New BIOTECHNOLOGY 74 (2023) 16–24

Contents lists available at ScienceDirect

New BIOTECHNOLOGY

journal homepage: www.elsevier.com/locate/nbt

AI for life: Trends in artificial intelligence for biotechnology

Andreas Holzinger [a,b,d,*], Katharina Keiblinger [a], Petr Holub [c], Kurt Zatloukal [b], Heimo Müller [b]

[a] University of Natural Resources and Life Sciences Vienna, Austria
[b] Medical University Graz, Austria
[c] Masaryk University Brno, Czech Republic
[d] Alberta Machine Intelligence Institute Edmonton, Canada

ARTICLE INFO

Keywords:
Artificial Intelligence
Biotechnology
Deep Learning
Digital Transformation
Machine Learning

ABSTRACT

Due to popular successes (e.g., ChatGPT) Artificial Intelligence (AI) is on everyone's lips today. When advances in biotechnology are combined with advances in AI unprecedented new potential solutions become available. This can help with many global problems and contribute to important Sustainability Development Goals. Current examples include Food Security, Health and Well-being, Clean Water, Clean Energy, Responsible Consumption and Production, Climate Action, Life below Water, or protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss. AI is ubiquitous in the life sciences today. Topics include a wide range from machine learning and Big Data analytics, knowledge discovery and data mining, biomedical ontologies, knowledge-based reasoning, natural language processing, decision support and reasoning under uncertainty, temporal and spatial representation and inference, and methodological aspects of explainable AI (XAI) with applications of biotechnology. In this pre-Editorial paper, we provide an overview of open research issues and challenges for each of the topics addressed in this special issue. Potential authors can directly use this as a guideline for developing their paper.

1. Introduction and motivation

Artificial intelligence (AI) is already widely used in biotechnology to solve a variety of problems. These include, for example, drug discovery [1], drug safety [2], functional and structural genomics [3],[4], proteomics [5],[6], metabolomics [7], pharmacology [8], pharmacogenetics [9], and pharmacogenomics [10], among many others [11], [12], [13]. Future advances in this domain depend critically on the ability of biotechnology researchers to use advanced AI solutions effectively. The biotechnology industry currently relies heavily on data storage, filtering, analysis and sharing. Biotechnology companies and various healthcare organizations around the world already maintain huge databases. Drug manufacturing, chemical analysis of various compounds, sequencing of RNA and DNA, enzyme studies, and other similar biological processes all require strong support from AI software solutions to move faster and reduce manual errors.

It is important to emphasize at the very beginning that all the successful AI we are describing today relies entirely on digital technology to function. Digitalization is therefore the very first step towards any AI application. In many cases, AI systems are integrated with other digital technologies such as sensors, actors (cyber-physical systems (CPS), often just called robots), and technology to enable the automation of tasks and the collection and analysis of data.

Overall, the development and use of AI is dependent on digital technology - the basis for it is digital computers. Digital transformation refers to the use of digital technologies to fundamentally change the way companies, organizations, research institutions and universities operate. In the context of biotechnology, digital transformation can involve the introduction of new technologies and processes to improve the efficiency, accuracy, and speed of research and development and enable the

List of Abbreviations: ABS, Access and Benefit Sharing; API, Application Programming Interface; AxAI, Actionable explainable Interface; AI, Artificial Intelligence; ChatGPT, Chat Generative Pretrained Transformer; CNN, Convolutional Neural Network; CRF, Conditional Random Fields; CPS, Cyber-Physical Systems; DL, Deep Learning; DSL, Domain Specific Language; GNN, Graph Neural Network; LCA, life cycle analysis; LSTM, Long Short-Term-Memory; LRP, Layer Wise Relevance Propagation; NLP, Natural Language Processing; PGM, Probabilistic Graphical Model; RL, Reinforcement Learning; TCGA, The Cancer Genome Atlas Research Network; UAS, Unmanned Aircraft System; XAI, Explainable Artificial Intelligence.

* Corresponding author at: University of Natural Resources and Life Sciences Vienna, Austria.
E-mail address: andreas.holzinger@boku.ac.at (A. Holzinger).

https://doi.org/10.1016/j.nbt.2023.02.001
Received 12 January 2023; Received in revised form 5 February 2023; Accepted 5 February 2023
Available online 6 February 2023

development of entirely new and disruptive products and services. Digital transformation can help accelerate the development and use of AI in biotechnology by providing access to big data and automating certain tasks, which can help improve the efficiency and accuracy of research and development. In this Editorial, we begin by answering the question "What is AI?" and explain the differences between AI, machine learning, and deep learning to provide a good common understanding. Subsequently, we introduce important domains of biotechnology where AI is being used or may be used in the future. Thereafter, we present some cross-cutting challenges where it is particularly important to advance future work. Finally, we provide a summary of hot topics in AI in biotechnology and end with a brief Conclusion.

2. What is AI ?

AI has a long tradition in computer science centred on the general goal of creating "intelligent" machines [14], but the term intelligence is *not* clearly defined and even measuring "intelligence" is extremely difficult [15].

The AI field was initiated in 1956 by a group of computer scientists during a workshop at Dartmouth College. The goals were extremely ambitious: *"The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves [16]".*

AI, machine learning (ML), and deep learning (DL) are all related but distinct. Here are some key differences between these fields:

- AI is an umbrella term and a broad field that refers to the creation of intelligent systems that can perform tasks that would normally require human intelligence, such as learning, problem-solving, and decision-making.
- ML is a subfield of AI that involves training digital computers to perform tasks without explicit instructions, using patterns and insights from data.
- DL is a subset of ML that uses artificial neural networks with many layers to learn and make decisions. It is particularly useful for tasks that involve analyzing large amounts of data, such as images (e.g., DALL-E2) or text (e.g., ChatGPT).

Symbolic AI is a type of AI that involves representing knowledge symbolically (Example: dog (is-a mammal) (has-property fur) (has-property four-legs)) and using logical rules to manipulate those symbols to solve problems. It is distinct from machine learning and deep learning, which do not rely on explicit rules but rather learn to recognize patterns in data. Symbolic AI is less commonly used today, as machine learning and deep learning have become more popular for tasks such as image and speech recognition. However, symbolic AI is still used in some applications, such as natural language processing and expert systems [17], [18].

AI has been an extremely broad discipline since its inception, ranging from philosophical considerations to concrete real-world applications [19]. More than any other subject, AI has experienced many ups and downs since its formal introduction as an academic discipline six decades ago. After expectations that were far too –"gh - *"within 10 years, anything humans can do, a machine will be able to do"* - the initial hype ended in a bitterly cold AI winter in the 1980 s [20]. The disappointed industry, as well as many scientists turned away from AI during this time - and even the term AI was nearly banned for a while.

The astonishing successes in statistical data-driven ML helped AI to regain (great) interest again. Two things were and still are responsible for its practical success: (1) increasingly available data sets and (2) increasingly more computational power. It was around 2010, when the family of ML algorithms called DL became particularly successful both in industry and daily life (Siri, Alexa, DeepL, etc). This triggered a second AI spring - and the best proof of what AI can do today is the newest natural language technology developed by OpenAI, called ChatGPT [21], which demonstrates what AI can do (and cannot do, because what still is missing is human common sense, in German "Hausverstand").

The grand goal of AI is to provide the theoretical fundamentals for ML to develop software that can learn autonomously from previous experience, automatically and with no human-in-the-loop [22]. Ultimately, to reach a level of *usable intelligence,* we need to (1) learn from prior data, (2) extract knowledge, (3) generalize, (4) fight the curse of dimensionality, and (5) disentangle the underlying explanatory factors of the data [23]. Machine learning is about understanding intelligence for designing and developing algorithms that can learn from data to gain knowledge from experience and improve their learning behavior over time. The challenge is to discover relevant structural and/or temporal patterns ("knowledge") in data that are often hidden in arbitrarily high-dimensional spaces and thus inaccessible to humans [24].

