021” in which they suggest exploiting the opportunities offered by digitalization and digitization to secure and expand medical biotechnology production sites in Germany [7]. Digitalization describes the transformation of processes using digital technologies; digitization is simply converting paper-based materials into a digital format. Additionally, administrative hurdles in clinical trials should be reduced, conditions for venture capital improved, and the networking of all stakeholders in the healthcare systems should be strengthened. In general, digital tools are expected to significantly reduce the average 13 years until a drug is approved [9].

Although two-thirds of German biotech companies plan to acquire new employees for digitization and digitalization, employees fear being replaced by machines. On the other hand, over 80% of biotech companies face challenges due to lack of qualified staff, too little budget, or too little employee acceptance [9].

Due to exceptional biotechnological advances, the complexity of processes and restrictions in drug development change. At the same time, with the development of new therapies, the complexity of clinical trials is also increasing, and documentation, reproducibility, supply chain management, and containment become a significant matter [10].

Overall, global challenges such as pandemics and democratic distribution of treatments exert pressure on the pharmaceutical industry. In addition to sustainable drug development, the industrial nations are responsible for supporting low- and middle-income countries with affordable, safe, and effective treatments. To this aim, the WHO started to support local production with a prequalification program for vaccines, which is progressively expanding by other medicines in 1996. The German Federal Ministry for Economic Cooperation and Development and all G8 countries (USA, Italy, Japan, Canada, Great Britain, France, Germany, Russia) support the program. Moreover, pharmaceutical companies offer self-cost products or licenses for local generics production. In many countries, even the cost price medicine is not affordable and if it is free, people do not have money to pay the doctor [11]. These countries need structural support for a public health system. It is essential that help is provided and, at the same time, the country's ideas should be implemented.

#### 1.4. Current Status of Digitalization

Automation and prediction are already playing a significant role in the pharmaceutical industry to speed up the drug discovery process and CMC. Companies operate semi-automated and accomplish automated screening processes using liquid handling stations. Additionally, predicting chemical reactions and the potential mode of action of new drugs is realized.

Machine learning software is used to build connections between diagnoses and therapies. Using medical databases, it is also possible to simulate the progress of a disease based on initial symptoms. Consequently, scientists gain an insight view of physiological processes in the body, signal transduction, and possible reactions to new therapies. Personalized medicine is just realized by linking information from large Biobanks containing information from clinical studies, treatments, and genetic profiles. Using clinical patterns, machine learning tools can find new biomarkers to diagnose and treat new diseases [12,13].

To date, different machine learning tools are being used in industry, i.e., to predict chemical reactions or protein folding [14,15]. Digitized literature research enables technical literature that summarizes news from one's discipline [9]. The availability of digital medical tools is gradually being democratized, as people in low-density medical care regions can also benefit from medical apps [12]. The rising number of mobile phone connectivity and mobile health initiatives in emerging countries support better healthcare globally [16–18]. Machine learning methods are also used for screening processes, process design, and optimization, and control of running processes [19,20]. Their use can shorten one of the most extended steps in drug development, which is chemical synthesis [15].

A large portion of CMC functions has started implementing automation. Pharmaceutical production plants are situated all over the world. The status of digitalization and use of machine learning depends on the location, size, and budget of CMC functions. Additionally, the quality of digitalization can be revealed by different criteria: laboratory integration and communication, automatic writing of reports, and number of fully digitalized processes. Thus, specific numbers of digitalized production plants cannot be provided, e.g., HSIEH et al. demonstrated using machine learning without digitalization. Moreover, there are small enterprises producing antibodies in Germany with low automation levels and fully automated laboratories. However, there is a general lack of a central software platform capable of mapping an entire laboratory, its devices, and processes.

The traditional CMC operating model includes product testing, defining its characteristics, proving its safety, and ensuring the manufacturing process keeps consistency between batches. This process is often managed paper-based and, as the required tests depend on the kind of drug, many lab experiments are performed manually. Furthermore, the preclinical testing includes method validation, which is then to be continued during clinical trials. Additional tests must be performed when moving into the clinical stage, and a proper scale-up strategy must be identified. Defining a proper scale-up strategy for small molecules that meets the validation requirements is already crucial in chemical processes but even more complicated when biologicals (biologically derived pharmaceuticals, Figure 3) are involved. At the same time, a fixed time point for harvesting has to be determined for validation. This may result in, e.g., mammalian cells being cultivated for longer than growth can be observed. To correctly estimate extensive scale conditions, it is necessary to predict large-scale parameters as precisely as possible. Once a drug candidate faces the clinical stage, no common understanding of how the data acquired during clinical studies are organized. A standard data format is indispensable to perform a standardized analysis and enable multiple CMC departments to share their findings. Drug development steps rarely occur in connected laboratories. Often, special departments are entrusted with individual tasks which hardly draw on the results of the previous working groups. So, there are many advanced labs worldwide working in a semi-automated manner. The introduction of a digital picture from the process is rare. This so-called digital twin is hard to compute and complicated to implement in a highly regulated field [21].

**Small molecules****Biologics****Novel treatment modalities**

**Figure 3.** Evolution of therapeutic medicinal agents from small molecules to novel treatment modalities, adapted from Cauchon et al., 2019 [22]. Remarkable advances in therapy options have been made, starting with small molecules, accelerating further with biologics such as therapeutic proteins to novel treatment modalities such as gene therapy. Many more classifications are made for validation [22–25].

### 1.5. State of Research: Industry 4.0 and Sustainability

The change toward a sustainable business model is expected from the press, politics, and society and promoted by various rankings [26–29]. Therefore, pharmaceutical development of the future should operate according to the sustainable development goals (Figure 1, [6]). We hope that the chemical industry continues to take significant steps toward climate neutrality to fulfill regulations and for a better tomorrow. At the same time, disruptive innovations of therapeutic and digital technologies are shaping the future of CMC. The environmental impact of current business practices, modeling of future sustainable development, and its impact on the economy were discussed in [30]. Smart manufacturing and the integration of machine learning, including molecular modeling in biotechnology and its benefits for the environment, are discussed in recent reviews [31,32]. In particular, the authors of [31] point out that combining neural networks with mechanistic models enables them to work with smaller amounts of training data. In addition, the use of these models can improve the operational procedures during fermentation. However, there is no unified approach that suits all use cases.

The modeling of product formation was also shown to accelerate further ecological sustainability [33]. The benefit of artificial intelligence (AI) in general and in all areas of sustainability—economic, environmental, and social—was also predicted recently to an overall positive impact of 79% [34].

However, to the best of our knowledge, the current literature does not include specific use cases that combine innovation, machine learning, and sustainability.

## 2. Materials and Methods

### 2.1. Concept

The systematic review is based on the PRISMA reporting standard [4]. Basic definitions of the regulatory field in drug development and sustainability and data science terms are provided. Further, an overview of status, restrictions, and advancement in the pharmaceutical industry is linked with the level of the newest research tools. Currently, machine learning and sustainability are playing a central role. It is shown that data science and sustainability can improve productivity and depend on each other. We describe the current state and challenges within the pharmaceutical industry and present data science tools commonly used in academic research. We also report milestones and limitations of our industry implementation. The authors included specific studies and reviews about data science tools in bioprocess engineering, diversity research, and descriptive publications for a sustainable business model. As there are, especially in sustainability, multiple publications, just the latest could be included to highlight the importance of data science and diversity in future business models. The authors specialize in biotechnology, machine learning, chemical engineering, and mathematics. The assessment of the analyzed data

is mainly qualitative; however, we include the calculation of the relative error to validate the models.

## 2.2. Theoretical Basics and Definitions

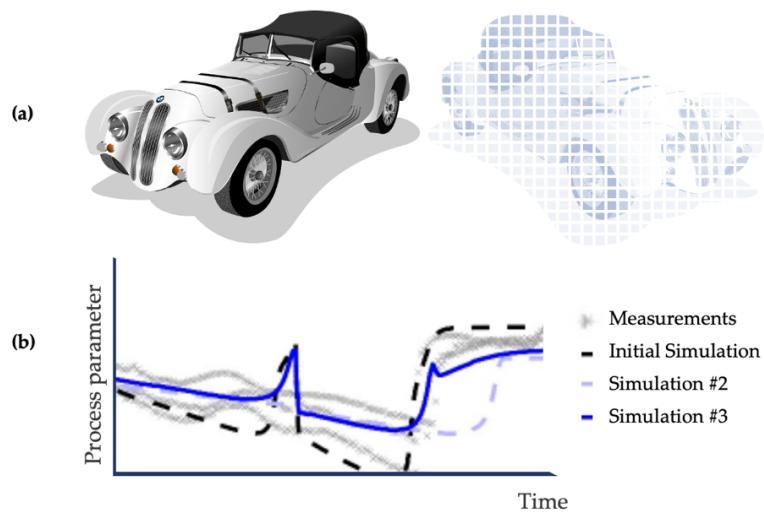
**CMC—drug development in the pharmaceutical industry.** To ensure standardized manufacturing of pharmaceuticals, drug characteristics must be defined appropriately, and the production process should be controlled tightly to guarantee consistent product quality. Following drug discovery, all phases of drug development are defined as CMC [35]. CMC comprises the preclinical drug development where proper analytical methods, stability testing, and physicochemical properties are described. Further characterization of the drug and its process design is required in the clinical stages. The process design must ensure the end product will always meet the specifications needed in the quality guidelines. After successfully passing clinical trials, the scale-up process must also ensure consistency between batches.

Depending on the future market of the drug, the regulatory authorities are setting specifications for the drug process, the U.S. Food and Drug Administration (FDA) and the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use in the EU (European Union), Japan, and the U.S. (United States) [36]. Further quality guidelines are defined by the European Medicines Agency (EMA) and the German Federal Institute for Drugs and Medicinal Devices [37]. The drug development process is highly controlled and depends on the kind of drug applied in humans.

**Personalized medicine.** There is no standard definition for personalized medicine, but it aims to solve the challenge of treating a disease differently in different patients [38,39]. It was shown, and it is still under investigation, that these differences can be linked to genetic and molecular profiles. So, genetic variations between patients can indicate different responses to drugs.

Personalized medicine becomes more critical every year using advanced knowledge from diseases and therapies. It is interdisciplinary to find the best appropriate treatment for individuals and improve medical care by a higher success ratio. On the other hand, more screening processes in advance will be necessary. In the end, medical and molecular knowledge is used to model and adapt the right therapeutic strategy.

**Digital twin.** The new catchword “digital twin” is the digital representation of a particular process (Figure 4). In general, twinning is the systematic comparison between actual and computer-generated methods [40]. The frequency of synchronization depends on the application requirements, i.e., a slower process such as the growth of mammalian cells would require less frequent measurements for synchronization than a chemical process. In truth, a digital twin is not achievable; the process and its measurements would have to be perfectly the same and the produced data indistinguishable from calculated data. The twin is used for simulation, creating data similar to the actual process. Typically, in chemical manufacturing, the process can be mimicked by profiting from knowledge of essential chemical reactions. Naturally, mechanistic models become more complex the more parameters are involved. The outcome is deterministic—using mechanistic models, the course of a process can be estimated. A perfect twin of the production process and accurate surveillance via sensors is necessary for accurate prediction. At this level, the proper device integration is obligatory.



**Figure 4.** A digital twin is the model of a specific process or technology which is easier to perform with static technologies such as a house or a car. Dynamic models are expected to predict the process flow when it comes to process modeling. (a) A digital twin of a device such as a car is easier to calculate because it is physical in nature. (b) A digital representation of a process estimates its performance. Here, the simulation containing the most data, Simulation #3, is closer to the measurements than earlier simulations. Images adapted from [41,42].

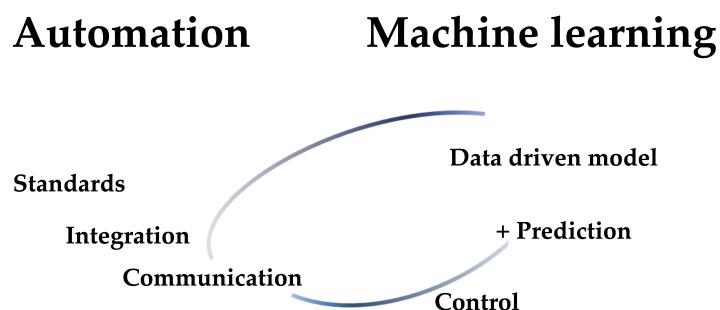
**Design of experiments (DoE).** Chemical and biological processes depend on multiple variables, i.e., induction time, feed, metabolite concentrations. The conventional one-variable-at-a-time (OVAT) method is laborious and cannot represent the optimum of all input variables. All variables but one are kept constant during the experiment [43]. DoE originated from statistical planning and analysis of experiments to find the optimal condition in a minimum number of experiments. It is broadly used in industrial process design to improve efficiency and obtain a detailed map of the dependencies of multiple process parameters. The roadmap of DoE encompasses randomization, statistical replication, blocking, orthogonality (statistical independent), and factorial (with more than one factor) experiments. Model-based DoE further improves the standardization and automation of clinical trial conduction in the pharmaceutical sector [32].

**Small molecule, biologicals, and gene therapy.** There are multiple evolving therapies and drugs in the market with game-changing potential as gene therapies, immune-oncology, or the just-validated synthetic mRNA (messenger ribonucleic acid) vaccine. On the other hand, these remarkable medicinal advances require more resources than developing a small molecule. Referring to the EMA, small molecules refer to chemical substances defined by a single molecular structure that is not a protein or nucleic acid [37]. Chemical substances are generally considered “small” molecules with associated salts, solvates, or ions and may be described using a single definitive or representative structure. Biologicals are drugs, therapies, and vaccines produced using genetically engineered hosts. Biologicals include various drug products produced from living organisms derived from humans, animals, or microorganisms using biotechnology. Vaccines and recombinant proteins and synthetic mRNA also belong to the group of biologicals.

