Received: 21 January 2021 | Revised: 26 July 2021 | Accepted: 28 July 2021

DOI: 10.1002/bies.202100025

BioEssays

PROBLEMS & PARADIGMS

Prospects & Overviews

AI models and the future of genomic research and medicine: True sons of knowledge?

Artificial intelligence needs to be integrated with causal conceptions in biomedicine to harness its societal benefits for the field

Harald König[1] | **Daniel Frank**[2] | **Martina Baumann**[1] | **Reinhard Heil**[1]

[1] Karlsruhe Institute of Technology, Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe, Germany

[2] Chair for Ethics, Theory, and History of the Life Sciences, University of Tübingen, Tübingen, Germany

Correspondence
Harald König, Karlsruhe Institute of Technology, Institute for Technology Assessment and Systems Analysis (ITAS), PO box 3640, Karlsruhe, 76021, Germany.
Email: h.koenig@kit.edu

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: 16ITA201A

Abstract

The increasing availability of large-scale, complex data has made research into how human genomes determine physiology in health and disease, as well as its application to drug development and medicine, an attractive field for artificial intelligence (AI) approaches. Looking at recent developments, we explore how such approaches interconnect and may conflict with needs for and notions of causal knowledge in molecular genetics and genomic medicine. We provide reasons to suggest that—while capable of generating predictive knowledge at unprecedented pace and scale—if and how these approaches will be integrated with prevailing causal concepts will not only determine the future of scientific understanding and self-conceptions in these fields. But these questions will also be key to develop differentiated policies, such as for education and regulation, in order to harness societal benefits of AI for genomic research and medicine.

KEYWORDS
artificial intelligence, causality, genomic medicine, molecular genetics, policy implications, scientific understanding

INTRODUCTION

Most practical or commercial technology developments that stand for the change, promise and fears ascribed to artificial intelligence (AI), such as in computer vision, robotics or financial modeling, are based on new machine learning (ML) techniques like deep learning models in particular, that rapidly evolved in the last decade.[1–3] Current such techniques can analyze large and complex data sets based on statistical modeling, using correlative associations from observational data for predicting outcomes.[4] Deep learning has proven to be particularly powerful for flexibly deriving patterns and predictive models from such data sets and for independently optimizing models (for AI and ML, see Box 1). Furthermore, certain methods such as deep learning are "black boxes" as it is hard for humans to recapitulate how and why predictive outcomes are achieved (Box 2).

Given increasingly large and complex data sets from biomedical research (such as on genome sequences or gene expression) and clinical medical practice (including from electronic health records and biobanks), academic research institutions as well as biotech and technology companies have developed and used AI/ML in various areas.[8,9] These are mainly the prediction of pharmaceutical properties of drug targets and drug candidates,[10] pattern recognition on medical images (such as magnetic resonance imaging) or histopathological analyses for diagnosis or monitoring disease states.[9] Another important application area is the analysis of multimodal data such as from genomics and

Abbreviations: AI, artificial intelligence; DL, deep learning; CPG, Clinical Practice Guidelines; EBM, evidence-based medicine; GWAS, genome-wide association studies; ML, machine learning; RCTs, randomized controlled trials; R&D, research and development

BOX. 1
Artificial intelligence (AI) and machine learning (ML)

The terms AI and ML are often used interchangeably. However, though there appears to be no strict definition of AI, it may be best described as the (broader) vision of science and engineering to generate computers and software that can perform in ways that were thought to require human intelligence. In contrast, ML constitutes a subfield of AI—with deep learning (DL) as a subset of ML—in which software and modeling automatically improve through experience (which is not a necessary condition for AI overall).[3]