One grand challenge remains open: to make sense of the data in the *context* of an application domain. The quality of data and appropriate features matter most, and previous work has shown that the best-performing methods typically combine multiple low-level features with high-level context [25]. However, the full effectiveness of all AI/ML success is limited by the algorithm's inabilities to re-trace the results, to interpret and explain its results to human experts [26]. This is a big issue in the life sciences generally and specifically in biotechnology.

Whether the rapid development and proliferation of AI is a good thing or not, the fact is that AI will permeate, influence and change virtually every area of biotechnology in the future.

3. AI in biotechnology

3.1. AI in agricultural biotechnology

Biotechnology firms are now leveraging AI/ML solutions to develop autonomous robots that handle important agricultural tasks such as harvesting crops at a much faster pace than humans. Computer Vision and DL algorithms are leveraged to process and analyze the data captured by drones. This helps in monitoring crop and soil health. ML algorithms help in tracking and predicting various environmental changes including weather changes that impact the crop yield. Digital transformation is also having a strong impact on the field of smart agriculture [27]. Numerous isolated, often non-interoperable solutions exist in digital ecosystems in agriculture. This is where an "Agricultural Data Space", such as that used in the Fraunhofer lighthouse project "Cognitive Agriculture" (COGNAC) [28], can offer great added value. One application example is the evaluation of ecological and economic sustainability via the nutrient cycle. In agriculture, a balanced and appropriate nutrient cycle is at the core of efficient, productive and sustainable production of crop and livestock products. In addition to documentation, e.g., for regulatory monitoring, the focus is increasingly on optimizing the nutrient cycle. Typical examples include dairy farming with arable and grassland management. Various suppliers of sensor systems record soil, plant and weather data. On dairy farms, conclusions about nutrient inputs and outputs can be made by collecting data on feeding along with the milk yield. However, to optimize nutrient cycling, the relevant data must be complete and of sufficient quality. This is often the key problem. Important components here are interoperability, uniform ontologies, and the cognitive processing of the data. Missing data must, for example, be interpolated or modelled accordingly. By representing the nutrient cycle in the form of a digital twin, the farmer can obtain information about current nutrient balance and thus identify possible problem areas.

AI in agriculture can provide a solution for food security by adapting agricultural management to a changing climate. This includes identification of resistant crops that are more resilient to environmental changes and extremes such as drought periods. This would allow maintaining crop yields under abiotic stresses, which can severely effect crop productivity. Extreme temperatures can lower wheat yields by 6% per C [29]. As Rubisco activity and hence the photosynthetic process is sensitive to high temperature and stops at temperatures above 35 °C [30]. Stress physiology of crops effected by water and nutrient limitation can be directly addressed by advances in remote sensing beyond solutions that are already used in agriculture for biomass estimation, crop type classification, and mapping of soil characteristics [31]. The use of AI together with low-cost multi-channel sensors as well as remote sensing to collect big data requires infrastructure with data security [32].

These technologies can also be applied to identify new crop phenotypes that are more efficient in resource use and resistant to highly variable climate conditions. Phenotyping has become a key discipline of plant sciences in the last decade [33]. Sensor developments and "big data" technologies have driven phenotyping as a main field of application in breeding to meet better the challenges of global change with increasing pressure from abiotic and biotic stresses [34].

As change in climate goes along with increased pressure of pests and disease, it is crucial that plant breeding programs also cover natural defence mechanisms of crops including wild relatives which often show a more diverse rhizomicrobiome and root exudates. Wild crop relatives should be considered in breeding programs, as they have a proven pool of genetic diversity to enhance adaptive capacity of agricultural systems to limiting resource availability and disease pressures. While climate change and severe events such as droughts and heavy precipitation causes plant stress with serious consequences for food production (yield potential and stability) and in particular also decreases plant fitness, identification of efficient natural defence mechanisms are thus highly needed. However, the stress from low doses of herbicide for example, however can also have a positive effect on plant fitness through hormesis [35]. New imaging technologies combined with AI [36] are at the nexus to process big data and support data interpretation for climate change effects. Besides image-based phenotyping in plant breeding, imaging technologies have also been proposed to identify the effect of biostimulants which are of increasing interest to enhance plant fitness [37].

Combining data from image-based phenotyping with information on stress responses on a molecular level, such as genomic variation, gene and protein expression, and metabolite biosynthesis from stressors and their doses to develop hormesis management protocols [38] could be highly rewarding to align different types of information sources.

The potential of big data leverage has increased the significance of AI in agricultural disciplines. Data availability increases exponentially with technologies that are already widely applied in agriculture, from proximal to remote sensing, such as image-based phenotyping platforms in greenhouses, unmanned aircraft systems (UAS) at the field scale and satellite based remote sensing up to the landscape and global scale. To make use of the information, advances in computer vision algorithms are critical, while offering unprecedented opportunities to advance our system understanding [39]. For example, the data produced from UAS with appropriate sensors are of high accuracy and predictability for management decision, and generally more objective than traditional visual scoring methods or point sensors that show limited samples sizes or do not allow detecting in-field variations. The data generation is cost-efficient and hence can deliver not only fine spatial but also high temporal resolution on, e.g., plant growth [39]. This high-throughput technology and its benefits have been combined with other technologies such as plant phenotyping in the field, yield and biomass estimation, crop disease monitoring, water stress of crops, as well as weed mapping. Crop characteristics such as vegetation indices, canopy height and cover, or selection of genomes are also possible at high temporal resolution [39]. The huge benefit of UAS based phenotyping is its combination with AI, as the huge datasets created increase the performance and robustness of models via better training conditions. For the application of the genomics tools available to improve crops, precise phenotypic measurements at the field level are required.

The application of molecular biology tools, in particular the genetic manipulation of DNA to improve animal or plant traits, is termed molecular breeding. The tools include molecular marker assisted or genomic selection, and gene manipulation or genetic engineering [40]. Plant tissue culture technique is a useful technique for plant propagation at a commercial scale. In recent years it has been applied for (1) rapid production of plants regardless of season, (2) production of (heat, drought, salinity) tolerant varieties, (3) disease resistant plants (4) for conservation of endangered species, and (5) genetic transformation among others. These applications aim to improve crop performance in agriculture and consist of a wide array of techniques, such as, tissues, organs and in-vitro regeneration of plants, aseptic growth of cells, molecular genetics, genome analyses, transfer of genes, and recombinant technologies. Among the benefits of these technologies used in agriculture are, genetic preservation of desirable genetics, uniform plant growth, genetic enhancement for increased plant efficiency, and year-round production regardless of season.

The combined use of tissue culture with AI and other optimization algorithms has been demonstrated as a field of technology for optimization of production efficiency [41]. Plant tissue culture, which is based on "totipotenti", the ability of a stem cell to differentiate into all kind of cell types, provides the basis for "micropropagation" [41]. In this process plants are grown in vessels containing culture media from different explants. This in vitro culture contains nutrients and growth regulators. The smaller growth of plants in vitro compared to in vivo, leads to the term "micropropagation". *In vitro* culture is among the most significant technologies for propagation and breeding of various crop species, enabling different methods such as shoot multiplication, or production of plants from cells and tissues via the development of somatic embryos or adventitious shoots. The particular nutrient demands of different plant cells and tissues differ with plant species and hence the improvement of culture media is a time-consuming process that needs an enormous number of media compositions. In this case AI models are highly beneficial in overcoming the problem of complex interactions with multiple factors of in vitro culture which cannot be solved with an unrealistic quantity of treatments and traditional statistics. These AI models can simulate and forecast development and growth of plant tissue in vitro under various conditions to optimize the media, with realistic number of treatments [41]. The potential of AI models to cope with different in vitro systems and results of the cultures have turned them into a common method applied among plant tissue scientists [41].