According to the EMA, gene therapies belong to the cell therapy group. Genetic engineering delivers nucleic acids to cells to alter their genetic information. Naturally, there are many more types of therapy, and this is a very superficial description aimed at distinguishing between current treatments and products. This report focuses on CMC’s relevant activities in the drug development process.

**Insights into machine learning and artificial intelligence.** To reveal the power of an intelligent lab, the benchmarks of machine learning techniques have to be pointed out. Mathematical concepts and language can describe processes such as chemical or biological production. A model can help predict process behavior and study the effects of

different parameters. Furthermore, experimental data can be expected to achieve process optimization (Figure 5).



**Figure 5.** A lab corresponding to Industry 4.0: A fully automated and digitalized lab is just one part of the answer toward Industry 4.0. Intelligent process optimization in all its facets is required to orchestrate more important CMC parameters, e.g., higher patient-centric drug considerations and health care demands place significant pressure on the industry, and bureaucratic hurdles are slowing innovation.

To develop an efficient production process, process behavior can be predicted. The re-design in real time and minimization of experiments help to reduce the number of wet lab processes drastically [41,44–46]. There are different ways of predicting the behavior in processes, i.e., using mechanistic or empirical modeling. Briefly, in mechanistic modeling, scientists use known differential equations from the production process, i.e., reaction rate, resulting in a deterministic outcome. In contrast, empirical models gain their structure from historic data and are applied to newly captured data. However, data capture for empirical modeling can be very laborious in biotechnological processes and requires multiple experiments. In addition, biological systems are rarely behaving in the exact same manner. Thus, constant adaption during the process is necessary.

Taking earlier process behavior and chemical and biological rules into account can speed up process development. Mathematical concepts and language can describe processes such as chemical or biological production. A model can help predict process behavior and study the effects of multiple influencing variables (such as temperature and glucose concentration). Here, naturally, model-based DoE also plays a role. Machine learning can predict the flow of the process and adapt influencing variables for optimal production. The re-design of a running process by machine learning improves the efficiency and the feasibility of the process.

Machine learning can predict the efficacy of substances before they are produced [19,47,48]. That saves superfluous tests, resources, and time. To foresee the behavior of drug candidates within the body or in combination with other drugs, molecular modeling and computational biology in drug design is already used.

The scaling-up strategy can be predicted to develop an efficient production process in CMC. Here, process modeling is discussed to perform intelligent process design with the Internet of Things (IoT), resulting in a self-driving laboratory. There are different ways of predicting behavior in processes.

Captured data enable the computing of empirical models. Sufficient empirical data could help estimate the progress of a current process with the same operating conditions. However, data capture for empirical modeling using a specifically designed experiment could take a year in biotechnological processes. A model can consist of artificial neuronal networks (ANN) and a DoE approach. Although neural networks can efficiently model complex data, every new process will obtain a different network. Consequently, it is difficult to support using a neural network in a highly regulated environment.

Hybrid modeling can be used to reduce weaknesses from each of these modeling techniques and compensate for unknown mechanisms [49–51]. Here, a deterministic

model is complemented by other machine learning techniques such as ANN. Continuous bioprocesses consist of hundreds of components.

Recurrent neural networks (RNN) are of great use when it comes to modeling the respective trajectories as long as all measurements are taken in the same time interval. Recently, neural networks combined with the integration of ordinary differential equations, and neural ordinary differential equations (neural ODEs) were used in biotechnological applications [52,53]. These models consist of a neural network that defines the dynamics of differential equations. They outperform recurrent neural networks when predicting time-series with not equidistant samples, as they do not need discrete information but can work with continuously defined dynamics [54]. In recent years, many data have been collected in the chemical and biotechnology industries, so there is enough data to realize a shift from mechanistic to hybrid models and neural ODEs to simulate and control processes.

Real-time process behavior can be characterized, predicted, and process control can be adapted following prediction. The certainty of prediction can be displayed by the coefficient of variation, calculated by dividing the sample standard deviation by the mean value and the relative error, which describes the normalized deviation from the measured value.

### 2.3. Validating Machine Learning Models

As introduced above, several model groups exist, including statistical or empirical, chemometric, mechanistic, and hybrid models. While models can help gain more knowledge about a system, they involve the subjective choice of model design. Oversimplifying this design results in a model with limited flexibility, which does not represent the data adequately. This phenomenon is called underfitting. On the other hand, overfitting occurs when the design involves too many model parameters, and a model starts to learn patterns that originate from noise [55]. Such a model performs excellently on training data but fails when presented with an unseen dataset. One can identify these issues by following appropriate validation methods such as using an independent test dataset [56].

Smiatek, Jung, and Bluhmki thus stress the importance of consistent terminology and appropriate model validation strategies [57]. Comparing different model types against each other while being loose on the validation methodology may lead to overestimating the models' capabilities [58,59]. For the use case of supervised learning in biology, Walsh et al. and Chicco proposed recommendations for machine learning best practices [56,58]. The goal is to enable a critical assessment, comparisons between different studies, and reproducibility. An overview of validation methods for the different model types in the context of biopharmaceutical production is given in [60]. This validation setting is essential for comparing different pre-processing steps or model architectures against each other. In particular, one should challenge one's model by comparing its performance to a more straightforward model type to ensure that the increase in model complexity comes with an absolute benefit [58]. Mowbray et al. describe the use of ML methods and their impact within biochemical engineering from the last two decades [31]. Establishing these model validation practices will help identify the areas within biology that benefit from machine learning.

Sometimes, a model works just fine on some parameter, as it strongly depends on the quality of the data. Model parameters, while relevant, might not be identifiable with the present data [61], e.g., the parameters of a model related to substrate-limited growth of microorganisms cannot be identified during a batch process. A guide on the development and statistical assessment of (mechanistic) models in this context is given in [62,63].

### 2.4. Methodologies

Methodology to assess the impact of new technologies and business methodologies on the SDGs. The authors perform a qualitative evaluation of reaching the SDGs by practicing a sustainable business model. The 17 SDGs can be divided into economic, ecological, and social goals [34]. A general impact, such as impact on climate, not on specific goals such as SDG number 14, life below water, could be evaluated. Thus, the authors decided to

reduce the number of goals to those that sustainable pharmaceutical production directly influences. Further, targets were reduced to a more general evaluation on sustainability, i.e., climate action in general and not the impact on life below water or life on land.

**TINS-D Analysis.** The constellation analysis of technology, individuals, nature, society, and democracy (TINS-D) allows us to analyze the reciprocal relations of these coordinates about innovations, societal developments, etc. [64]. The coordinates of nature and technology and individuals and society are placed at opposite sides, forming a cross with democracy placed in the middle. This placement allows us to analyze especially the dialectic relations between the two polar coordinates, how they support, block, affect, and constitute each other. It also allows us to analyze how the two polar coordinates affect the other two coordinates. Democracy is at the center to call for analysis from this viewpoint of the entire reciprocal relations of the other four coordinates and their democratization. Overall, the TINS-D analysis allows a more general view on sustainability than the 17 SDGs.

### 3. Results

#### 3.1. Challenges in the Pharmaceutical Industry Require Innovation and Data Science

Drug discovery, i.e., finding a promising therapy or vaccination, is already lengthy and time-consuming. Different strategies are tested, millions of molecules are screened. CMC comes into play if there is a promising therapy and another long journey of characterization, validation, and process development starts. Would a wholly digitalized and intelligent pharmaceutical plant be far more efficient and comprehensible, resulting in a process that can be easily translated to the different plants in the world?

Scientific developments in the pharmaceutical sector are currently boosting the number and quality of potential new therapies. Significant advances are, e.g., the Human Genome Project, the ENCODE project, a humanized mouse model, nanobodies, considerable progress in diagnostics, and drug delivery techniques [19,20,65]. Accordingly, we are moving into an era of high precision medicine combining diagnostics with predictive and personalized medicine applications [22]. Mapping and calculating these large amounts of data is only possible by digitalization and the use of machine learning technologies.

Artificial neural networks are used for developing green pharmaceutical methods. Once a new compound is found, CMC is responsible for characterizing it qualitatively and deciding how to measure it quantitatively. The number of experiments to determine the physicochemical properties of compounds and behavior under different conditions can be reduced by using multivariate statistic tools [66]. This is important, as drugs need to be mixed with other components during formulation to become a medicinal product, and their behavior in the human body has to be predicted as precisely as possible. Models from the quantitative structure–activity relationship (QSAR) family are used to estimate the environmental impact of new drugs [67]. As using these models in the regulated environment will become necessary, Tong and colleagues characterized the prediction accuracy of QSAR models. Moreover, in chemical applications, ANN and genetic algorithms are used to identify patterns in data, single and combined [66]. The decision for the perfect solvents depends on their physicochemical, toxicological, and hazard characteristics; there are green alternatives. Using correspondence analysis and PCA, the number of wet lab measurements to find the optimal solvent can be decreased drastically [66]. Korany and colleagues established a simultaneous determination of the concentration of two chemicals in tablets using zero-crossing derivative spectrophotometry and a spectrophotometric ANN chemometric calibration model. This method saves solvents for separation methods, which would have been performed in standard procedures. The ANN method was superior to the derivative technique because ANN can determine both drugs under nonlinear experimental conditions [68].

Therapeutic medicinal agents evolved from simple small molecules produced chemically or therapeutic proteins extracted from natural resources to complex treatments such as gene or cell therapies. Their production becomes more and more complex. An acceleration of the development process by companies is required to survive on the market.

This encompasses all product development steps from the initial experiments through the individual development and production stages to the final purification stage. However, such a concept of consistent process improvement can only be implemented in a comprehensively networked laboratory. To handle the complexity and number of diseases and test many more therapy candidates, automation is necessary and improving the situation for validation. However, in the end, the rising experimental effort can only be solved by intelligent process design, beginning with model-based DoE and subsequently using data science technologies to monitor and control processes.

The evolving personalized, patient-specific medicine consisted of niche products and low-turnover drugs until recently. Advanced Therapy Medicinal Products, ATMPs, are treatments including recombinant nucleic acids, engineered cells, or tissues. From the manufacturing perspective, a huge change for their production is necessary and can be managed by flexible and redundant process management in parallel with modeling and simulation [69,70].

More remarkable techniques have been developed to identify the best therapy option without touching the patient in the field of personalized medicine. The challenge of personalized medicine is synchronizing patient and treatment (see Section 2.2). The increased understanding of genomes and a better understanding of rare and hard-to-treat diseases makes personalized medicine necessary, and first therapies are already showing significant higher success. In 2018, the first CAR T-cell (chimeric antigen receptor T-cell) therapy was approved in the European Union and in the U.S. in 2017. This gene therapy drug product comprises engineered T-cells from the patient. Another T-cell-based therapy is the allogeneic cytotoxic T-cell immunotherapy. A donor delivers healthy T-cells that are activated and enriched if they recognize antigens expressed by cancerous cells. The perfect T-cell line for the patient can be identified using data science. The technology must be performed and optimized for every single patient in a single bioreactor. To this aim, the Cocoon™ technology, a highly scalable method for T-cell therapy, obtained its patent for research use in 2020 and is already used for medical treatment [71–75]. Other evolving cell therapies like adoptive T-cell transfers, from which more than 500 are already in clinical trials, will benefit from the further development of smaller high throughput systems. High precision medicine combines diagnostic tools to treat a specific disease of a single patient. In oncology, cell samples from the patient can be taken and used for diagnostic purposes such as single-cell sequencing and multiplex bioimaging. Drugs or distinct gene therapies can be tested using cultures of the patient's cells. Finally, automation and machine learning combine all these tools, and their computation results in the optimal treatment [22,76].

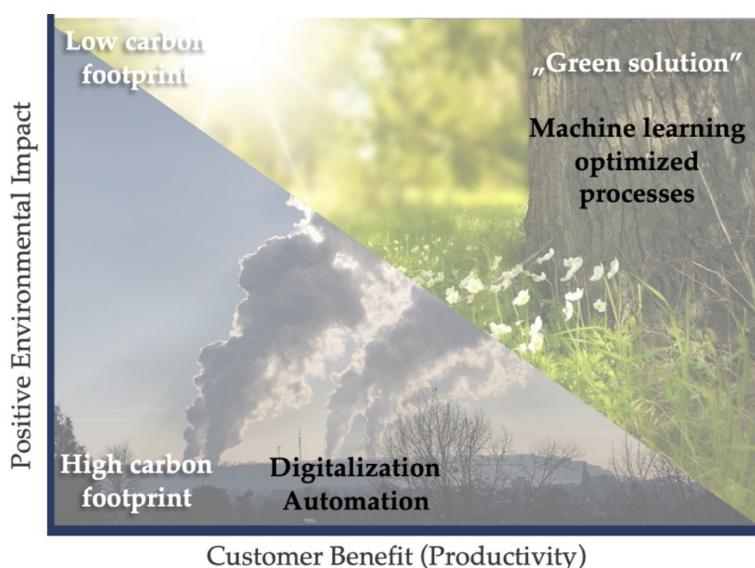
Evolving biologicals comprise proteins, gene therapy products, and antibodies. They are produced by hosts such as bacteria, yeast, mammalian cell lines, or even primary cells from the patient. Small-scale production is performed in minibioreactor systems. Scaling up for large-scale production is a scientific area embracing process simulation and mimicking extensive processes in small production systems. Naturally, a digital twin comes into play for process simulation to generate a specific computer-generated process, ideally combined using a small production system like minibioreactors for bacterial, fungal, and mammalian cells.