ML encompasses three major classes: supervised learning, unsupervised learning and reinforcement learning. Supervised learning aims to predict (as output) a classification or label of data points (e.g., a property of an item) by using a given set of labeled training examples (providing input features known about the item). In unsupervised learning, the aim is to learn inherent patterns within the data themselves. Reinforcement learning is based on rewarding desired and/or punishing undesired behavior of software agents (following a trial-and-error approach). The main difference between "standard" ML methods such as logistic regression or support vector machines and DL is that DL models have a higher capacity and are much more flexible (with typically millions of trainable parameters). Thus DL is very flexible in the kinds of relationships between inputs (such as genetic variants or epigenetic marks, in the case of genomics) and outputs (e.g., cell-type specific expression of protein forms) that they can model and has proven to be particularly powerful in deriving patterns and models for making predictions from large and complex data sets.[1,5,6] DL models are based on software-simulated multiple layers of artificial neurons ("deep neural networks") and can have different architectures, corresponding to different assumptions about data and different tasks. For example, convolutional neural networks can capture special spatial dependencies (e.g., to analyze medical images or patterns in biological sequences), while recurrent neuronal networks are suited to handle sequential or time-series data (such as for genomic splicing code analysis or EEG-based prediction of epileptic seizures).[5–7]

BOX. 2
Black boxes in AI and medicine

Sophisticated forms of ML are especially powerful and flexible in the kinds of relationships between inputs that they can model (Box 1). In deep learning models this typically involves automatic adjusting of millions of parameters to create a network that most accurately transforms the inputs into output predictions.[5,6] Due to this automatic adjustment or "learning" and the sheer size of the resulting networks, however, these models are "black boxes": they are hard to "explain" or to "interprete" by humans with respect to how and/or why an outcome is achieved. No human may step through the vast number of operations or non-linear associations (taking the input data and model parameters) to recapitulate the model predictions, at least not in reasonable time.[12,13]

There is an increasing number of techniques from "explainable AI" research to provide insight about the internal operation of such networks (such as automatic-rule extraction), or networks built to explain themselves. But although such methods may help in providing relevant information, it appears still unclear what the best type of explanation metric should be for different purposes, such as risk assessment and oversight by experts or regulators, or the evaluation of recommendations by health care practitioners.[12,13]

Though the black-box character of certain AI/ML system is broadly discussed as a key challenge in relation to applications in medicine, the black-box issue has not been introduced to medicine through AI/ML. The most important instrument in evidence-based medicine (EBM) for testing the efficacy and safety of drugs or treatments (and for approving them), namely randomized controlled trials (RCTs) with their underlying difference-making, probabilistic conception of causation, can usually only provide black-box causal claims. For they establish causal relationships between interventions and measured end points on patient outcomes, without providing a pathophysiological, mechanistic explanation for why the interventions worked.[14–17]

other omics fields, and their combination with clinical data, in order to generate new diagnostic and predictive models for diseases (like in cancer liquid biopsies[11]) and/or for their underlying genetic causes.[5,6]

The black-box character of some important AI/ML methods is often seen as a main challenge for their use. The de facto inability by humans to "explain" or "interprete" how these models generate predictive outcomes has widely been argued to be especially important in the medical domain, mainly based on two grounds. First, there may be high risks linked to potential flaws and biases in models and data, and a corresponding need for system verification and improvement,[18,19,12] including in systems that may constantly retrain and change over time.[20] Such risks have been broadly discussed in relation to the utility and safety of AI systems as well as to ethical and legal issues (such as non-discrimination, privacy or accountability),[12] all of which have made "interpretability" or "explainability" also an issue for regulation. The second reason often invoked is that explanations on how these systems work were needed for trust in, and adoption of AI/ML-based approaches and innovations, including by users such as physicians and patients.[18,21–23]

In this perspective, we analyze how modern AI/ML systems interrelate and may conflict with needs for and accounts of causal knowledge

in genomic research and medicine, and point out possible implications for the future of these areas. Our analysis suggests that current discussions and policy proposals too narrowly focus on the black-box issue and that its relevance for trust in and adoption of AI/ML applications is far less clear than previously proposed. Rather the future development and possible societal benefits will be determined by the extent to which knowledge from AI/ML models is perceived to need experimental verification and on whether such verification is possible.