Pest and disease control, via intelligent and precise application of plant protection agents, is expected to enhance sustainability in conventional agricultural systems. Imaging of plant disorders in real time, and the use of UAS and AI, allow a targeted and automated spraying of pesticides and fertilizers in high precision, and reduced risks of contamination for crops, animals, humans and other environmental resources (i.e., water bodies). Beside identification of disorders of plants/crops, targeting their maturity and timing of harvest is also a field of computer vision and AI in precision agriculture. Already this technology is at a stage, where computer vision models can outperform human observation accuracy [42]. After harvest, AI assisted computer vision with imaging algorithms can be used for sorting and grading of agricultural products, by identifying disease and defects, or products that do not meet the demands regarding size, shape and colour [43]. In addition to external values, as well as the internal ones, the quality of the products can be assessed with AI machine learning and image data [44]. By automating this process, AI can overcome labour shortages that are already being faced in many industrialized countries.

Coming back to the ground, AI can also be of excellent use in soil health monitoring. On the one hand, computer vision is able to characterize efficiently soil organic matter and texture at large spatial scales, providing a source of information to improve existing soil maps. The identification of soil health parameters is usually a time and cost demanding effort. Therefore, identifying key soil variables underlying soil health functions for larger scale monitoring is of high importance to define effective monitoring schemes. Laboratory data as well as low cost handheld devices can be used to train algorithms to identify soil health or potentials for improvement.

A digital hub could focus on developing tools and algorithms to better measure and process data characterizing agricultural ecosystem complexity based on the interaction of microorganisms, plants and animals and feedback to humans and their health. The focus is resilient biodiversity in production systems. In particular the role of fungi and bacteria in soils frames the design and the bandwidth of driving processes of the system. High resolution data sets with biostatistical-mathematical modelling are the tools to identify key species for dynamic buffering or stabilising ecosystems.

Soil health maps for different land use areas are a key challenge in advancing towards sustainable management, as they require precise spatial data acquisition and thus improved usage of remote sensing information beyond the current state of the art. Beyond novel remote and proximal sensors with higher spatial resolution and extended spectral bands, better learning algorithms to infer on belowground processes from data fusion are highly needed. Multimodal data processing to merge soil biodiversity data from the increasing databases with other types of environmental information (e.g., soil maps, meteorological data) can advance our understanding and monitoring capabilities of soil health and underlying natural and human drivers. Implementing these data in the evaluation of soil health and ecosystem service measures could provide a novel approach for environmental risk assessment in terms of biodiversity and losses in ecosystem functions in agricultural systems.

Digitalization in agriculture can improve data collection and recording on the *status quo* of soil health and future application of regenerative agricultural practices. Soil attributes and functions, such as soil organic matter, pore volume, aggregate stability, water holding capacity, activity of microorganisms, and nutrient (mainly nitrogen) availability, have positive impacts on the environment and crop yield. These soil functions are strongly affected by soil management practices (e.g., tillage, mechanical weed control) and crop rotations (i.e. legumes, grass-legume, field grass). Positive effects of crops on soils by providing nutrients and as a source of organic matter are often complex and not considered due to a lack of immediate responses. Nevertheless, sustainably managed soils can play an important role in climate change mitigation by storing stable soil organic carbon and reducing greenhouse gas emissions to the atmosphere [45]. This has caused a rethinking of agricultural practices, as soil fertility and soil health is the most important resource. Soil health monitoring assessment, using soil microbiome-based diagnostics and AI models proved most accurate when trained with the highest taxonomic resolution [46]. This is becoming increasingly important to observe impacts of management and to ensure improvements in crop productivity and sustainable agricultural systems. However, there is still a lack for a universal parameters or simple measures that enable high throughput analyses for soil quality or health [47]. Soil health or quality requires optimal physicochemical characteristics, which are less sensitive to soil degradation; and soil biological characteristics which respond quickly, but often are also sensitive to seasonal changes. Soil microbiotas produce enzymes and actively contribute to the formation of biophysical structures that are essential for soil functions, and thereby drive "quality", "health" and "fertility" of soils. This integral information from soil microbial data can be a strategy identifying an integrated measure for soil health, via machine learning algorithms, that is currently lacking. In particular, it is crucial to also capture sites specific factors that often limit current model predictions, and to predict soil health for a wide scale. For a robust and universally applicable soil health indices a collaborative effort among environmental, biological and computer science disciplines is needed for sustainable ecosystems and agricultural management practices.

Nowadays awareness of the factors that influence human health is increasing and the concept termed "one health" considers the environment of a human being and the functioning of the surrounding ecosystems as a pre-requisite for healthy communities. These ecosystem functions include the provision of fresh water, clean air, food security and medicine [48]. Extinction of species above and belowground is a fundamental problem, as our soils are a potential source for new antibiotics in human medical applications [49]. Hence, restoring and sustaining healthy soils and their biodiversity may have numerous benefits inducing preserving future opportunities for human health. AI can assist in identifying key drivers for ecosystem functions and how they can be regulated via measures to be implemented in land use and in particular agricultural systems. The loss of key taxa is one concern, but AI is also able to identify the complex interplay of the food-web belowground, as environmental disturbances, not only cause a reduction in the abundance of some organisms this has also consequences for others, as there are multiple interactions within food webs. A reservoir of genetic resources for crops, livestock and soil biota is only provided in biodiverse ecosystems that are fundamental for a nutritious variety, which are an essential determent for health, i.e., via micronutrient availability. There is a large use of traditional medicine by 60% of the world's population, and this originates from plant for medical use from wild populations and cultivation. Manifold communities depend on natural products collected from ecosystems not only for medical treatment, but also for cultural activities.

3.2. AI in forest biotechnology

Wood is an increasingly important resource for humanity and natural forests are of enormous ecological value. However, these slow-growing forests are unable to meet current demand, resulting in loss and degradation of forest resources. This is where forest biotechnology, especially genetic engineering, can help. This is very important because plantation forests, for example, are urgently needed to sustainably meet global demand for wood [50]. There are many potential applications for AI, including:

- Predictive modelling: AI can be used to analyze data from satellite imagery, drone imagery, and other sources to predict the growth and yield of different species of trees in different locations. This can help to optimize the planting and management of forests for maximum productivity [51].
- Disease and pest management: AI can be used to analyze data on the presence and spread of diseases and pests in forests, as well as to predict their likely impact on the health and productivity of trees. This can help to identify areas that are at risk and to implement preventative measures to protect forests.
- Environmental monitoring: AI can be used to analyze data from sensors and other sources to monitor the health of forests and identify potential environmental impacts, e.g., wildfire [52]. This can help to identify areas that are at risk and to implement measures to protect forests, also.
- Resource management: AI can be used to optimize the use of resources, such as water and nutrients, in forests to maximize productivity and minimize waste.
- Inventory management: AI can be used to optimize the management of forests for different purposes, such as timber production, conservation, and recreation. This can involve the use of AI to analyze data on the location, age, and species of trees, as well as the availability of resources and the demand for different products and services.