Compared to broadly used shake flasks, minibioreactors can suitably simulate process parameters such as oxygen supply, feed rate, and glucose concentration which can then be transferred to larger systems [77]. In combination with liquid handling stations, advanced control of cultivation is feasible. The potential of cost-effective minibioreactor systems is becoming more popular, although a successive approach is often inevitable. In both scenarios, the attempt to combine minibioreactor systems with twinning is promising [41,44,45].

Advanced process analytical technologies (PAT) are used further to facilitate chemical and biological process optimization in bioreactors and synchronize the twin with the actual process. The FDA published PAT guidelines to improve upstream processes, as they already have influence on product quality [78]. The rising technologies for at-line, in-line, and on-line analysis, i.e., on-line oxygen, pH, and glucose measurement, using physical-chemical sensors,

are essential tools for automated process control and its transfer to higher and more minor scales. RAMAN spectroscopy, as well as the implementation of fiber-optical photonic devices for the analytical characterization of chemical or biotechnological processes, are promising wet lab technologies to estimate the real-time state of the process [79,80]. Additionally, soft sensors improve real-time monitoring of bioprocesses [81].

The demand for drug development processes that meet the environmental sustainability goals is rising. To realize a bioeconomy, a fast and safe path to highly efficient production processes is needed. Green alternatives exist for many ecologically harmful chemical processes. However, these are not implemented because the research and development costs do not pay off. Process optimization by model-based DoE, machine learning-based process design, and real-time control can reduce development costs to a minimum. The difference an intelligently designed process can make in terms of productivity and carbon footprint can be visualized in an Eco-Care Matrix (Figure 6): intelligent and straightforward automation stand far apart. Automation enables the industry to realize more processes than manually possible, providing much data, and can thus be found in the lower middle part (Figure 6). Although a high throughput is reached, the actual productivity is relatively low as the vast number of experiments still need to be directed by humans. Clearly stated, subsequently, the success of an experiment depends on the number of attempts. As machine learning makes processes more effective and saves work and resources, you can find it in the upper right corner, where nature and productivity are profiting from the technology. Thus, automation can speed up the number of operations at the expense of the environment. The combination with machine learning will reveal an ecological and efficient design.



**Figure 6.** Eco-Care Matrix (ECM): The Eco-Care Matrix must be applied in the early stages of the product life cycle. Global regulations and policy also need to be considered. The ECM shows the environmental and economic impact of the new product or process. Products placed in the lower left corner need to be adapted further to move into the upper right corner demonstrating a green solution, implicating customer and environment profit comparably. Adapted from [82].

### 3.2. Machine Learning and Laboratory Automation in Line with the Idea of Industry 4.0 Promotes Process Efficiency

To outline a vision and possibilities of the pharmaceutical industry of the future, we present the latest advances in machine learning and insights from more sustainable business practices. However, putting the efficiency in numbers is not trivial. Ultimately, modeling supports and promotes both sets of methods. On the one hand, it can predict, plan, and control processes to increase efficiency and sustainable business practices. On the other, it can forecast their benefits for society and the environment. Data science is the correct

term for describing what most headlines want to express when using “digitization” and “digitalization”. The apparent advantages of digitalization are easy to list—facilitated data collection and data analysis—creating the basis for modeling, applying DoE, an improved control strategy, and data sharing. Additionally, the speedup of product testing, characterization, quality control, and validation of analytical methods is optimized. Furthermore, the scale-up strategy can be calculated in advance, and digital twins can control the process.

An initial step to lab digitalization is device integration. Proper integration enables real-time data capture and remote control of lab devices. As the machines use different drivers, have other data, and have machine-to-machine (M2M) communication standards, the proper integration of a new lab device can take up to six months. In the end, lab automation can be reached by a holistic device control and data management approach. Suitable device integration results in connection, integration of devices, interface with other devices and labs, and data transfer to the storage unit.

The fourth industrial revolution was already describing what the objectives of future enterprises were: automation, Internet of Things (IoT), and M2M communication are resulting in intelligent control of devices without human intervention [83,84]. The self-driving lab using data science goes one step further: software can plan and execute the resulting crucial experiments [85].

**Increasing the efficiency of bacterial scale-up/scale-down processes through modeling bioprocesses.** The TU Berlin (Technische Universität Berlin) and, in particular, the department for bioprocess engineering has played a vital role in shaping the field of bioprocess development in recent years [41,44–46,86–88]. A high throughput laboratory was established, and extensive automation, process control, and soft- and hardware integration was developed. Thus, it represents an ideal laboratory for bioprocess optimization.

In 2019, the setup of the liquid handling stations and automated screening of the bacterium *E. coli* (*Escherichia coli*) using minibioreactors in combination with online monitoring and control was described in detail by Haby et al. [77]. Combined with a second liquid handling station, at-line analysis was performed. A fully connected laboratory enabled central real-time data storage and model-based process control. In the end, the authors show that using a model-based framework, the digital twins, enormous potentials can be exploited in developing new products [44].

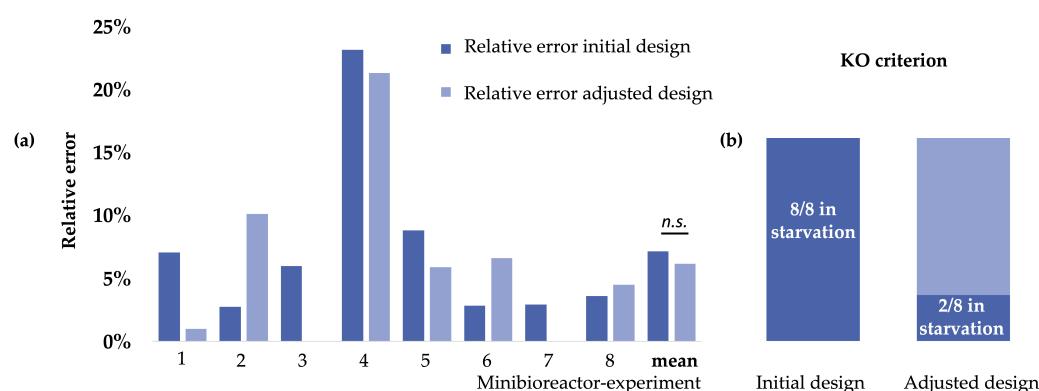
To describe the progress in the wet labs using liquid handling stations, we are analyzing two publications from the high throughput laboratory performing screening experiments with strains from the bacterium *E. coli*. We will point out the feasibility of the case studies and compare the errors of the predictions compared to the predictions made beforehand. Overall, the objective of the experiments is to define the optimal feeding strategy or the best performing clone by performing just one run in the liquid handling station. The liquid handling station can run up to 48 minibioreactor experiments in parallel. Their unique capability is to mimic the conditions of large fermenters, which are usually used to produce the final biological product.

In 2017, Cruz et al. performed a case study and determined the optimal growth conditions for one clone of *E. coli* using eight minibioreactors [45]. During the experiment, 23 parameters such as glucose consumption rate, growth rate, etc., were repeatedly predicted while the experiment was running. The prediction was based on Monod-type differential equations. Naturally, the more information the digital twin obtains, the more precise the running experiment could be predicted. The authors compared the deviation with the measured values by determining the coefficient of variation. Compared to an experiment using the initial computed conditions, the average coefficient of variation was 50 times lower. Thus, the paper showed that repeated re-computing is needed to control multiple biotechnological experiments in parallel.

In 2020, Hans et al. managed to screen eight triplicates of *E. coli* clones in parallel [41]. The objective was to find the fastest-growing strain with oscillating glucose concentrations, as would be found at an industrial scale. They adapted the optimal conditions (feeding profile, pH, oxygen supply) during the experiment to reveal the best performing strain.

Just the process parameters dissolved oxygen tension (DOT) and pH could be measured in real time. Together with at-line analyzed parameters such as cell dry weight and glucose and acetate concentrations, measurements were written with more than one hour delay into the database, critical parameters were computed, and the experimental conditions were changed and updated. The experiment lasted eight hours. The glucose consumption of a batch cultured is one crucial parameter to be predicted as this determines the start of the feeding and the cultivation time (Figure 7). As the glucose measurement cannot be performed in real time, scientists estimate the consumption using mechanistic models.

During the cultivation of *E. coli*, it is favorable that the cultivation is glucose-limited, meaning that there is no or only temporary accumulation of glucose. Assume *E. coli* cells are subjected to glucose amounts that exceed their cellular uptake capacities. In that case, the cells enter the overflow metabolism and produce chemical compounds that are harmful to themselves [44,45]. Both overfeeding and starvation must be prevented, and the start of the feed and its rate must be calculated. The course of the experiments was once predicted initially and continuously adapted during the cultivation. To show the importance of model calibration during the experiments, we will compare the initial model's errors calculated with the adjusted model's mistakes made to the observed value. The adaptive and specific experimental design enabled an experiment where the start time of feed and feed rate were calculated correctly in all experiments. The relative error calculation did not reveal a significant difference. Compared to the study from Cruz et al., an earlier point in time with much fewer data available was analyzed [45].



**Figure 7.** Prediction performance of eight optimized minibioreactor experiments. Error analysis of initial model, adjusted model (parameter estimation after 1.88 h), and observed/measured time points for glucose depletion. (a) No significant improvement of prediction time after 1.88 h, paired T-TEST,  $p$ -value = 0.578. (b) The re-calibration of the model enabled the rescue of 6 from 8 experiments by predicting an earlier time-point of glucose depletion compared to the initial model. Presented and analyzed data from Table 1 in [41].

Although the change of the relative error was not significant in the presented study, all *E. coli* cultures would have starved too long and would have been unusable if the glucose consumption had not been re-designed. Moreover, the calculations prevented overfeeding. The analyses of the re-designed experiment positioned the time point of total glucose consumption just in two experiments too late, leading to starvation (clones 4 and 8, Table 1). In two cases, the exact time point was estimated. In one case, the depletion was estimated only one minute too late, and in the remaining three cases, the point in time was calculated too early so that the cultivation was not endangered [41]. In one triplicate, the batch end was predicted 22.8 min too late. In that triplicate, the variance of glucose consumption was very high. The batch end of a single minibioreactor was very late and altered the overall prediction. The authors inform that without the leading sample the prediction was closer to the actual batch end.

**Table 1.** Overview of time-point of glucose depletion from eight minibioreactor experiments. Prediction overview of initial and adjusted experimental design. The relative error of predicted parameters vs. observed parameters was calculated from columns 2–4. Values of glucose prediction and depletion are derived from Table 2 in [41].

Strain/Glucose Consumption (min)	Initial Design	Adjusted Design	Observed Glucose Depletion	Deviation Initial	Deviation Adjusted	Relative Error Intitial Design	Relative Error Adjusted Design
<i>E. coli</i> . clone 1	106	100	99	-7	-1	7.07%	1.01%
<i>E. coli</i> clone 2	112	98	109	-3	11	2.75%	10.09%
<i>E. coli</i> clone 3	106	100	100	-6	0	6.00%	0.00%
<i>E. coli</i> clone 4	133	131	108	-25	-23	23.15%	21.30%
<i>E. coli</i> clone 5	111	96	102	-9	6	8.82%	5.88%
<i>E. coli</i> clone 6	109	99	106	-3	7	2.83%	6.60%
<i>E. coli</i> clone 7	106	103	103	-3	0	2.91%	0.00%
<i>E. coli</i> clone 8	115	116	111	-4	-5	3.60%	4.50%
<b>Mean relative error (Glucose consumption, low number desired)</b>						7.14%	6.17%
<b>Amount of cultures in starvation (low number desired)</b>						all	2
<b>Paired T-TEST of relative errors initial vs. adjusted design</b>						<i>p</i> -value = 0.578	

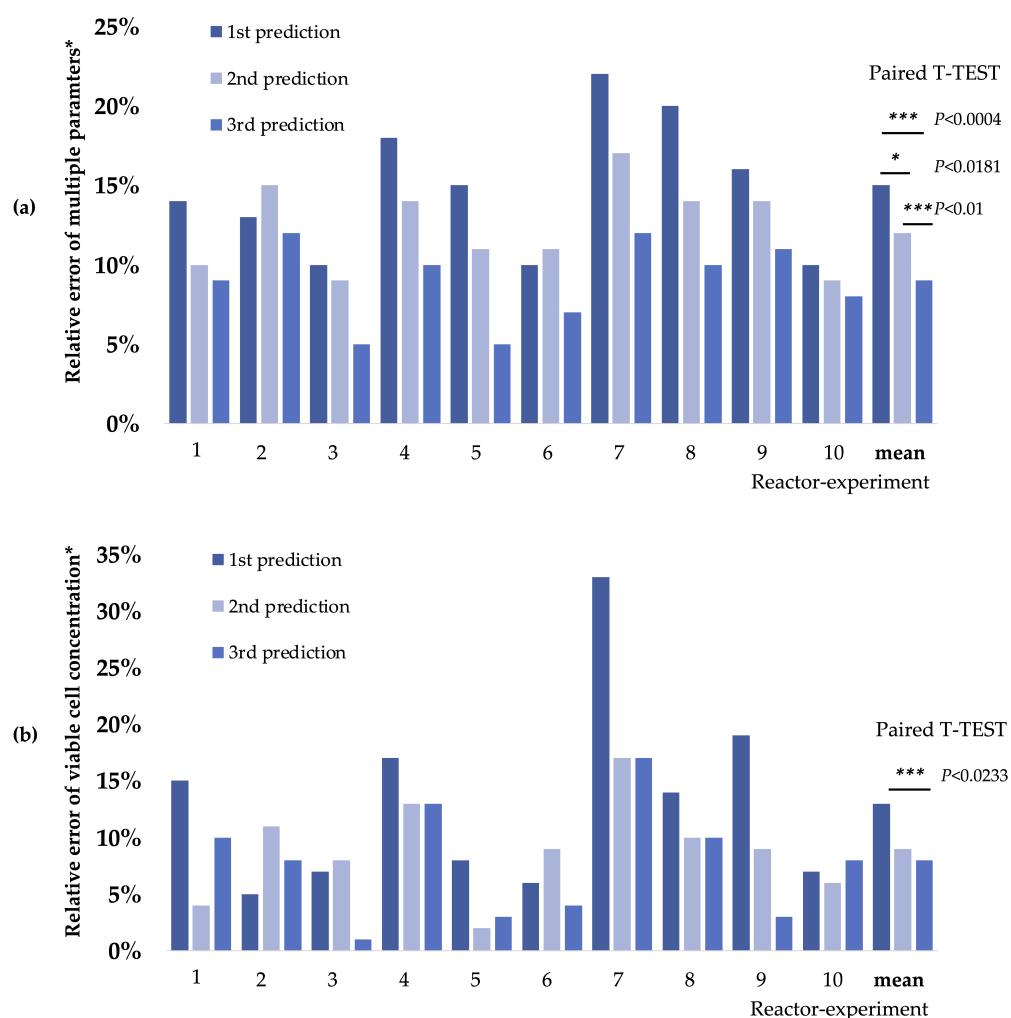
Compared to the case study from the same group, published in 2017, three times more clones could be screened automatically, resulting in the first model-based process of 24 fed-batch minibioreactor cultivations. Although the error of the re-designed calculation did not differ significantly from the initial model, the re-calibration improved six out of eight experiments. In this context, it is essential to mention that the initial model was developed further compared to previous versions [41,45].

**Increasing the efficiency of mammalian cell cultivations by modeling.** Hernández Rodríguez et al. performed, predicted, and redesigned an industrial cell culture using mammalian cells. Chinese hamster ovary cells were used to produce monoclonal antibodies, a biological drug with a market value of USD 114.43 billion in 2021 and applied to the treatment of cancer, autoimmune diseases, and multiple sclerosis [89].

Compared to bacterial fermentation, the cultivation and control of mammalian cells is more complex as cells are transferred to different, increasingly larger bioreactors, the so-called seed trains. The time point for transfer must be determined individually. Additionally, many more parameters such as viability, cell density, glucose, lactate, and ammonia concentration affect cultivations and need to be predicted. On the other hand, mammalian cells are growing slower, and decision making is not so time-critical.

The group also used mechanistic modeling based on Monod-kinetics for their predictions and extended them by Bayesian updating. Bayesian parameter estimation is based on previous knowledge, i.e., previous experiments, and is one method of empirical machine learning. To this end, eight simple experiments were performed to improve the prediction of model parameters. These predictions had an average coefficient of variation of 30% (Table 3, [89]).

As a result, prediction intervals containing a set of possible true values and their relative errors were calculated. The aim was to have small prediction intervals with high precision. Due to the variability of the cultivations, not all parameters and seed trains could be predicted with a high probability. Still, the accuracy of cell density prediction had a relative error reduced from 13% in the first to 8% in the last calculations in a 90% prediction interval. The average relative error of the other parameters was reduced from 15% to 9% (Figure 8). The authors demonstrated that in mammalian cell culture, without automation and relatively simple shake flask experiments, critical time points of experiments can be determined using machine learning. The model improves significantly the more information is available. The relative error of the prediction compared to the measured value was finally reduced to 8%.



**Figure 8.** Error analysis of parameter estimations of ten seed train predictions. Training performed with data from three consecutive bioreactor scales up to 2160 L. Bayesian parameter estimation was performed in combination with mechanistical modeling. Same mechanistical model used in all scales. Analyzed parameter predictions: (a) viable and total cell concentration; glucose, glutamine, lactate, ammonia concentration and (b) viable cell concentration. Paired T-TEST revealed a  $p$ -value  $<0.05$ , so the mean relative error decreased significantly from the first to third seed train. Analyzed data are derived from Table 5 from [89].

### 3.3. Economic Efficiency through Diversity and Equality

To survive in the fast-changing economy, companies must innovate continuously. This means adapting to new markets and inventing new processes, products, and technologies. A survey of 18,200 professionals in 27 countries and 150 C-level executives from 8 countries found that the willingness to innovate is 11 times higher in companies with the strongest focus on equality. Thereby, it is essential not to hire people with multiple backgrounds and genders but also to give them a genuine chance of equal careers. Innovative mindsets are six times higher in environments where cultures are almost identical compared to least equal ones [90,91].

Hsieh and colleagues analyzed the economic effect of changes in diversity, especially considering the gender and background of staff in highly skilled occupations, and published their findings in 2019 [92]. They examined the proportion of white men, women, and people from diverse backgrounds in the U.S. in different occupations using census data and American Community Survey data to define parameters that describe the impact on the labor force distribution in other fields. Data from 1960 to 2010 were considered and

evaluated. For example, they calculated the propensity of a group to work in a particular occupation in a certain period.

As a basis, they were using the Roy model of occupational choice [93]. It writes that everyone performs the work for which they have the most talent and inclination. They extended the model by clarifying that everyone has an additional occupational preference, such as liking to work close to home. Moreover, they include strengths influencing the distribution of personnel. Their simulations include and define different parameters such as gender, background, occupation, occupational preferences, human capital, and human capital barriers. The parameters were determined to match the empirical targets. To describe the uncertainty of their predictions, they estimated the earnings per year in 2009 and compared them with the census data as displayed in Figure 9d. Our analysis revealed a mean average relative error of their predictions of 2.3%. Therefore, the model works with sufficient accuracy to describe the impact of lowering gender and background-specific barriers.

In the U. S., the proportion of white men in the occupations of doctor and lawyer dropped from 94% to 62% between 1960 and 2010. The model assumes that talent is distributed equally across different backgrounds. The time series decomposition of GDP growth (gross domestic product) per person was calculated and other parameters outlining different personnel distributions were estimated. As the overall GDP using the actual parameters is continually rising in the given period (Figure 9c, dark blue line), the green line demonstrates the GDP development if no changes in staff diversity were accomplished. It is likely that by including talents of all genders and backgrounds, the effect on productivity rises, and if female and black staff had been excluded from the 1970s on, growth would have stagnated. As illustrated in Figure 9a,b, the authors' modeling shows that the propensities to work in a particular job are the same for all groups considered over the period, but the barriers differ significantly. For example, it was still tough for a white woman to work as a doctor or lawyer in 1960, but the wall had fallen considerably by 2010. Nevertheless, the barriers for white women in various professions are still higher than for white men in 2010. The authors conclude that the inclusion of women and people from diverse backgrounds accounted for a 40% increase in GDP in 2010 (Figure 9c) [92].

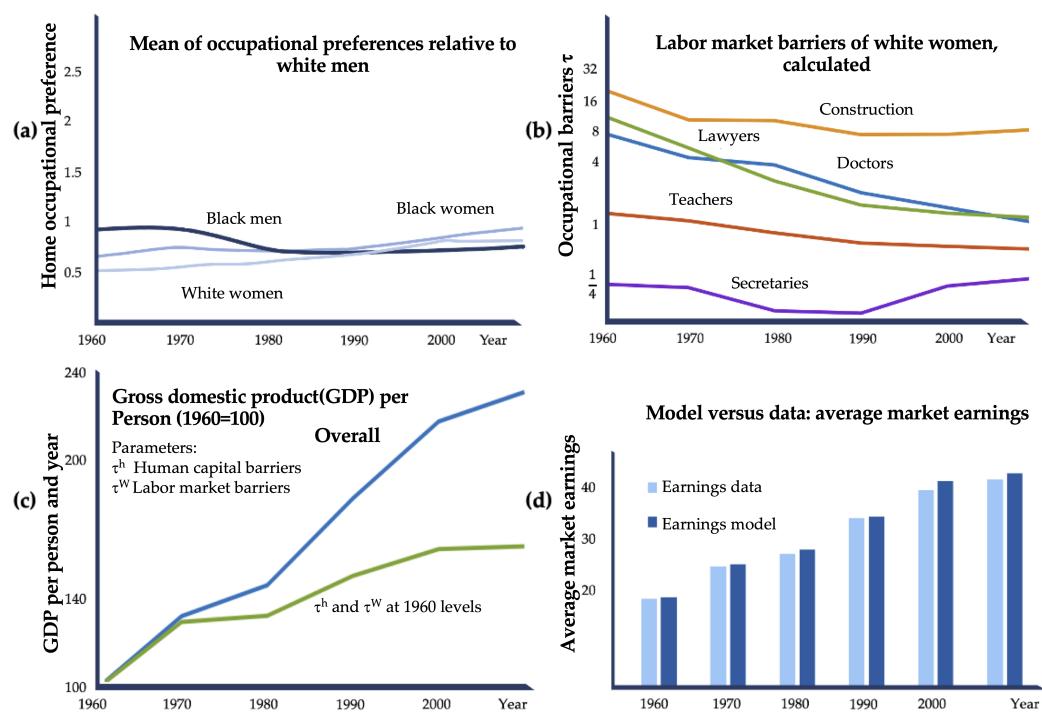
A study from German companies performed by the journal *Handelsblatt* confirmed these findings. Six out of eight departments found that diverse teams perform better than homogeneous teams [94,95]. The authors interviewed 1000 employed participants of different genders and backgrounds. They found that a large proportion of the directors of the companies surveyed were not female. In addition, respectively, 59% and 60% of the employees surveyed are missing equal opportunities for people with an immigrant background and women [94].

### 3.4. Industry 4.0 is in Line with the Sustainability Goals

We have presented the impact of the (new) technologies and business practices on the achievement of the 10 SDGs in Table 2. Naturally, each technology has a different impact depending on the company and process, but a general calculation reveals that future CMC departments act with more throughput while supporting sustainability goals. Compared to other industries, the development of new treatments depends not only on costs, turn over, and ecological input but primarily on the well-being and health of humankind.

Sustainability analysis is conducted to demonstrate the importance of intelligent laboratory automation to achieve the 17 development goals.

The TINS-D analysis is a tool to understand the effect of a new product or technology on the other four coordinates, that is, on nature, individuals, society, and democracy (Figure 10) [64]. Digitization, digitalization, automation, and machine learning are expected to contribute to sustainable development and democratization. Regarding the implementation of Industry 4.0 in CMC, the TINS-D analysis shows the impact on sustainability and the need for action by looking at each of these components:



**Figure 9.** Analysis of economic effect of changes of gender and background of staff in distinctive occupations. (a) Mean of occupational preferences relative to white men. (b) Occupational barriers for white women; authors' calculations based on model equation using census data. (c) GDP (gross domestic product) per person, data and counterfactual. The graph shows the cumulative growth in GDP per person in the data (overall) and in the model with no changes in former barriers. (d) Average market earnings per worker in 2009 in USD in the Census/American community survey data alongside the corresponding model values per year. The mean relative error of the estimations is 2.3%. Adapted from Hsieh et al., 2019: Figures 4, 6 and 7 and values from table IV in [92]. Parameters:  $\tau^h$  occupational barriers,  $\tau^h$  human capital barriers,  $\tau^W$  labor market barriers.

**Table 2.** Assessment of the impact of new technologies on sustainability goals published by the WHO. The KIWI-biolab fulfills most reported sustainability goals.

Technology and Methods	Good Health	Gender Equality	Decent Work, Economic Growth	Industry, Innovation, and Infrastructure	Reduced Inequality	Responsible Consumption and Production	Climate Action	Partnership for the Goals
High precision medicine [48,76,96,97]	✓			✓	✓	✓	✓	✓
Innovations in CMC [22]	✓			✓	✓	✓	✓	✓
Diversity [92]	✓	✓	✓	✓	✓	✓	✓	✓
<b>Bioprocess:</b>								
Automation [46]	✓		✓	✓	✓	✓	✓	✓
Design of Experiments [43]	✓		✓	✓	✓	✓	✓	✓
Model-based DoE [98]	✓		✓	✓	✓	✓	✓	✓
Knowledge management [99]	✓	✓	✓	✓	✓	✓	✓	✓
Wet lab PAT [78]	✓		✓	✓		✓	✓	
Soft sensors [81]	✓		✓	✓		✓	✓	
Modeling [77]	✓		✓	✓		✓	✓	
Hybrid models [50]	✓		✓	✓		✓	✓	
Online re-design [41,45]	✓		✓	✓		✓	✓	
<b>KIWI-biolab</b> [100]	✓	✓	✓	✓	✓	✓	✓	✓

**High-end technology.** From the technological point of view, automation and digitalization are the basis for developing a framework that can compute a digital twin. It is the next step toward an optimized process, the highest goal to reach from the technological and ecological point of view. The new technology supports data integrity which is essential in the industry. The new efficient technology affects nature by saving resources. The technology promotes an inclusive and democratic ecosystem in the company by making the

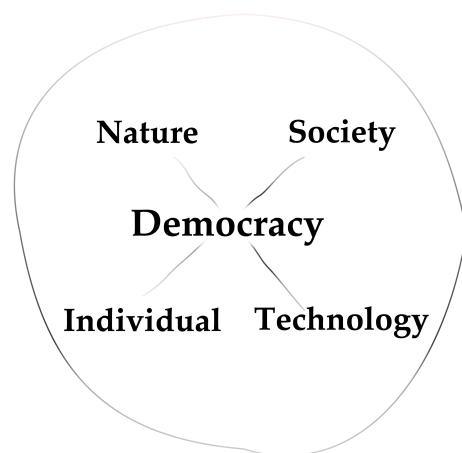
equipment easily accessible to people with all levels of training and people with disabilities. Further, digitalization, automation, and machine learning can improve user interaction and sustainability by reducing the number of experiments, saving resources, and reducing the failure rate.