3.3. AI in medical biotechnology

The European In Vitro Diagnostics Regulation (IVDR) explicitly includes software and thus AI algorithms in its requirements. This poses significant challenges for in vitro diagnostics (IVD) companies that use AI for data analysis and decision support [26]. However, if the ethical and legal issues are taken into account and addressed well, we see enormous potential for AI to revolutionize medical biotechnology by enabling the faster, more accurate and more cost-effective identification and development beyond new drugs. Some specific ways in which AI can be used in medical biotechnology include:

- Drug target identification: AI can be used to analyze data from various sources, such as genomic data and protein-protein interaction data, to identify potential therapeutic targets for the treatment of diseases. This can involve the use of machine learning algorithms to identify patterns and correlations that may not be apparent to humans.
- Drug screening: AI can be used to analyze data on the activity of potential drugs against different targets to identify those that are most likely to be effective. This can involve the use of ML algorithms to predict the likelihood of a particular drug being effective based on its characteristics and the characteristics of the target.
- Image screening: AI can be used to analyze medical images, such as CT scans and MRI images, to identify abnormalities and diagnose diseases. This can involve the use of DL algorithms to automatically segment and classify structures in medical images [53].
- Predictive modelling: AI can be used to analyze data from various sources, such as electronic health records and wearable devices, to make predictions about an individual's health. This can include the use of machine learning algorithms to predict the likelihood of an individual developing a particular disease or the likelihood of a particular treatment being effective.

3.4. AI in Animal biotechnology

Livestock production systems have intensified in recent years in terms of productivity per animal or unit of land or labour [54]. This intensification of various livestock farming practices has also led to social concerns in terms of consumer acceptance, concerning food and nutrition security, food safety, sustainability, animal welfare, animal health and human health [55]. Livestock is globally the largest user of land resources, as almost 80% of the total agricultural land is dedicated to the production of feed and pasture, despite the large proportion of land, as well as the energy (and water), used to produce animal protein, by converting crops which are dedicated to feed livestock by 50–60%, for feed production on declining arable land per person [56]. Along the whole value chain, livestock related direct and indirect emissions are 16.5% of the total, or 8.1 GtCO2 equivalents according to the FAO. In addition, climate change, with more severe extreme weather events such as droughts and heavy precipitations, may result in adverse effects on animal health and welfare, as well as increased GHG emissions, reduced quality and quantity of feed and food, and hence human health. To tackle these challenges, agricultural systems must produce more with the use of fewer resources and focus on the reduction of food waste with more closed production chains. Emerging technologies such as "Precision Livestock Farming" can be beneficial for animal health and well-being together with information on environmental and economic aspects in agricultural production. Sensor-based monitoring in animal husbandry can provide optimal quantitative input data for life cycle analyses. A combination with communication and information technologies build a fundament for agriculture 4.0 [56].

Life cycle analyses (LCA) of agricultural production systems are of crucial importance considering the environment, natural resources and human health which are classified as areas of protection in life cycle impact assessment (LCIA) [57]. Product chain related life cycle analyses also include the evaluation of precision life stock farming technologies. These technologies provide an opportunity for the animal biotechnology sector, by improved performance from replacement of manual labour with intelligent labour techniques [56] enhancing production sustainability and animal welfare by reducing costs and environmental impact.

Animal farming involves tracing of animals that allows tracking the whole production process to be made available even to consumers and stakeholders via data collection and evaluation. Complete data tracking to be evaluated via LCA includes sustainable crop production for feed, tracing real-time positioning and health monitoring of animals, as well as the transportation and food processing and storage, resulting in a full tracking from farm to fork, for consumer health and safety, and could improve their awareness of behaviours. In particular, the data on environmental and health conditions will ensure animal welfare and wellbeing. Tracing (animal) products, creates big data which could facilitate more closed cycles, by reducing inputs and saving resources and costs, as well as reductions in GHG emissions (including manure management) that feed back onto the global climate [58],. considering in particular that carbon taxes would increase beef prices by more than 100% [59] without considering actual effects of the energy crisis that accelerate costs for fossil fuel and fertilizer as well.

3.5. AI in bioinformatics

While ML is already well established in medical research integrating multi omic approaches for system biology [60], there are still challenges in environmental sciences. For example the use of soil metaproteomics and the link to other omic data, or even the lack of this information, are consuming computational power and time, as the size of general databases is increasing dramatically. Here, DL algorithms may be a solution to save resources, as ML is particularly helpful for the prediction of large datasets, and human-in-the-loop can increase the explanatory power by excluding hits that are not plausible for the studied ecosystem. The combination of omics data with bioinformatics and ML will enable moving from data that are of explanatory nature to application in fields such as medicine, but also to agriculture and forestry.

One example is breeding for improved crops via soil rhizomicrobiome selection where a combined approach with AI in bioinformatics can enhance detection of genotypes with improved biotic and/or abiotic stress resistance (e.g., pest/herbivore resistance, water and nutrient use efficiency) via association with a specific rhizosphere community to promote plant growth/health and thereby reduce the input of agrochemicals. This underlines that AI provides multiple combined applications for breeding programs and are a potential biotechnology to adjust to the current production needs [61].

Another example where crop-soil microbial interactions are targeted by using large sequencing data is the effort of targeted design of microbial products, such as biostimulants, biofertilizers and biopesticides, to tackle functions such as improved nutrient uptake or plant immune system [62]. Also for revealing untargeted effects of biopesticides on changing the soil microbial community composition [63] and their functions (such as N-fixation) [64] sequencing data are a potential source for understanding the relevant changes in microbial structure to underlying observed functional effects, under a given combination of climatic, soil type and crop species.

Finally, in global change research, large datasets are crucial on global (soil) biodiversity, and the drivers for biodiversity losses and ecosystem functioning are critical to sustain stable ecosystem health. The implementation of computomics, with modern high-throughput omic measurement platforms, is fundamental to unravel environmental system understanding and discover keystone taxa that are crucial in sustaining ecosystem functions that are vital for human life and wellbeing [62].

4. Crosscutting challenges

4.1. Enabling trustworthy AI model development

Biotechnological and biomedical research has suffered from reproducibility issues over recent decades [65], [66], [67], [68], [69], [70], [71]. Studies looking systematically into the sources of these problems, such as [72] show that the problems are not localized to one specific part, but rather in various parts of the whole chains of research objects, from biological material through data generated/collected to data processing/analysis. Development of AI models is a specific type of data processing that brings particular reproducibility challenges [73].

It is broadly accepted that providing detailed documentation of research objects and their chains can substantially improve the situation with the traceability, reproducibility and trustworthiness of research results, and multiple life science and biotechnology domains are seeking to improve the current situation [74], [72], [75], [76]. This has led to the development of *provenance models* [77] which allow the documented history of research objects in a machine-readable way in distributed multi-institutional environments [78], including standard development in ISO as 23494 standard series. Once domain-specific model extensions are available, and provenance systematically generated, they can be used for programmatic assessment of data quality, i.e., fitness for specific purposes. Machine-readable provenance can also support various regulatory purposes, such as Access and Benefit Sharing for genetic resources (ABS, also known as the *Nagoya protocol*) [79], [80], or documentation for purposes of compliance with In Vitro Diagnostic Regulations (IVDR) [81] or clinical trials.

Specifically for development of AI models, the limitations stem from availability of sufficient quantities of data which are fit for purpose, i.e., of adequate quality. Availability of the data in many cases then depends on the availability of fit-for-purpose biological material. Availability of data in general has been stimulated by broad adoption of FAIR principles [82] and their domain or topical extensions, such as FAIR-Health [83] or FAIR4RS [84]. In some domains, there are focused initiatives to generate high-quality data repositories for AI purposes, such as the TCGA Research Network (https://www.cancer.gov/tcga) with over 2.5 petabytes of data publicly available for anyone in the research community to use. However, other domains still lack large reference data sets that have adequate quality for various AI-related purposes.

Another aspect specific to AI model development is availability of software and reproducibility of its runs. While availability of code has been greatly improved with adoption of public repositories such as GitHub or BitBucket and their adoption of AI frameworks[1] and models development,[2] the AI models themselves may not necessarily be trained in a reproducible way. These reproducibility concerns led recently to developing mechanisms for their deterministic operations.