**Positive impact on individuals.** The influence of digital change on individuals must be addressed by analyzing the effects on employees and customers. In the pharmaceutical industry, the definition of “customer” is multifaceted, as patients, physicians, pharmacies, and insurance companies consume products to varying degrees depending on each country’s healthcare system. The capabilities of the staff, who are working in interdisciplinary teams, and their way of communication need to be considered. A culture of equality and regular workshops promote innovation and thus adaptability to changes. Economic growth should be distributed equally in the hierarchy. Customer relationship management (CRM) ensures that the digital transformation also meets the customer’s expectations and that the marketing strategy fits the current trends and services. Individual impact results in a positive impact on society, enabling treatment of rare diseases and extending the reach to more patients in areas with less advanced infrastructure.

**Positive impact on nature.** The effect on nature can be revealed by demonstrating the green impact of the change. Measuring environmental sustainability is made more straightforward in an entirely digitized company. Parameters such as reducing resources to a minimum, the CO<sub>2</sub> footprint, and tools such as the Eco-care matrix can measure the overall positive effects. As a result of these positive effects, new resources and data will become available, which then can be integrated into the calculations for further insights.

**Positive impact on society.** Effects to society must be included in change management as pharmaceutical companies often have a negative social reputation. The societal impact and changes must be communicated transparently. These include the shortening of the development time of drugs, the companies’ research activities, promoting education, well-being, and the fact that the number of employees is not decreasing due to new technologies. The teams are diverse in terms of life-style, age, and skills. The customers need to understand why something is changing and benefit from it. It must also be communicated that economic growth is distributed evenly across society.

**Industry 4.0 is promoting democracy.** The digital transformation in the pharmaceutical industry is also leading to the increasing democratization of technology. By improving the functionality and user interface of the software, even without special qualifications, nearly anyone can use even the most advanced technologies. Resource-saving manufacturing is more environmentally friendly, less costly, and increases the availability of therapies worldwide.



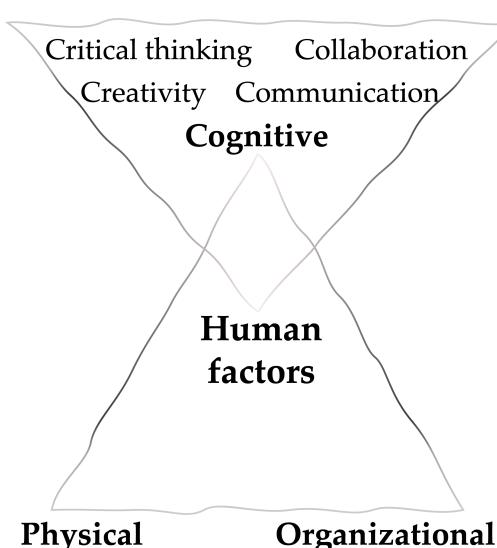
**Figure 10.** Sustainability analysis tool TINS-D (technology, individual, nature, society, democracy). The future project or technology is examined in terms of its effect on the subjects and their consequence for democracy. Further, it will be analyzed if the technology promotes equality. Adapted from André Baier, TU Berlin [64].

**A different outlook: Human factors are a central component in next-generation CMC.** Typically, change in management starts with the analysis of whether people are ready for the difference. The following factors must be reflected to enable the implementation of intelligent technologies and enable a sustainable company business. Proper change management must accompany the introduction of new technologies. The fear of reducing the workforce by future technologies is unfounded in biomanufacturing, as companies grew by 5.4% in 2020 [7]. Although new technologies should ease use, learning and adapting to change are necessary. At the same time, it is essential that the staff feels free enough to propose changes and develop new technologies [101]. As Ellyn Shook and Julie Sweet showed in 2019, diversity and a culture of equality are critical building blocks for business success and innovation in organizations [90,91]. To accelerate creativity and innovation in a company, staff must be encouraged to create and share information. An exemplary implementation of knowledge management, which can be considered the link between data governance and quality risk management, enables organizations to make decisions more efficiently and encourages an organizational culture of learning [99].

As not only are specialist disciplines mixed but employees and customers are located worldwide, it is advantageous to employ multilingual staff. The preferred skills include creativity, collaboration, critical thinking, and communication (4 C's) that are comprehensible to colleagues; therefore, we extended the traditional human factors in Figure 11 [102].

To cover the ever-increasing demand for specialized staff, companies started to change their way of recruitment by buying smaller companies, offering training and workshops, and public challenges scientists can apply for. Another way to gain new scientists is to let them optimize processes in academic laboratories by students and then transfer the process back to the company. The students are offered a job, and companies stay up to date by close collaboration. In the KIWI-biolab, pharmaceutical companies are optimizing their processes with students from TU Berlin; this is a win-win situation for both sides: students learn to work with the industry and under GMP-conform settings within the companies, and the companies can profit from the newest research results.

The authors think a common, half public and half industrial research institute for CMC processes would be beneficial as many companies face the same challenges. In summary, more interdisciplinary, diverse teams and the establishment of future skills in public institutes for drug development could drastically increase the number of qualified personnel.



**Figure 11.** Overview of the three main domains of human factors and ergonomics by International Ergonomics Association (IEA): The impact of cognitive, physical, and organizational factors and their

relationships to each other must be considered when introducing a new technology or planning a new subunit or when introducing new data management, i.e., perception and memory and their interplay with physical factors such as the anatomy are playing a major role in many occupations. Figure adapted from IEA and extended by 4 C's future skills [102–104].

### 3.5. Next-Generation Drug Development will Combine the Best Therapies, Data Science, and Sustainability

Automation, digitization, and digitalization will become a standard in all CMC functions. Companies and institutes aim to reduce the paper-based work to a minimum. There will be no manual copy-paste processes, but there will also be no drones flying around the company, no co-bots completely replacing human labor. People will perform challenging and fulfilling work, and most of them will be situated in front of a computer or tablet. Reducing paper-based work such as certification, validation, and data capture by new technologies is relatively simple with an enterprise-wide data structure, i.e., for clinical reports. The digital integration of all paper-based processes seems fairly easy to be managed; the highly regulated environment controls the digital transformation. In the end, employees will coordinate bots for end-to-end business processes. Paper-based processes will be entirely digitized at a highly standardized level. In the wet lab, where product characterization, formulation, and process scaling are performed, scientists will benefit from fully integrated labs and a central control center. All equipment in the wet lab runs under a standardized communication language. They are controlled remotely; automated and generated data are stored in a laboratory management system or electronic laboratory notebook. The devices can also communicate with each other, e.g., to indicate when a process step has been completed.

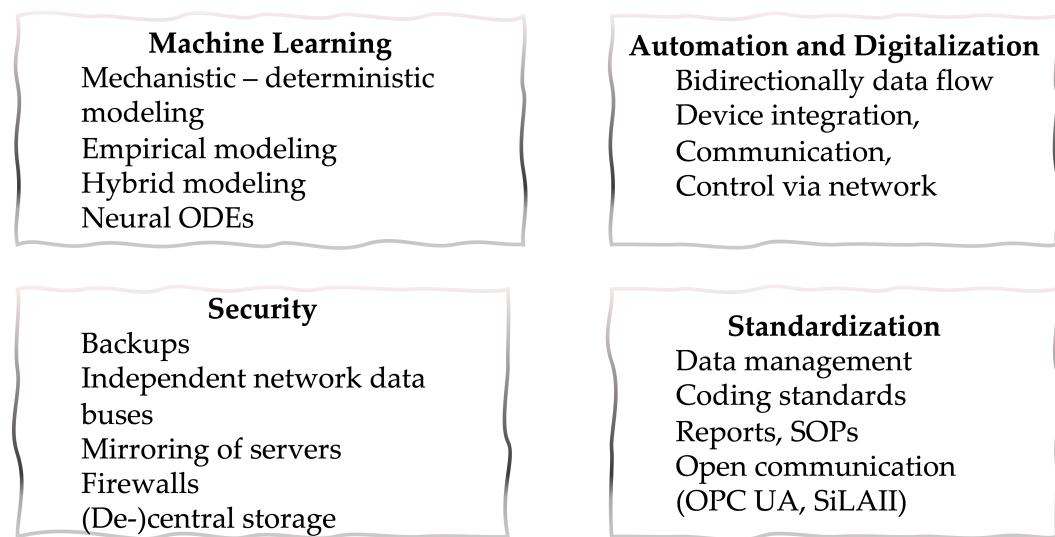
Machine learning will be integrated into the design and control of all processes. Automation and machine learning promote privileged skills and reduce the work of tedious, repetitive, and fixed operations. However, the great advantage of intelligent automation is that it will replace repetitive tasks, shape the current design and the following process. Future scientists can orchestrate multi-functional platforms to perform all necessary steps in drug development in a self-directed manner and at the most resource-efficient level.

Data are structured, available, and analyzed in real time. There is a central software that orchestrates the inputs from the wet lab, clinical trials, and previous therapy outcomes. Simple analyses are immediately available, and monitoring for documentation is automated. The data structure, storage, and handling are strictly standardized. Insights can be easily shared, and new therapy development will require a fraction of the effort compared to 2020. If this is how your CMC department wants to operate in the future, you need to start with this roadmap right away (Figure 12).

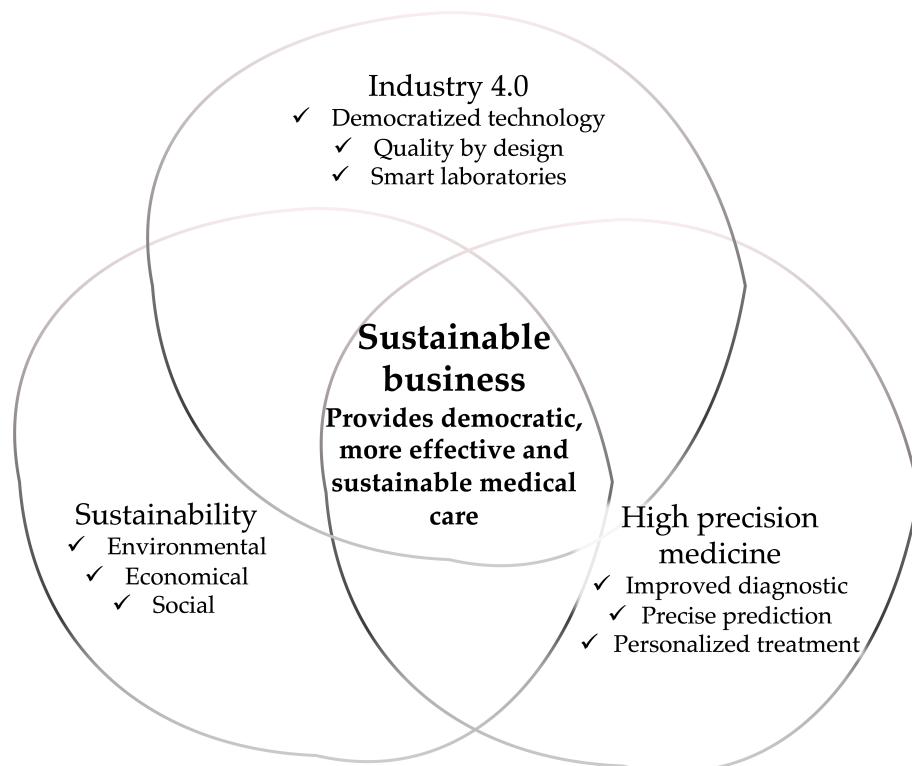
Process optimization will reach an unimaginable level. The software framework for process simulation and parameter estimation can generate a digital representation of specific processes, the digital twin. It will control automated test stations where new substances are run through and determine critical properties. A machine learning tool will decide which further tests need to be performed and schedule new experiments. Other tests and process optimizations will be carried out depending on their results.

High throughput devices are essential to finding the optimal process and screening conditions in combination with the methodology model-based DoE. The combination of intelligent experiment preparation including parameter dependencies and real-time optimization will speed up all processes in drug development [89]. The thoughtful planning of experiments by model-based DoE can reduce the wet lab work to 25% (Figure 4 in [98]).

An overall sustainable business model will be established (Figure 13). Ecological life cycle analysis will be automated and visualized via an Eco-Care matrix (Figure 6). A sustainability analysis will be performed in advance with established software tools taking a holistic view of different innovation paths and digital progress [26]. Ultimately, a self-driving lab changes the pace of drug development and meets the need for green solutions.



**Figure 12.** Roadmap to intelligent lab automation. This is a bold sketch showing which steps need to be taken toward digitization, digitalization, automation, and machine learning. The individual steps are not interdependent; machine learning can also be used without automation. We recommend building a new infrastructure, thinking about the security of the data, their transport between different devices and into a central platform. The communication as well as the storage of the data must comply with a standardized data format. Only the central storage of all experimental laboratory data as well as the results from clinical studies and administration makes intelligent process planning and control possible.



**Figure 13.** Sustainability, medicinal innovations, and data science go hand in hand in future CMC departments.

Talent acquisition will be inclusive, and the company will follow bold leadership. This creates teams with very different backgrounds and areas of expertise (interdisciplinary, intercultural, different experience levels). Global networking is feasible, e.g., robots could be controlled from another continent via cloud-based applications. The high degree of automated processes, especially regarding fixed and highly regulated tasks, makes employees like their job and concentrate on more creative and demanding tasks, saves money, and increases precision.