4.2. Understanding AI methods

In addition to the necessary robustness, explainability is an essential factor in ensuring the trustworthiness of an AI solution [85], [86]. In the biotechnology domain explainability allows researchers, policymakers and other stakeholders to understand how the AI system makes decisions and whether those decisions are aligned with their values and goals. In the biotechnology domain, AI systems are often used to analyze large amounts of data and make predictions or recommendations that can have significant implications for public health, environmental safety, and other important areas. If an AI system is unable to provide a clear explanation for how it arrived at a particular decision or prediction, it can be difficult for people to have confidence in the accuracy and reliability of those outputs.

In addition to helping to build trust, explainability is also important to identify and address potential biases or errors in an AI system. If an AI system is making decisions that are not aligned with its intended purpose or that are harmful to certain groups of people, it is important to be able to understand how and why those decisions are being made so that appropriate corrective action can be taken.

There are several ways that an AI developer can ensure that different stakeholders can understand how a given AI model works in general and in a particular situation on specific data. Some strategies that can be effective include:

- Providing clear documentation and explanations of the AI model's architecture and training process. This can help stakeholders understand how the model was designed and how it is intended to function.
- Using interpretable models [87] or, if this is not possible, incorporating interpretability techniques. There are methods of explainable AI methods that can be used to make AI models interpretable, for an overview see [88].
- Visualizing the model's outputs and decision-making process: graphical representations [89] of the model's output or decision-making process can be helpful in explaining how it is functioning and help stakeholders understand how it arrived at a particular decision [90].
- Engaging with stakeholders and answering their questions: it is important to be responsive to stakeholders' questions and concerns and to be willing to engage with them to help them understand how the model works. This can involve providing additional explanations or conducting demonstrations of the model's functioning.

To evaluate the boundary conditions in which an AI model is known to work with a given level of statistical performance, it is important to carefully assess the model's performance under different conditions and with different data [91]. This may involve conducting experiments or simulations to test the model's performance under various scenarios or comparing its performance to other algorithms or baselines. It is also important to consider the robustness and generalizability of the model, as well as any potential biases or limitations in the data or analysis. By carefully evaluating the model's performance under different conditions, it is possible to identify the boundary conditions in which it is known to work with a given level of statistical performance.

Explainability is an essential factor in ensuring the trustworthiness and ethical use of AI in the biotechnology domain, as it allows people to understand and evaluate the decision-making processes of the AI system and to ensure that it is being used in a responsible and beneficial manner. It is important for AI developers to be proactive in communicating with stakeholders and providing them with the information and resources they need to understand how the AI model works and how it is being used. Moreover, explainability is a central factor for data quality and in the In Vitro Diagnostic Regulation (IVDR) because it allows stakeholders to understand how the data was collected and how it is being used. In the context of the IVDR, data is used to support the safety, performance, and clinical relevance of in vitro diagnostic (IVD) medical devices. If the data used to support these claims is not of high quality or if it is not used in an appropriate manner, it can have serious consequences for patient safety and the accuracy of diagnostic results [26]. Explainability is important for data quality because it allows stakeholders to understand how the data was collected and whether it is representative of the intended population. It also allows stakeholders to understand how the data is being used and whether it is being used appropriately to support the claims made about the IVD medical device.

In the IVDR, explainability is also important for ensuring that the data used to support the safety, performance, and clinical relevance of IVD medical devices is of high quality and that it is being used appropriately. The IVDR requires that the data used to support these claims be transparent, traceable, and verifiable, and explainability is a key factor in achieving these goals. By providing clear explanations of how the data was collected, how it is being used, and what claims are being made based on the data, stakeholders can have confidence in the quality and appropriateness of the data being used to support the safety, performance, and clinical relevance of IVD medical devices.

[1] https://www.tensorflow.org/api_docs/python/tf/config/experimental/enable_op_determinism

[2] https://suneeta-mall.github.io/2019/12/22/Reproducible-ML-tensorflow.html

4.3. Human-AI interfaces

Human-AI interfaces are essential for the success of any AI use in the biotechnology domain because they allow humans and AI systems to interact and collaborate in a way that is intuitive and efficient and supports explainability and causability [92]. Rapid prototyping involving all stakeholders can be enormously helpful before anything is implemented [93], [94]. The key is user-centered design (UCD) [95], which has been proven to ensure good usability and the quality of the AI solutions developed, and to ensure the satisfaction of the end users in biotechnology, who are just mostly not computer nerds. A relatively new approach is to involve humans directly in the AI pipeline, which is known as human-in-the-loop approach [96]. The human-in-the-loop (HITL) approach is a principle that involves incorporating human input and guidance into the development and operation of AI algorithms. The goal of HITL is to enable humans and AI systems to work together effectively and efficiently to achieve common goals. There are several ways in which the HITL principle can be implemented in the development of AI algorithms. For example:

- Human oversight: in this approach, humans are involved in the development and operation of the AI algorithm, providing oversight and guidance to ensure that the algorithm is functioning as intended.
- Human feedback: in this approach, humans are involved in the development and operation of the AI algorithm by providing feedback and guidance to the algorithm as it is learning and making decisions.
- Human-AI collaboration: in this approach, humans and AI systems work together as equals, collaborating to achieve common goals and using their complementary strengths and capabilities.

4.4. AI ethics, fairness and trust

Finally, there is no question that AI ethics [97], fairness [98] and trust [99] are very important research topics in the application of AI for biotechnology. Some examples of (open) research questions include:

- How can we ensure that AI systems are designed and used in ways that are ethically and socially responsible, and that respect fundamental human rights and values?
- How can we ensure that AI systems are fair, and do not perpetuate or amplify existing biases or discrimination?
- How can we ensure that AI systems are transparent and explainable, so that they can be trusted by users and stakeholders?
- How can we ensure that AI systems are secure, and do not expose individuals or organizations to risks or harms?
- How can we ensure that the development and deployment of AI systems is inclusive and involves diverse perspectives and voices?
- How can we address the ethical and societal implications of emerging technologies such as artificial general intelligence, machine learning, and autonomous systems?
- How can we develop and implement effective policies, regulations, and governance frameworks for AI?
- How can we foster dialogue and collaboration between researchers, policymakers, industry, civil society, and other stakeholders to address the ethical and societal implications of AI?
- How can we educate and raise awareness about AI ethics, fairness, and trust among the general public, as well as among those who design, develop, and use AI systems?

5. Summary of hot topics of AI in biotechnology

There are several hot topics in the field of AI and biotechnology that are currently being actively researched and are likely to continue to be areas of focus in the future:

- AI/ML and data analytics: These are becoming increasingly important in biotechnology, as they can be used to analyze large datasets and make predictions about complex biological systems. This includes the use of AI techniques to analyze genomic data, proteomic data, all sorts of *omics and many other types of biological data to better understand the underlying mechanisms of diseases and to identify potential therapeutic targets.
- Drug discovery and development: AI can be used to analyze large amounts of data to identify patterns and relationships that may not be apparent to humans. This can be used to help identify new drugs and drug targets, as well as to optimize existing therapies.
- Personalized medicine: AI can be used to analyze an individual's genomic data and other types of health data to develop personalized treatment plans that are tailored to their specific needs. This includes the use of machine learning algorithms to predict an individual's response to a particular treatment and to identify potential adverse reactions.
- Diagnostics and disease prediction: AI can be used to analyze data from various sources, such as electronic health records and wearable devices, to identify patterns and correlations that may indicate the presence of a particular disease. This can help to improve the accuracy of diagnoses and enable earlier interventions to prevent the progression of diseases.
- Biomedical image analysis: AI can be used to analyze medical images, such as CT scans and MRI images, to identify abnormalities and diagnose diseases. This includes the use of deep learning algorithms to automatically segment and classify structures in medical images.