Employees will be rewarded with flexible working models, and the workforce structure will change fundamentally in the direction of flat hierarchies, i.e., shorter decision-making paths. Enthusiasm for change is maintained through sensible change management, e.g., workshops and further training of the employees. As is already practiced, family-friendliness is not only applied by the possibility of a home office but also by childcare options. The creative and equal atmosphere will accelerate innovation and help companies compete. Not just legal constraints but also stakeholders are demanding a sustainable business model, quantification of resource consumption, a diverse product strategy, fair treatment and compensation of personnel, fair prices, fair distribution of products, and a positive environmental impact. Implementing machine learning techniques in all areas of a sustainable business, to run all processes optimally, together with attracting and motivating top talent, would multiply the performance of drug development departments.

#### 4. Discussion

This review discusses different impacts on future drug development performance. High precision medicine, automation, machine learning, and sustainability will shape the next generation of CMC departments. Sustainability goals, new technologies, and new corporate governance are mutually supportive and intertwined. Machine learning is supported by automation and digitalization. The sustainability of new processes can thereby be improved and measured.

Data science supports biological therapies' breakthroughs, and these make up an ever-increasing proportion of newly approved therapies. Biologicals reached sales of 14.6 billion EUR in 2020 and a market share of 30.8% in Germany. In 2013, the inclusion of statistical methods such as model-based DoE and Principal Component Analysis to reduce wet lab work and build an efficient manufacturing method for microbial processes was previously discussed [86]. As presented here and earlier, the prediction of scale-up/scale-down processes improved dramatically over the past years by using past data to train models for the adjustment process [32]. Purely empirically-obtained models have a high consumption of resources in biotechnology. By implementing advanced machine learning techniques, which use chemical and biological mechanistic models in combination with empirical methods, the process from active pharmaceutical ingredient to the finished drug will be accelerated [52,53].

Machine learning-based process optimizations in CMC require validation. The economic benefit of ANN combined with mechanistic models or ODEs is obvious: saving resources. Still, their usefulness when it comes to validation in the final part of the value chain is not yet proven. From the scientific point of view, processes controlled using machine learning can be designed so that the resulting products will always meet the same quality. However, on the other hand, the construction of the digital twin must be transparent. This can only be guaranteed when using mechanistic/deterministic models and becomes more complicated when using ANN.

In 2017, an article about a holistic production control strategy was published aiming for robust and flexible production and integrated life cycle documentation [105]. This article covers the methodology of Data Integrity by Design and takes a holistic approach to parallel pharmaceutical development and management tasks. The acceleration of model-based solutions in every step of the value chain and the fact that drug development is the most significant part was highlighted by Narayanan et al. in 2019 [50].

A sustainable business model was announced as a competitive advantage when focusing on constant innovation. Later, when the business world called for disruptive innovations in the pharmaceutical industry, these took place through the invention of breakthrough therapies [96,97,106]. All-embracing sustainability is promoted through the creation of various indices such as the Dow Jones Sustainability Index or the Euronext Vigeo Eiris indices, which were discussed in 2020 by Buchholz et al. [26]. It measures environmental impact, economic value, consumer benefit, ethical aspects, societal engagement, governance, and plans to integrate digitalization (Figure 1 in [26]). The *Wall Street Journal* ranking also gives information on how sustainable a company is. There, human capital, social, environmental, and business and innovation rank is valued [107].

The idea of a sustainable business model was discussed from different perspectives as a continuous adaption to the market, innovative business models, environmental sustainability, or overall sustainability, including diversity in talent acquisition [27–29,108]. In 2009, an article in the *Harvard Business Review* was published claiming that there is no other way than sustainable business development in which they are focusing on environmental sustainability [28]. In general, companies should act more sustainable because they are better prepared for future legal regulations, enjoy better standing with stakeholders and customers, save resources, and can thus shorten the time to market, especially in the pharmaceutical industry. In the end, sustainability is expected to be integrated into the business development process, a conclusion this review also reached.

Involving machine learning to ensure sustainability and its measurement was promoted by Coeckelbergh et al. in 2020 [109]. A detailed assessment of the impact of AI on the Sustainable Development Goals is discussed in Vinueasa et al., who used consensus-based data and concluded that AI and machine learning techniques enable most of the development goals (79%) [34,109]. Moreover, the WHO published a detailed guide for policymakers, industry, and academia on how to contribute to the 17 SDGs in the E4As (Engage, Assess, Align, Accelerate, Account) Guide for Advancing Health and Sustainable Development. Among other things, the guide advises using lean management, promoting equality, and using digital technologies to save resources [2,3].

The KIWI-biolab, a future biotech lab situated at the TU Berlin, meets most criteria of the SDGs for sustainable drug development, as the lab is working efficiently by inter-connectivity and transparency and does a lot for the education of people and longevity of technologies (Table 2).

Finally, we are recommending a modified corporate sustainability measurement tool comparable to the published evaluation tool from Buchholz et al. in 2020, combined with a holistic Industry 4.0 approach to predict, adapt, and perform processes [26]. In addition, local policymakers and stakeholders should constantly study and implement the proposals of the local WHO office to achieve the health-related SDGs.

Scientific advances to accelerate high precision medicine and data science approaches in clinical trials are studied in publicly funded projects at the moment [22,100]. The project MELLODDY (MachinE Learning Ledger Orchestration for Drug DiscoverY) would cover the need for a proven, reliable, and data-secure platform using blockchain technologies through which companies could extract information from competition-related data [110]. This technology could solve the problem of using patient-related data for training machine learning models.

**Statistical validation.** Still, the validation issue is not solved yet in terms of validation from the authorities. If the model represents the real world, the mathematical proof has to be determined before. Rajamanickam et al. discussed model validation in bioprocessing as models are encouraged by the Quality by Design (QbD) initiative to improve process performance by model predictive control [60]. Here, statistical validation methods which are essential for authorization are described. To evaluate the predictability and interpretability, a large variety of processes and thus a large diversity of model validation methods have to be analyzed. A model can generalize a complex natural phenomenon, but not all critical influencing factors are known. When underfitting a model, too few influence

factors are included; when overfitting, too many influence factors are included, and the model learns irrelevant information. The validation approach depends on model type and structure; the easiest method is the data-driven validation and is widely accepted by the health authorities.

Data-driven validations are Coefficient of Determination ( $R^2$ ), Root Mean Squares Error (RMSE), the calculation of accuracy and precision (Closeness of Prediction to Real Value and Uncertainty), specificity (true negative rate), and sensitivity (actual positive rate), cross-validation (dataset is split into training and validation dataset), repeatability, Maximum likelihood, optimizing information criteria by Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), Goodness-of-Fit, to prove Model Uncertainty, Model Robustness, Credibility Score, and Continuous Testing. The authors agree that many points have to be considered when implementing a method for model validation as there is no joint procedure: the nature of the dataset, number, and kind of replicates, variation, sample size, model state, and type (statistical/mechanistic/hybrid) have to be considered [60].

**Regulatory validation.** The validation of machine learning models is included in official guidelines and standards. To date, computational modeling is widely described in the regulation of medical devices but less regulated in drug development of small molecules and biologicals [111]. In medical apps, where software as a medical device is used, twenty-nine apps have been approved by the FDA [111,112]. This includes exclusively medical applications that are not changing their calculations after marketing, so-called "locked" algorithms [112]. On the contrary, adaptive algorithms are used in bioprocessing and computed to learn from process data. Adaptive machine learning-based applications are used in drug discovery—what precedes the drug development in the value chain.

The FDA's Office of Science and Engineering Laboratories is investigating software reliability and AI/ML-based devices and working in various new research fields [113]. Moreover, the American Society of Mechanical Engineers and the FDA developed a framework for the validation of computational models for specific use in medical device development [114]. Bideault et al. performed a use case in bioprocess engineering based on this regulatory framework [115]. The fundamental problem in the validation of machine learning models in bioprocessing is that they learn and adapt during the process [41,115]. The authorities and the used models aim to maintain product quality (SQUIPP: Safety, Quality, Identity, Purity, and Potency) [116]. To reach this aim, adapting the original model might be necessary. The framework allows specific changes of used models if approved—that can happen by an approved change protocol, by a culture of quality, or real-world performance and documentation [115]. The tracking of changes could be realized with technology compared to the MELLODDY project [110]. The authors show that the proposed framework lacks methods to learn and adapt in real time, precisely what is needed in bioprocessing [115].

ICH guidelines describe model validation as part of model development, and as described before, verification has to be performed throughout the life cycle of the medicinal product. They further suggest categorizing models by their impact on the process. The FDA expects verification of the prediction by a reference method [117]. In process analytical technologies (PAT), statistical methods can be used for mathematical validation [118]. In computational models like the digital twin or digital health applications, the models must be described and their output validated by other analytical methods [119]. If the model is part of the production system, the corresponding software has to be validated and must be part of the quality system [120].

The HMA/EMA workshop on AI in 2021 medicines informed about AI applications in medication and its regulation, presented the latest innovations in AI applications, engaged industrial, academic participants, and authorities, and communicated with them on the prioritization of AI-specific recommendations. The use of standard data models to facilitate the external validation of prediction models across multiple electronic health records databases was discussed. Moreover, digital twins to enable personalized medicine were

presented. Moreover, the use of machine learning in the life cycle of medicinal products was shown. It was addressed that these will be used very likely in the future to profit from real-world data, so after marketing, to further improve personalized medicine. Moreover, more use cases were shared using AI for data reconciliation, regulatory portfolio planning, pharmacovigilance, and statistical table quality control.

## 5. Conclusions

The change toward a sustainable business model results in an innovative, long-living business, but the change process should be organized precisely. Moreover, the limitations of fully integrated laboratories should be mentioned. First, for small companies, the costs may exceed the benefits: liquid handling stations, servers, and computational power are expensive and must be maintained. Additional personnel is necessary to maintain the network and make sure communication and control of the devices are secured. Moreover, a pause in production must be accepted until the new system is implemented. A running system can be attacked, and the company has to ensure data protection requirements (see Figure 12).

The regulations by the authorities are constantly changing, and they are incorporating and accepting more and more machine learning-based processes. Moreover, if not financially beneficial in the first place, digitalization and intelligent lab automation will pay off to ensure the best product quality, the most transparent production process, and a sustainable business.

Therefore, it is wise to start with one use case, hire specially trained personnel, and test the system beforehand. It may be possible to establish an Industry 4.0 connected lab parallel to the old system in larger companies.

Ideally, new self-driving laboratories are built from scratch. This chance can be taken by supporting healthcare in low-income countries as the costly implementation of new technologies should be supported worldwide to reduce inequality. Especially in countries where more and more generics are being produced, companies can take the chance to use the latest technologies and profit from the innovative potential of the target regions. The local production is supposed to save costs and support the area by creating jobs. The new companies need appropriate structural and technical equipment trained staff to reach GMP-compliance and low-cost drug production for prequalification by the WHO [121]. The WHO prequalification program aims to support affordable treatments in resource-limited regions.

**Author Contributions:** K.P. is the main author of the presented research. S.S., C.L., A.K. and A.B. worked on several sections of the manuscript. M.N.C.-B. is the principal investigator of the KIWI-biolab. D.H. supervised the manuscript creation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Bundesministerium für Bildung und Forschung, BMBF, grant number: 01DD20002A.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We are thanking Peter Neubauer for the possibility to work in and with the KIWI-biolab and scientific discussions. Thanks to the HTBD group of the TU Berlin for substantially shaping the future lab. We would like to express our special thanks to Sabine Lühr-Müller, Thomas Högl, and the lab management team for technical help. Another thanks goes to Irmgard Schäffl for proofreading.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	artificial intelligence
ANN	artificial neural network
CAR T-cell	chimeric antigen receptor T-cell
CMC	chemistry, manufacturing, and control
CO <sub>2</sub>	Carbon Dioxide
CRM	Customer relationship management
CV	coefficient of variation
DoE	Design of experiments
DOT	dissolved oxygen tension
<i>E. coli</i>	<i>Escherichia coli</i>
ECM	Eco-Care Matrix
EMA	European Medicines Agency
EU	European Union
FDA	Food and Drug Administration
GMP	good manufacturing practice
IoT	Internet of Things
M2M	machine to machine
MELLODDY	MachinE Learning Ledger Orchestration for Drug DiscoverY
mRNA	messenger ribonucleic acid
Neural ODEs	neural ordinary differential equations
OVAT	one-variable-at-a-time
PAT	process analytical technologies
PCA	Principal Component Analysis
QbD	Quality by Design
QSAR	quantitative structure–activity relationship
RNN	Recurrent Neural Networks
TU Berlin	Technische Universität Berlin
U. S.	United States
UN	United Nations
vfa bio	Die forschenden Pharma-Unternehmen bio

## References

1. Wouters, O.J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA* **2020**, *323*, 844. [[CrossRef](#)]
2. World Health Organization. *E4As Guide for Advancing Health and Sustainable Development. Resources and Tools for Policy Development and Implementation*, 1st ed.; WHO Regional Office for Europe: Copenhagen, Denmark, 2021.
3. World Health Organization. *Report of the Regional Director: The Work of the WHO Regional Office for Europe in 2020–2021*; Technical Report; WHO Regional Office for Europe: Copenhagen, Denmark, 2021.
4. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D. Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. *Dmw-Dtsch. Med. Wochenschr.* **2011**, *136*, e9–e15. [[CrossRef](#)]
5. 17 Goals for a Sustainable Future by the Welthungerhilfe.de. Available online: <https://www.welthungerhilfe.org/our-work/focus-areas/civil-society-and-advocacy/sustainable-development-goals/> (accessed on 10 September 2021).
6. United Nations General Assembly. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: New York, NY, USA, 2015.
7. Lücke, J.; Bädeker, M.; Hildinger, M.; The Boston Consulting Group, München. *Medizinische Biotechnologie in Deutschland 2021*; Vfa-Publikationen, Deutschland, The Boston Consulting Group: München, Germany, 2021.
8. European Federation of Pharmaceutical Industries and Associations. The Pharmaceutical Industry in Figures. Available online: [https://www.efpia.eu/media/554521/efpia\\_pharmafigures\\_2020\\_web.pdf](https://www.efpia.eu/media/554521/efpia_pharmafigures_2020_web.pdf) (accessed on 15 February 2022).
9. KPMG. *Digitalization in Life Sciences*; KPMG International: Birmingham, UK; Berlin, Germany, 2018; p. 48.
10. D/A/CH Affiliate: Containment Manual (English Translation). International Society for Pharmaceutical Engineering. 2017. Available online: <https://ispe.org/publications/guidance-documents/dach-affiliate-containment-manual-english-translation/> (accessed on 22 March 2022).
11. Kaplan, W.; Laing, R. *Local Production of Pharmaceuticals: Industrial Policy and Access to Medicines, an Overview of Key Concepts, Issues and Opportunities for Future Research*; World Bank: Washington, DC, USA, 2005.