6. Conclusion

AI is a very broad term that is used today generally and practically for everything where any Digital Information Processing System processes any data. Digitization and digital transformation are thus central to the beginning of any application of AI. It is precisely the availability of large and high-quality data volumes and the rapid increase in computing power that have been and will continue to be the decisive factors. These will continue to be the driving forces of AI in the future. We are still in the middle of this development process and there is currently no end in sight. The future will look different than the professional futurologists and the media describe it, however it is pretty certain that AI will become increasingly important in the future - whether we like it or not. We do not need to fear a second AI winter this time, because (a) the field of AI applications is now much broader; (b) in many domains AI is already very successful in everyday life (speech processing like Alexa, Siri etc., DeepL translator etc., or ChatGPT the latest system from OpenAI that impresses with amazing performance), all of these systems being also available for biotechnology; and (c) entire AI ecosystems are already being built today that are so large that they have the power to renew themselves.

Our wish for the future is AI fairness, open science, and open data - AI ecosystems for the benefit of all people on our planet.

Credit authorship contribution statement

All authors contributed to conceptualization, writing, review and editing.

Declaration of Competing Interest

The Authors declare that there are no conflicts of interests.

Acknowledgements

This work does not raise any ethical issues. Parts of this work have been funded by the Austrian Science Fund (FWF), Project: P-32554 explainable Artificial Intelligence. Parts of this work have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 874662 (HEAP), No. 101057062 (AIDAVA) and No. 101079183 (BioMedAI). This publication reflects only the authors' view, and the European Commission is not responsible for any use that may be made of the information it contains.

References

[1] David L, Thakkar A, Mercado R, Engkvist O. Molecular representations in AI-driven drug discovery: a review and practical guide. J Chemin- 2020;12(1):1–22. https://doi.org/10.1186/s13321-020-00460-5.

[2] Diaw MD, Papelier S, Durand-Salmon A, Felblinger J, Oster J. AI-assisted QT measurements for highly automated drug safety studies. IEEE Trans Biomed Eng 2022. https://doi.org/10.1109/TBME.2022.3221339.

[3] Caudai C, Galizia A, Geraci F, et al. AI applications in functional genomics. Comput Struct Biotechnol J 2021;19:5762–90. https://doi.org/10.1016/j.csbj.2021.10.009.

[4] Lin J, Ngiam KY. How data science and AI-based technologies impact genomics. Singap Med J 2023;64(1):59–66. https://doi.org/10.4103/singaporemedj.SMJ-2021-438.

[5] Xiao Q, Zhang F, Xu L, et al. High-throughput proteomics and AI for cancer biomarker discovery. Adv Drug Deliv Rev 2021;176:113844. https://doi.org/10.1016/j.addr.2021.113844.

[6] Mund A, Coscia F, Hollandi R, et al. AI-driven Deep Visual Proteomics defines cell identity and heterogeneity. BioRxiv 2021. https://doi.org/10.1101/2021.01.25.427969.

[7] Petrick LM, Shomron N. AI/ML-driven advances in untargeted metabolomics and exposomics for biomedical applications. Cell Rep Phys Sci 2022;3:7. https://doi.org/10.1016/j.xcrp.2022.100978.

[8] van der Lee M, Swen JJ. Artificial intelligence in pharmacology research and practice. Clin Transl Sci 2023;16(1):31–6. https://doi.org/10.1111/cts.13431.

[9] Roche-Lima A, Roman-Santiago A, Feliu-Maldonado R, et al. Machine learning algorithm for predicting warfarin dose in caribbean hispanics using pharmacogenetic data. Front Pharmacol 2020;10:1550. https://doi.org/10.3389/fphar.2019.01550.

[10] Lin E, Lin C-H, Lane H-Y. Precision psychiatry applications with pharmacogenomics: artificial intelligence and machine learning approaches. Int J Mol Sci 2020;21(3):969. https://doi.org/10.3390/ijms21030969.

[11] Oliveira AL. Biotechnology, big data and artificial intelligence. Biotechnol J 2019;14(8):1800613. https://doi.org/10.1002/biot.201800613.

[12] Goh WWB, Sze CC. AI paradigms for teaching biotechnology. Trends Biotechnol 2019;37(1):1–5. https://doi.org/10.1016/j.tibtech.2018.09.009.

[13] Kim H. AI, big data, and robots for the evolution of biotechnology. Genom Inform 2019;17(4):e44. https://doi.org/10.5808/GI.2019.17.4.e44.

[14] Turing AM. Computing machinery and intelligence. Mind 1950;59(236):433–60. https://doi.org/10.1093/mind/LIX.236.433.

[15] Holzinger A, Kickmeier-Rust M, Müller H. Kandinsky patterns as IQ-test for machine learning. Lecture Notes in Computer Science LNCS 11713. Cham: Springer/Nature; 2019. p. 1–14. https://doi.org/10.1007/978-3-030-29726-8-1.

[16] McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. AI Mag 2006;27(4):12–4. https://doi.org/10.1609/aimag.v27i4.1904.

[17] Bratko I, Muggleton S. Applications of inductive logic programming. Commun ACM 1995;38(11):65–70. https://doi.org/10.1145/219717.219771.

[18] Muggleton SH, Schmid U, Zeller C, Tamaddoni-Nezhad A, Besold T. Ultra-strong machine learning: comprehensibility of programs learned with ILP. Mach Learn 2018;107:1119–40. https://doi.org/10.1007/s10994-018-5707-3.

[19] Russell S.J., Norvig P. Artificial intelligence: a modern approach (4th edition). Upper Saddle River: Prentice Hall; 2020.

[20] Hendler J. Avoiding another AI winter. IEEE Intell Syst 2008;23(2):2–4. https://doi.org/10.1109/MIS.2008.20.

[21] King MR. The future of AI in medicine: a perspective from a chatbot. Ann Biomed Eng 2022:1–5. https://doi.org/10.1007/s10439-022-03121-w.

[22] Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N. Taking the human out of the loop: a review of bayesian optimization. Proc IEEE 2016;104(1):148–75. https://doi.org/10.1109/JPROC.2015.2494218.

[23] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35(8):1798–828. https://doi.org/10.1109/TPAMI.2013.50.

[24] Holzinger A. Introduction to machine learning and knowledge extraction (MAKE). Mach Learn Knowl Extr 2019;1(1):1–20. https://doi.org/10.3390/make1010001.

[25] Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR); 2014; Columbus (OH). 10.1109/CVPR.2014.81.

[26] Müller H, Holzinger A, Plass M, Brcic L, Stumptner C, Zatloukal K. Explainability and causability for artificial intelligence-supported medical image analysis in the context of the european in vitro diagnostic regulation. N Biotechnol 2022;70:67–72. https://doi.org/10.1016/j.nbt.2022.05.002.

[27] Holzinger A, Saranti A, Angerschmid A, et al. Digital transformation in smart farm and forest operations needs human-centered ai: challenges and future directions. Sensors 2022;22(8):3043. https://doi.org/10.3390/s22083043.

[28] Kalmar R, Rauch B, Dörr J, Liggesmeyer P. Agricultural data space. In: Otto B, Hompel Mt, Wrobel S, editors. Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Cham: Springer/Nature; 2022. https://doi.org/10.1007/978-3-030-93975-5_17.

[29] Naqvi RZ, Siddiqui HA, Mahmood MA, et al. Smart breeding approaches in post-genomics era for developing climate-resilient food crops. Front Plant Sci 2022:13. https://doi.org/10.3389/fpls.2022.972164.

[30] Barnabás B, Jäger K, Fehér A. The effect of drought and heat stress on reproductive processes in cereals. Plant, Cell Environ 2008;31(1):11–38. https://doi.org/10.1111/j.1365-3040.2007.01727.x.

[31] Barnes ML, Breshears DD, Law DJ, et al. Beyond greenness: detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data. PLoS One 2017;12(12):e0189539. https://doi.org/10.1371/journal.pone.0189539.

[32] Holzinger A, Weippl E, Tjoa AM, Kieseberg P. Digital transformation for sustainable development goals (SDGs) - a security, safety and privacy perspective on AI. Springer Lecture Notes in Computer Science, LNCS 12844. Cham: Springer; 2021. https://doi.org/10.1007/978-3-030-84060-0_1.

[33] Fiorani F, Schurr U. Future scenarios for plant phenotyping. Annu Rev Plant Biol 2013;64:267–91. https://doi.org/10.1146/annurev-arplant-050312-120137.

[34] Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. Science 2010;327(5967):818–22. https://doi.org/10.1126/science.1183700.

[35] Jalal A, de Oliveira Junior JC, Ribeiro JS, et al. Hormesis in plants: physiological and biochemical responses. Ecotoxicol Environ Saf 2021;207:111225. https://doi.org/10.1016/j.ecoenv.2020.111225.

[36] Holzinger A, Haibe-Kains B, Jurisica I. Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. Eur J Nucl Med Mol Imaging 2019;46(13):2722–30. https://doi.org/10.1007/s00259-019-04382-9.

[37] Rouphael Y, Spíchal L, Panzarová K, Casa R, Colla G. High-throughput plant phenotyping for developing novel biostimulants: from lab to field or from field to lab. Front Plant Sci 2018;9:1197. https://doi.org/10.3389/fpls.2018.01197.

[38] Rico-Chávez AK, Franco JA, Fernandez-Jaramillo AA, Contreras-Medina LM, Guevara-González RG, Hernandez-Escobedo Q. Machine learning for plant stress modeling: a perspective towards hormesis management. Plants 2022;11(7):970. https://doi.org/10.3390/plants11070970.

[39] Jung J, Maeda M, Chang A, Bhandari M, Ashapure A, Landivar-Bowles J. The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. Curr Opin Biotechnol 2021;70:15–22. https://doi.org/10.1016/j.copbio.2020.09.003.

[40] Ribaut J, De Vicente M, Delannay X. Molecular breeding in developing countries: challenges and perspectives. Curr Opin Plant Biol 2010;13(2):213–8. https://doi.org/10.1016/j.pbi.2009.12.011.

[41] Hesami M, Jones AMP. Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. Appl Microbiol Biotechnol 2020;104:9449–85. https://doi.org/10.1007/s00253-020-10888-2.

[42] Zhu Y, Cao Z, Lu H, Li Y, Xiao Y. In-field automatic observation of wheat heading stage using computer vision. Biosyst Eng 2016;143:28–41. https://doi.org/10.1016/j.biosystemseng.2015.12.015.

[43] Deng L, Du H, Han Z. A carrot sorting system using machine vision technique. Appl Eng Agric 2017;33(2):149–56. https://doi.org/10.13031/aea.11549.

[44] Iraji MS. Comparison between soft computing methods for tomato quality grading using machine vision. J Food Meas Charact 2019;13(1):1–15. https://doi.org/10.1007/s11694-018-9913-2.

[45] Lal R. Soil carbon sequestration to mitigate climate change. Geoderma 2004;123(1–2):1–22. https://doi.org/10.1016/j.geoderma.2004.01.032.

[46] Wilhelm RC, van Es HM, Buckley DH. Predicting measures of soil health using the microbiome and supervised machine learning. Soil Biol Biochem 2022;164:108472. https://doi.org/10.1016/j.soilbio.2021.108472.

[47] de Andrade VHGZ, Redmile-Gordon M, Barbosa BHG, Andreote FD, Roesch LFW, Pylro VS. Artificially intelligent soil quality and health indices for 'next generation'food production systems. Trends Food Sci Technol 2021;107:195–200. https://doi.org/10.1016/j.tifs.2020.10.018.

[48] Marselle MR, Hartig T, Cox DT, et al. Pathways linking biodiversity to human health: a conceptual framework. Environ Int 2021;150:106420. https://doi.org/10.1016/j.envint.2021.106420.

[49] Blum WE, Zechmeister-Boltenstern S, Keiblinger KM. Does soil contribute to the human gut microbiome? Microorganisms 2019;7(9):287. https://doi.org/10.3390/microorganisms7090287.

[50] Fenning TM, Gershenzon J. Where will the wood come from? Plantation forests and the role of biotechnology. TRENDS Biotechnol 2002;20(7):291–6. https://doi.org/10.1016/S0167-7799(02)01983-2.

[51] Nothdurft A, Gollob C, Kraßnitzer R, et al. Estimating timber volume loss due to storm damage in Carinthia, Austria, using ALS/TLS and spatial regression models. Ecol Manag 2021;502:119714. https://doi.org/10.1016/j.foreco.2021.119714.

[52] Holzinger A, Stampfer K, Nothdurft A, Gollob C, Kieseberg P. Challenges in artificial intelligence for smart forestry. Eur Res Consort Inform Math (ERCIM) N 2022;130(July):40–1. ⟨https://ercim-news.ercim.eu/en130/r-i/challenges-in-artificial-intelligence-for-smart-forestry⟩.

[53] Janisch M, Adelsmayr G, Müller H, et al. Non-contrast-enhanced CT texture analysis of primary and metastatic pancreatic ductal adenocarcinomas: value in assessment of histopathological grade and differences between primary and metastatic lesions. Abdom Radiol 2022:1–9.

[54] Thornton PK. Livestock production: recent trends, future prospects. Philos Trans R Soc B: Biol Sci 2010;365(1554):2853–67. https://doi.org/10.1098/rstb.2010.0134.

[55] Scholten MT, De Boer I, Gremmen B, Lokhorst C. Livestock farming with care: towards sustainable production of animal-source food. NJAS: Wagening J Life Sci 2013;66(1):3–5. https://doi.org/10.1016/j.njas.2013.05.009.