12. Die forschenden Pharmaunternehmen | vfa. Was Digitalisierung für die Pharma-Industrie bedeutet. Available online: <https://www.vfa.de/de/wirtschaft-politik/pharma-digital> (accessed on 10 September 2021).
13. Die forschenden Pharmaunternehmen | vfa. Big Data und KI in der Pharma-Industrie. Available online: <https://www.vfa.de/de/wirtschaft-politik/pharma-digital/zukunft-und-debatte/big-data-und-ki-fuer-die-pharmaindustrie> (accessed on 9 September 2021).
14. Rifaioglu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Doğan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Briefings Bioinform.* **2019**, *20*, 1878–1912. [CrossRef]
15. Merck KGaA SYNTHIA™, die KI-Software für die Wirkstoffentwicklung. Available online: <https://www.merckgroup.com/de/research/science-space/envisioning-tomorrow/future-of-scientific-work/synthia.html> (accessed on 9 September 2021).
16. Silver, L.; Smith, A.; Johnson, C.; Taylor, K.; Jiang, J.; Anderson, M.; Rainie, L. Mobile connectivity in emerging economies. *Pew Res. Cent.* **2019**, *7*, 72.
17. Poushter, J. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Res. Cent.* **2016**, *22*, 1–44.
18. Kay, M.; Santos, J.; Takane, M. *mHealth: New Horizons for Health through Mobile Technologies*; World Health Organization: Geneva, Switzerland, 2011.
19. Smyth, P.; de Lannoy, G.; Stosch, M.V.; Pysik, A.; Khan, A. Machine learning in research and development of new vaccines products: Opportunities and challenges. *Comput. Intell.* **2019**, *6*.
20. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437–1452.e17. [CrossRef]
21. Purdie, F.P. *Data Integrity and Compliance with CGMP Guidance for Industry*; FDA: Rockville, MD, USA, 2016; p. 13.
22. Cauchon, N.S.; Oghamian, S.; Hassanpour, S.; Abernathy, M. Innovation in Chemistry, Manufacturing, and Controls—A Regulatory Perspective from Industry. *J. Pharm. Sci.* **2019**, *108*, 2207–2237. [CrossRef]
23. Free Image on Pixabay—Caffeine, Molecule. Available online: <https://pixabay.com/vectors/caffeine-molecule-chemical-structure-148821/> (accessed on 21 September 2021).
24. Free Image on Pixabay—Vaccine, Cure, Medicine, Virus. Available online: <https://pixabay.com/photos/vaccine-cure-medicine-virus-5897391/> (accessed on 1 September 2021)
25. Free Image on Pixabay—Blood Cells, Cells, Biology, Liquid. Available online: <https://pixabay.com/illustrations/blood-cells-cells-biology-liquid-2177469/> (accessed on 21 September 2021).
26. Buchholz, H.; Eberle, T.; Klevesath, M.; Jürgens, A.; Beal, D.; Baic, A.; Radeke, J. Forward Thinking for Sustainable Business Value: A New Method for Impact Valuation. *Sustainability* **2020**, *12*, 8420. [CrossRef]
27. Geissdoerfer, M.; Vladimirova, D.; Evans, S. Sustainable business model innovation: A review. *J. Clean. Prod.* **2018**, *198*, 401–416. [CrossRef]
28. Ram Nidumolu, C.K.P. Why Sustainability Is Now the Key Driver of Innovation. *Harv. Bus. Rev.* **2009**, *87*, 56–64.
29. Schaltegger, S.; Hansen, E.G.; Lüdeke-Freund, F. Business Models for Sustainability: Origins, Present Research, and Future Avenues. *Organ. Environ.* **2016**, *29*, 3–10. [CrossRef]
30. Business Perspectives. *Environmental Economics*; LLC Consulting Publishing Company “Business Perspectives”: Sumy, Ukraine, 2021.
31. Mowbray, M.; Savage, T.; Wu, C.; Song, Z.; Cho, B.A.; Del Rio-Chanona, E.A.; Zhang, D. Machine learning for biochemical engineering: A review. *Biochem. Eng. J.* **2021**, *172*, 108054. [CrossRef]
32. Smiatek, J.; Jung, A.; Bluhmki, E. Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing. *Trends Biotechnol.* **2020**, *38*, 1141–1153. [CrossRef]
33. Radivojević, T.; Costello, Z.; Workman, K.; Garcia Martin, H. A machine learning Automated Recommendation Tool for synthetic biology. *Nat. Commun.* **2020**, *11*, 4879. [CrossRef]
34. Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Fuso Nerini, F. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* **2020**, *11*, 233. [CrossRef]
35. Hejnaes, K.R.; Ransohoff, T.C. Chemistry, Manufacture and Control. In *Biopharmaceutical Processing*; Elsevier: London, UK, 2018; pp. 1105–1136. [CrossRef]
36. U.S. Food and Drug Administration, Office of the Commissioner U.S. 2021. Available online: <https://www.fda.gov/home> (accessed on 9 September 2021).
37. European Medicines Agency. Detailed Guidance on the Electronic Submission of Information on Medicinal Products for Human Use by Marketing Authorisation Holders to the European Medicines Agency in Accordance with Article 57(2), Second Subparagraph of Regulation (EC) No. 726/2004. 2012. Available online: [https://www.ema.europa.eu/en/documents/other/chapter-3ii-xevprm-user-guidance-detailed-guidance-electronic-submission-information-medicinal\\_en.pdf](https://www.ema.europa.eu/en/documents/other/chapter-3ii-xevprm-user-guidance-detailed-guidance-electronic-submission-information-medicinal_en.pdf) (accessed on 9 September 2021).
38. EUR-Lex. *Council Conclusions on Personalised Medicine for Patients*; Doc ID: 52015XG1217(01); European Union: Brussels, Belgium, 2015.
39. International Consortium for Personalised Medicine | ICPerMed. *What Is Personalised Medicine?* International Consortium for Personalised Medicine | ICPerMed: Bonn, Germany. 2021. Available online: <https://www.icpermed.eu/en/icpermed-medicine.php> (accessed on 10 September 2021).
40. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. [CrossRef]

41. Hans, S.; Haby, B.; Krausch, N.; Barz, T.; Neubauer, P.; Cruz-Bournazou, M.N. Automated Conditional Screening of Multiple Escherichia coli Strains in Parallel Adaptive Fed-Batch Cultivations. *Bioengineering* **2020**, *7*, 145. [[CrossRef](#)]
42. Free Image on Pixabay—Bmw, Car, Roadster, Sports Car. Available online: <https://pixabay.com/vectors/bmw-car-roadster-sports-car-158703/> (accessed on 21 September 2021).
43. Bowden, G.D.; Pichler, B.J.; Maurer, A. A Design of Experiments (DoE) Approach Accelerates the Optimization of Copper-Mediated 18F-Fluorination Reactions of Arylstannanes. *Sci. Rep.* **2019**, *9*, 11370. [[CrossRef](#)]
44. Anane, E.; García, Á.C.; Haby, B.; Hans, S.; Krausch, N.; Krewinkel, M.; Hauptmann, P.; Neubauer, P.; Cruz Bournazou, M.N. A model-based framework for parallel scale-down fed-batch cultivations in mini-bioreactors for accelerated phenotyping. *Biotechnol. Bioeng.* **2019**, *116*, 2906–2918. [[CrossRef](#)]
45. Cruz Bournazou, M.; Barz, T.; Nickel, D.; Lopez Cárdenas, D.; Glauche, F.; Knepper, A.; Neubauer, P. Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities: Online Optimal Experimental Re-Design in Robotic. *Biotechnol. Bioeng.* **2017**, *114*, 610–619. [[CrossRef](#)] [[PubMed](#)]
46. Sawatzki, A.; Narayanan, H.; Haby, B.; Krausch, N. 2 Accelerated Bioprocess Development 3 of Endopolygalacturonase-Production with 4 Saccharomyces cerevisiae using Multivariate 5 Prediction in a 48 Mini-Bioreactor Automated 6 Platform. *Bioengineering* **2018**, *5*, 101.
47. Freedman, L.P.; Cockburn, I.M.; Simcoe, T.S. The Economics of Reproducibility in Preclinical Research. *PLoS Biol.* **2015**, *13*, e1002165. [[CrossRef](#)] [[PubMed](#)]
48. The ENCODE Project Consortium; Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shores, N.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A.; et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **2020**, *583*, 699–710. [[CrossRef](#)]
49. von Stosch, M.; Hamelink, J.M.; Oliveira, R. Hybrid modeling as a QbD/PAT tool in process development: An industrial *E. coli* case study. *Bioprocess Biosyst. Eng.* **2016**, *39*, 773–784. [[CrossRef](#)]
50. Narayanan, H.; Luna, M.F.; Stosch, M.; Cruz Bournazou, M.N.; Polotti, G.; Morbidelli, M.; Butté, A.; Sokolov, M. Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnol. J.* **2020**, *15*, 1900172. [[CrossRef](#)]
51. Narayanan, H.; Cruz Bournazou, M.N.; Guillén Gosálbez, G.; Butté, A. Functional-Hybrid modeling through automated adaptive symbolic regression for interpretable mathematical expressions. *Chem. Eng. J.* **2022**, *430*, 133032. [[CrossRef](#)]
52. Lagergren, J.H.; Nardini, J.T.; Baker, R.E.; Simpson, M.J.; Flores, K.B. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLoS Comput. Biol.* **2020**, *16*, e1008462. [[CrossRef](#)]
53. Yazdani, A.; Lu, L.; Raissi, M.; Karniadakis, G.E. Systems biology informed deep learning for inferring parameters and hidden dynamics. *PLoS Comput. Biol.* **2020**, *16*, e1007575. [[CrossRef](#)]
54. Chen, R.T.Q.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. Neural Ordinary Differential Equations. *arXiv* **2019**, arXiv:1806.07366.
55. Lever, J.; Krzywinski, M.; Altman, N. Points of significance: Model selection and overfitting. *Nat. Methods* **2016**, *13*, 703–705. [[CrossRef](#)]
56. Walsh, I.; Fishman, D.; Garcia-Gasulla, D.; Titma, T.; Pollastri, G.; ELIXIR Machine Learning Focus Group; Harrow, J.; Psomopoulos, F.E.; Tosatto, S.C.E. DOME: Recommendations for supervised machine learning validation in biology. *Nat. Methods* **2021**, *18*, 1122–1127. [[CrossRef](#)] [[PubMed](#)]
57. Smiatek, J.; Jung, A.; Bluhmki, E. Validation Is Not Verification: Precise Terminology and Scientific Methods in Bioprocess Modeling. *Trends Biotechnol.* **2021**, *39*, 1117–1119. [[CrossRef](#)] [[PubMed](#)]
58. Chicco, D. Ten quick tips for machine learning in computational biology. *Biodata Min.* **2017**, *10*, 1–17. [[CrossRef](#)] [[PubMed](#)]
59. Bishop, D. Rein in the four horsemen of irreproducibility. *Nature* **2019**, *568*, 435–436. [[CrossRef](#)] [[PubMed](#)]
60. Rajamanickam, V.; Babel, H.; Montano-Herrera, L.; Ehsani, A.; Stiefel, F.; Haider, S.; Presser, B.; Knapp, B. About Model Validation in Bioprocessing. *Processes* **2021**, *9*, 961. [[CrossRef](#)]
61. Guisasola, A.; Baeza, J.; Carrera, J.; Sin, G.; Vanrolleghem, P.; Lafuente, J. The Influence of Experimental Data Quality and Quantity on Parameter Estimation Accuracy: Andrews Inhibition Model as a Case Study. *Educ. Chem. Eng.* **2006**, *1*, 139–145. [[CrossRef](#)]
62. Anane, E.; Barz, T.; Sin, G.; Gernaey, K.V.; Neubauer, P.; Bournazou, M.N.C. Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. *Biochem. Eng. J.* **2019**, *150*, 107247. [[CrossRef](#)]
63. Franceschini, G.; Macchietto, S. Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci.* **2008**, *63*, 4846–4872. [[CrossRef](#)]
64. Baier, A. *Education for Sustainable Development within the Engineering Sciences—Design of Learning Outcomes and a Subsequent Course Evaluation*; Technische Universitaet Berlin: Berlin, Germany, 2019.
65. Holland, M. *Deepmind: KI schafft Durchbruch bei der Proteinfaltung*; Heise Medien: Hannover, Germany, 2020. Available online: <https://www.heise.de/news/Deepmind-KI-schafft-Durchbruch-bei-der-Proteinfaltung-4975964.html> (accessed on 21 September 2021).
66. Bystrzanowska, M.; Tobiszewski, M. Chemometrics for selection, prediction, and classification of sustainable solutions for green chemistry—A review. *Symmetry* **2020**, *12*, 2055. [[CrossRef](#)]
67. Tong, W.; Hong, H.; Xie, Q.; Shi, L.; Fang, H.; Perkins, R. Assessing QSAR limitations—A regulatory perspective. *Curr.-Comput.-Aided Drug Des.* **2005**, *1*, 195–205. [[CrossRef](#)]