[56] D'Agaro E, Rosa F, Akentieva NP. New technology tools and life cycle analysis (LCA) applied to a sustainable livestock production. Eur J 2021;5(3):130–41. https://doi.org/10.2478/ebtj-2021-0022.
[57] De Vries M, de Boer IJ. Comparing environmental impacts for livestock products: a review of life cycle assessments. Livest Sci 2010;128(1–3):1–11. https://doi.org/10.1016/j.livsci.2009.11.007.
[58] Cooper J, Noon M, Jones C, Kahn E, Arbuckle P. Big data in life cycle assessment. J Ind Ecol 2013;17(6):796–9. https://doi.org/10.1111/jiec.12069.
[59] Doelman JC, Stehfest E, van Vuuren DP, et al. Afforestation for climate change mitigation: potentials, risks and trade-offs. Glob Change Biol 2020;26(3):1576–91. https://doi.org/10.1111/gcb.14887.
[60] Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. Biotechnol Adv 2021;49:107739. https://doi.org/10.1016/j.biotechadv.2021.107739.
[61] Pérez-Jaramillo JE, Carrión VJ, de Hollander M, Raaijmakers JM. The wild side of plant microbiomes. Microbiome 2018;6:1–6. https://doi.org/10.1186/s40168-018-0519-z.
[62] Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. Plant–microbiome interactions: from community assembly to plant health. Nat Rev Microbiol 2020;18(11):607–21. https://doi.org/10.1038/s41579-020-0412-1.
[63] Keiblinger KM, Fuchs S, Zechmeister-Boltenstern S, Riedel K. Soil and leaf litter metaproteomics—a brief guideline from sampling to understanding. FEMS Microbiol Ecol 2016;92:11. https://doi.org/10.1093/femsec/fiw180.
[64] Schneider M, Keiblinger KM, Paumann M, et al. Fungicide application increased copper-bioavailability and impaired nitrogen fixation through reduced root nodule formation on alfalfa. Ecotoxicology 2019;28:599–611. https://doi.org/10.1007/s10646-019-02047-9.
[65] Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. Circ Res 2015;116(1):116–26. https://doi.org/10.1161/CIRCRESAHA.114.303819.
[66] Servick K, Enserink M. The pandemic's first major research scandal erupts. Science 2020;368(6495):1041–2. https://doi.org/10.1126/science.368.6495.104.
[67] Lagoze C. Big Data, data integrity, and the fracturing of the control zone. Big Data Soc 2014;1(2):1–11. https://doi.org/10.1177/2053951714558281.
[68] Mobley A, Linder SK, Braeuer R, Ellis LM, Zwelling L. A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. PLoS One 2013;8(5):e63221. https://doi.org/10.1371/journal.pone.0063221.
[69] Morrison SJ. Time to do something about reproducibility. eLife 2014;3:e03981. https://doi.org/10.7554/eLife.03981.
[70] Byrne JA, Grima N, Capes-Davis A, Labbé C. The possibility of systematic research fraud targeting under-studied human genes: causes, consequences, and potential solutions. Biomark Insights 2019;14. https://doi.org/10.1177/1177271919829. 1177271919829162.
[71] Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 2011;10(9). https://doi.org/10.1038/nrd3439-c1. 712-712.
[72] Freedman LP, Inglese J. The increasing urgency for standards in basic biologic research. Cancer Res 2014;74(15):4024–9. https://doi.org/10.1158/0008-5472.CAN-14-0925.
[73] Gundersen O.E., Kjensmo S. State of the art: Reproducibility in artificial intelligence. Paper presented at: Proceedings of the AAAI Conference on Artificial Intelligence2018. 10.1609/aaai.v32i1.11503.
[74] Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014;383(9912):166–75. https://doi.org/10.1016/S0140-6736(13)62227-8.
[75] Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature 2012;483(7391):531–3.
[76] Landis SC, Amara SG, Asadullah K, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 2012;490(7419):187–91. https://doi.org/10.1038/nature11556.
[77] Moreau L., Freire J., Futrelle J., McGrath R.E., Myers J., Paulson P. The open provenance model: An overview. Provenance and Annotation of Data and Processes: Second International Provenance and Annotation Workshop, IPAW 2008,; 2008; Salt Lake City, UT, USA, June 17–18, 2008. Revised Selected Papers 2. 10.1007/978–3-540–89965-5_31.
[78] Wittner R, Mascia C, Gallo M, et al. Lightweight distributed provenance model for complex real–world environments. Sci Data 2022;9(1):503. https://doi.org/10.1038/s41597-022-01537-6.
[79] Martins J, Cruz D, Vasconcelos V. The Nagoya Protocol and its implications on the EU Atlantic Area countries. J Mar Sci Eng 2020;8(2):92. https://doi.org/10.3390/jmse8020092.
[80] Nijar GS, Louafi S, Welch EW. The implementation of the Nagoya ABS Protocol for the research sector: experience and challenges. Int Environ Agreem: Polit, Law Econ 2017;17(5):607–21. https://doi.org/10.1007/s10784-016-9328-7.
[81] Spitzenberger F, Patel J, Gebuhr I, Kruttwig K, Safi A, Meisel C. Laboratory-developed tests: design of a regulatory strategy in compliance with the international state-of-the-art and the regulation (EU) 2017/746 (EU IVDR [in vitro diagnostic medical device regulation]). Ther Innov Regul Sci 2022;56:47–64. https://doi.org/10.1007/s43441-021-00323-7.
[82] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3:160018. https://doi.org/10.1038/sdata.2016.18.
[83] Holub P, Kohlmayer F, Prasser F, et al. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-health. Biopreservation Biobanking 2018;16(2):97–105. https://doi.org/10.1089/bio.2017.0110.
[84] Hong NPC, Katz DS, Barker M, et al. FAIR principles for research software (FAIR4RS principles). Res Data Alliance 2022. https://doi.org/10.15497/RDA00065.
[85] Holzinger A. The next frontier: AI we can really trust. In: Kamp M, editor. Proceedings of the ECML PKDD 2021, CCIS 1524. Cham: Springer Nature; 2021. https://doi.org/10.1007/978-3-030-93736-2_33.
[86] Holzinger A, Dehmer M, Emmert-Streib F, et al. Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence. Inf Fusion 2022;79(3):263–78. https://doi.org/10.1016/j.inffus.2021.10.007.
[87] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206–15. https://doi.org/10.1038/s42256-019-0048-x.
[88] Holzinger A, Saranti A, Molnar C, Biececk P, Samek W. Explainable AI methods - a brief overview. XXAI - Lecture Notes in Artificial Intelligence LNAI 13200. Cham: Springer; 2022. https://doi.org/10.1007/978-3-031-04083-2_2.
[89] Müller H, Reihs R, Zatloukal K, Holzinger A. Analysis of biomedical data with multilevel glyphs. BMC Bioinforma 2014;15(Suppl 6). https://doi.org/10.1186/1471-2105-15-S6-S5.
[90] Hund M, Boehm D, Sturm W, et al. Visual analytics for concept exploration in subspaces of patient groups: making sense of complex datasets with the doctor-in-the-loop. Brain Inform 2016;3(4):233–47. https://doi.org/10.1007/s40708-016-0043-5.
[91] Carrington AM, Manuel DG, Fieguth PW, et al. Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. IEEE Trans Pattern Anal Mach Intell 2023;45(1):329–41. https://doi.org/10.1109/TPAMI.2022.3145392.
[92] Holzinger A, Mueller H. Toward human-AI interfaces to support explainability and causability in medical AI. IEEE COMPUTER 2021;54(10):78–86. https://doi.org/10.1109/MC.2021.3092610.
[93] Holzinger A. Rapid prototyping for a virtual medical campus interface. IEEE Softw 2004;21(1):92–9. https://doi.org/10.1109/MS.2004.1259241.
[94] Holzinger A, Kargl M, Kipperer B, Regitnig P, Plass M, Müller H. Personas for artificial intelligence (AI) an open source toolbox. IEEE Access 2022;10(23732):23747. https://doi.org/10.1109/ACCESS.2022.3154776.
[95] Hussain Z, Slany W, Holzinger A. Current state of agile user-centered design: a survey. HCI and Usability for e-Inclusion, USAB 2009, Lecture Notes in Computer Science, LNCS 5889. Berlin, Heidelberg: Springer; 2009. https://doi.org/10.1007/978-3-642-10308-7-30.
[96] Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform 2016;3(2):119–31. https://doi.org/10.1007/s40708-016-0042-6.
[97] Mueller H, Mayrhofer MT, Veen E-BV, Holzinger A. The ten commandments of ethical medical AI. IEEE COMPUTER 2021;54(7):119–23. https://doi.org/10.1109/MC.2021.3074263.
[98] Angerschmid A, Zhou J, Theuermann K, Chen F, Holzinger A. Fairness and explanation in AI-informed decision making. Mach Learn Knowl Extr 2022;4(2):556–79. https://doi.org/10.3390/make4020026.
[99] Holzinger K, Mak K, Kieseberg P, Holzinger A. Can we trust machine learning results? Artificial intelligence in safety-critical decision support. ERCIM N 2018;112(1):42–3.