68. Korany, M.A.; Gazy, A.A.; Khamis, E.F.; Ragab, M.A.; Kamal, M.F. Analysis of closely related antioxidant nutraceuticals using the green analytical methodology of ANN and smart spectrophotometric methods. *J. Aoac Int.* **2017**, *100*, 8–17. [CrossRef] [PubMed]
69. Damerval, M.; Fagnoni-Legat, C.; Louvrier, A.; Fischer, S.; Clairet, A.L.; Limat, S.; Nerich, V.; Madeleine, I.; Kroemer, M. ATMP environmental exposure assessment in European healthcare settings, a systematic review of the literature. *Front. Med.* **2021**, *2162*. [CrossRef]
70. Christopher, J.S.; Boatright, M. *Challenges of ATMP Commercial Production: Facility Design for Newcomers*; ISPE | International Society for Pharmaceutical Engineering: Tampa, FL, USA, 2020.
71. Byrne, J. *Lonza Hails ‘Significant Milestone’ as First Lymphoma Patient Treated with Cocoon CAR-T System*; BioPharma Reporter: Rochdale, UK, 2020.
72. Kim, K.; Yuraszeck, C.; O’Connor, J. When using a closed and automated manufacturing platform, is there an option to maintain flexibility? *Cell Gene Ther. Insights* **2021**, *7*, 857–869. [CrossRef]
73. Neo, B.H.; Bandapalle, S.; O’Connor, J.; Lin, K.; Daita, K.; Sei, J.; Zander, M.; Gleissner, T.; Schroeder, J.; Abraham, E.; et al. *The Cocoon® Platform Combined with the 4D-Nucleofector™ LV Unit A Non-Viral Workflow for Modifying Primary T-Cells*; LONZA: Basel, Switzerland, 2020; p. 7.
74. Roth, T.L.; Puig-Saus, C.; Yu, R.; Shifrut, E.; Carnevale, J.; Li, P.J.; Hiatt, J.; Saco, J.; Krystofinski, P.; Li, H.; et al. Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **2018**, *559*, 405–409. [CrossRef] [PubMed]
75. Shi, Y.; McAfee, E.; Banda-Palle, S.; Siehoff, A.; Gleissner, T.; O’Connor, J.; Abraham, E.; Purpura, K.; Trainor, N.; Smith, T. End-to-End Cell Therapy Automation. WO 2019/046766 A2, 7 March 2019.
76. Baidak, B.; Hussain, Y.; Kelminson, E.; Jones, T.R.; Franke, L.; Haehn, D. CellProfiler Analyst Web (CPAW) - Exploration, analysis, and classification of biological images on the web. In Proceedings of the IEEE Visualization Short Paper (IEEE VIS), New Orleans, LA, USA, 24–29 October 2021.
77. Haby, B.; Hans, S.; Anane, E.; Sawatzki, A.; Krausch, N.; Neubauer, P.; Cruz Bournazou, M.N. Integrated Robotic Mini Bioreactor Platform for Automated, Parallel Microbial Cultivation With Online Data Handling and Process Control. *Slas Technol. Transl. Life Sci. Innov.* **2019**, *24*, 569–582. [CrossRef] [PubMed]
78. Food and Drug Administration. Guidance for Industry PAT—A Framework for Innovative Pharmaceutical Development, manufacturing, and Quality Assurance. 2004; p. 19. Available online: <http://www.fda.gov/cder/guidance/published.html> (accessed on 23 March 2021).
79. Kögler, M.; Itkonen, J.; Viitala, T.; Casteleijn, M.G. Assessment of recombinant protein production in *E. coli* with Time-Gated Surface Enhanced Raman Spectroscopy (TG-SERS). *Sci. Rep.* **2020**, *10*, 2472. [CrossRef]
80. Schlappa, S.; Brenker, L.J.; Bressel, L.; Hass, R.; Münzberg, M. Process Characterization of Polyvinyl Acetate Emulsions Applying Inline Photon Density Wave Spectroscopy at High Solid Contents. *Polymers* **2021**, *13*, 669. [CrossRef]
81. Krämer, D.; King, R. A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. *J. Process. Control.* **2019**, *82*, 91–104. [CrossRef]
82. Auer, J.; Bey, N.; Schäfer, J.M. Combined Life Cycle Assessment and Life Cycle Costing in the Eco-Care-Matrix: A case study on the performance of a modernized manufacturing system for glass containers. *J. Clean. Prod.* **2017**, *141*, 99–109. [CrossRef]
83. Moore, M. What Is Industry 4.0? Everything You Need to Know. 2019. Available online: <https://www.techradar.com/news/what-is-industry-40-everything-you-need-to-know> (accessed on 23 March 2021).
84. Parsad, B. The engineer’s perspective: Evolution in cell therapy bioprocess automation. *Cell Gene Ther. Insights* **2020**, *6*, 1615–1621. [CrossRef]
85. Häse, F.; Roch, L.M.; Aspuru-Guzik, A. Next-generation experimentation with self-driving laboratories. *Trends Chem.* **2019**, *1*, 282–291. [CrossRef]
86. Neubauer, P.; Cruz, N.; Glauche, F.; Junne, S.; Knepper, A.; Raven, M. Consistent development of bioprocesses from microliter cultures to the industrial scale. *Eng. Life Sci.* **2013**, *13*, 224–238. [CrossRef]
87. Neubauer, P.; Glauche, F.; Cruz-Bournazou, M.N. Editorial: Bioprocess Development in the era of digitalization. *Eng. Life Sci.* **2017**, *17*, 1140–1141. [CrossRef]
88. Neubauer, P.; Lin, H.Y.; Mathiszik, B. Metabolic load of recombinant protein production: Inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in *Escherichia coli*. *Biotechnol. Bioeng.* **2003**, *83*, 53–64. [CrossRef]
89. Hernández Rodríguez, T.; Posch, C.; Schmutzhard, J.; Stettner, J.; Weihs, C.; Pörtner, R.; Frahm, B. Predicting industrial-scale cell culture seed trains—A Bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using an MCMC method. *Biotechnol. Bioeng.* **2019**, *116*, 2944–2959. [CrossRef]
90. Shook, E.; Sweet, J. When She Rises, We All Rise. *Accenture* **2018**, 1–23.
91. Shook, E.; Sweet, J. Equality = Innovation. *Accenture* **2019**, 1–36.
92. Hsieh, C.-T.; Hurst, E.; Jones, C.I.; Klenow, P.J. The allocation of talent and U.S. economic growth. *Econometrica* **2019**, *87*, 1439–1474. [CrossRef]
93. Roy, A.D. Some thoughts on the distribution of earnings 1. *Oxf. Econ. Pap.* **1951**, *3*, 135–146. [CrossRef]
94. Dohmeyer, M.; Dittrich, C. *Vielfalt 2020-Die truffls Studie zur Wahrnehmung von Diversity in deutschen Unternehmen*; Handelsblatt: Düsseldorf, Germany, 2020.

95. Kewes, T. *Studie zu Diversity: Gerade in Krisen profitieren Unternehmen von Vielfalt*; Handelsblatt GmbH—ein Unternehmen der Handelsblatt Media Group GmbH & Co. KG: Düsseldorf, Germany, 2020; ISSN 0017-7296.
96. Allam, M.; Cai, S.; Coskun, A.F. Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics. *Npj Precis. Oncol.* **2020**, *4*, 11. [[CrossRef](#)]
97. Blasiak, A.; Khong, J.; Kee, T. CURATE.AI: Optimizing Personalized Medicine with Artificial Intelligence. *Slas Technol. Transl. Life Sci. Innov.* **2020**, *25*, 95–105. [[CrossRef](#)]
98. Möller, J.; Kuchemüller, K.B.; Steinmetz, T.; Koopmann, K.S.; Pörtner, R. Model-assisted Design of Experiments as a concept for knowledge-based bioprocess development. *Bioprocess Biosyst. Eng.* **2019**, *42*, 867–882. [[CrossRef](#)]
99. ISPE | International Society for Pharmaceutical Engineering GAMP RDI Good Practice Guide: Data Integrity by Design, U.S. 2020. Available online: <https://ispe.org/publications/guidance-documents/gamp-rdi-good-practice-guide-data-integrity-design> (accessed on 9 September 2021).
100. Neubauer, P. This Is the KIWI-Biolab, a BMBF KI Future Lab! Available online: <https://kiwi-biolab.de> (accessed on 23 March 2021).
101. Ho, C.F.; Hsieh, P.H.; Hung, W.H. Enablers and processes for effective knowledge management. *Ind. Manag. Data Syst.* **2014**, *114*, 734–754. [[CrossRef](#)]
102. Stauffer, B. What Are the 4 C's of 21st Century Skills? Available online: <https://www.aeseducation.com/blog/four-cs-21st-century-skills> (accessed on 8 September 2021).
103. International Ergonomics Association | IEA *Human Factors*; IEA: Geneva, Switzerland, 2021. Available online: <https://iea.cc/what-is-ergonomics/> (accessed on 10 September 2021).
104. Bridger, R. *Introduction to Human Factors and Ergonomics*, 4th ed.; CRC Press: Portsmouth, Hampshire, UK, 2017. ISBN 9781498795944.
105. Schepers, T.; Beutel, S.; McGuinness, N.; Heiden, S.; Oldiges, M.; Lammers, F.; Reardon, K.F. Digitalization and Bioprocessing: Promises and Challenges. In *Digital Twins*; Herwig, C., Pörtner, R., Möller, J., Eds.; Advances in Biochemical Engineering/Biotechnology; Springer International Publishing: Cham, Switzerland, 2020; Volume 176, pp. 57–69. [[CrossRef](#)]
106. June, C.H.; O'Connor, R.S.; Kawalekar, O.U.; Chassemi, S.; Milone, M.C. CAR T cell immunotherapy for human cancer. *Science* **2018**, *359*, 1361–1365. [[CrossRef](#)]
107. These Are the World's 10 Most Sustainably Managed Companies. *Wall Street Journal*. 13 October 2020, p. 1. Available online: <https://www.wsj.com/articles/these-are-the-worlds-10-most-sustainable-companies-11602624830> (accessed on 22 March 2022).
108. Kundu, S. Mathematical Modeling As A Tool For Sustainable Development. *J. Artic. Math. Educ.* **2018**, *5*, 348–350.
109. Coeckelbergh, M. AI for climate: Freedom, justice, and other ethical and political challenges. *AI Ethics* **2021**, *1*, 67–72. [[CrossRef](#)]
110. France, O. Melloddy, MachinE Learning Ledger Orchestration for Drug DiscoverY. Available online: <https://cordis.europa.eu/project/id/831472/de> (accessed on 23 March 2021).
111. Benjamins, S.; Dhunnoo, P.; Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. *NPJ Digit. Med.* **2020**, *3*, 118. [[CrossRef](#)]
112. Food and Drug Administration. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*; Food and Drug Administration: Silver Spring, MD, USA, 2019.
113. Food and Drug Administration. The FDA's Office of Science and Engineering Laboratories. Available online: <https://www.fda.gov/about-fda/cdrh-offices/office-science-and-engineering-laboratories> (accessed on 23 March 2021).
114. Morrison, T.M.; Hariharan, P.; Funkhouser, C.M.; Afshari, P.; Goodin, M.; Horner, M. Assessing computational model credibility using a risk-based framework: application to hemolysis in centrifugal blood pumps. *Asaio J.* **2019**, *65*, 349. [[CrossRef](#)]
115. Bideault, G.; Scaccia, A.; Zahel, T.; Landertinger, R.W.; Daluwatte, C. Verification and validation of computational models used in pharmaceutical manufacturing: Potential application of the ASME verification and validation 40 standard and FDA proposed AI/ML model life cycle management framework. *J. Pharm. Sci.* **2021**, *110*, 1540–1544. [[CrossRef](#)]
116. Evans, J.L.; Hillyer, C.D. Introduction to Quality Systems and Quality Management. In *Transfusion Medicine and Hemostasis*; Elsevier: London, UK, 2009; pp. 13–18.
117. FDA. Guidance for Industry, Q8, Q9, Q10 Questions and Answers. Available online: <https://www.fda.gov/media/83904/download> (accessed on 23 March 2021).
118. Food and Drug Administration. *Analytical Procedures and Methods Validation for Drugs and Biologics*; US Department of Health and Human Services: Washington, DC, USA, 2015.
119. Food and Drug Administration. *Reporting of Computational Modeling Studies in Medical Device Submissions—Guidance for Industry and Food and Drug Administration Staff*; Food and Drug Administration: White Oak, MD, USA, 2016.
120. Food and Drug Administration. *General Principles of Software Validation: Final Guidance for Industry and FDA Staff*; Tech. Rep.; Center for Devices and Radiological Health: Washington, DC, USA, 2002.
121. Blaschke, T.F.; Lumpkin, M.; Hartman, D. The World Health Organization prequalification program and clinical pharmacology in 2030. *Clin. Pharmacol. Ther.* **2020**, *107*, 68. [[CrossRef](#)]