CAUSAL REASONING AND PREDICTION IN BIOMEDICAL RESEARCH AND MEDICINE

Causal reasoning and causal accounts

While there are numerous ideas and theories throughout philosophy about what causality actually is and what role it plays in or to explain reality (or the physical world),[24–26] causal cognition processes appear to be evolutionary entrenched in how we think and act. Thus we unconsciously strive to learn about causal relations in our environment and we constantly use causal beliefs or knowledge to draw inferences or make predictions through causal reasoning.[27] Causal reasoning, and in particular enhanced grades of causal cognition that have occurred later in human evolution, appear to be also important in contributing to the development of technological innovations (including first "complex" technologies such as bow hunting or poisoned arrows).[28,29] Such enhanced grades of causal cognition seem to involve abstract causal understanding, integrating difference-making information from various sources (e.g., one's own interventions or interventions by others). This allows to imagine and hypothesize causal networks and their outcomes under varying circumstances.[28,30]

Given this strong evolutionary foundation of causal cognition processes it may not be surprising that causation is key to various models of explanation and associated conceptions of understanding in many special sciences,[31] including areas of modern biology, such as molecular biology, physiology or evolutionary and developmental biology, and (bio-)medicine.[32,33] In contrast, the place of causality in physics, and in fundamental physics in particular, has been controversially discussed.[31,34] Key issues include the nature of laws and time as well as how to reconcile the central role of causal concepts in the special sciences, and to identify effective strategies in practice, with the often supposed absence of causation in fundamental physical laws.[34]

In molecular biology and biomedicine two different types of causal accounts are common: causal-mechanistic and interventionist conceptions. Causal-mechanistic conceptions provide scientific explanation by revealing the causal network of processes and interactions that lead to the event to be explained, exploiting experimental interventions.[35] Against this, in interventionist conceptions the goal is to observe whether an action or a treatment causes an effect, without necessarily making assumptions on or looking at causal mechanisms.[36] Causal mechanistic accounts of understanding prevail in basic molecular biology and biomedical research[32] and provide insight into molecular and physiological mechanisms (e.g., linking genetic variants to pathophysiological changes in human cells or animal models).[37–39] Interventionist conceptions have become key in evidence-based medicine (EBM), for example, to judge the efficacy of treatments for a disease on patient outcomes, using randomized controlled trials (RCTs) as its most important tool.[14,15] Under both conceptions, causal claims can involve counterfactual dependencies and reasoning (i.e., allowing to answer "what-if-things-had-been-different" questions),[31,40] with counterfactuals (such as using control and treatment groups in RCTs) having become especially important to causal inference in EBM.[41]

Both conceptions on causation, but In particular the quest for causal understanding derived from specific experimental interventions in physiological processes, may exemplify a key notion underlying the transition to modern science involving experimentation in the seventeenth century, as famously called for by Francis Bacon: "to seek, not pretty and probable conjectures, but certain and demonstrable knowledge"—as "true sons of knowledge"[42] (Figure 1).

Causality, associative models and prediction

Causal concepts and knowledge derived from them can thus either establish and/or mechanistically explain (in retrospect) why an outcome occurred or, directed into the future, enable predictions on outcomes. To predict future outcomes is also the aim of predictive models, as in the ML field. However, these models are usually based on statistically significant, but not necessarily causal, associations in the data and thus not on knowledge about what makes outcomes happen.[4,44]

Correspondingly, causal concepts are important to basic biomedical research to uncover and understand physiological pathways, or to prove hypotheses on them.[38,39,45] Moreover, causality is critical to weigh interventions and their (observed or putative) effects in drug development.[15,46,47] In contrast, associative model approaches are often used in clinical practice to provide risk estimates, for example, to predict whether patients are at high risk for a disease or to inform prognoses.[48,49] Furthermore, in basic science such predictive models may help to hypothesize likely causes or physiological mechanisms by analyzing datasets containing complex patterns[5,44] (see also below).

CURRENT AI-BASED METHODS CONFLICT WITH PREVAILING CAUSAL ACCOUNTS IN MOLECULAR GENETICS

Though causality and causal modeling have become an active research field in AI/ML,[50,51] currently established ML methods for analyzing large and complex data are based on statistical modeling.[4,51] These methods do not reflect genuine causal properties of the variables they analyze or reconstruct. Instead correlative associations from observational data are used for predictive modeling of outcomes,[4,51] such as functional consequences of genetic variants, cancer diagnoses or properties of drug candidates.[5,6,52]

Thus, Bacon's call for "certain and demonstrable knowledge" and experiments[42,43] (see also Figure 1) as well as the strong

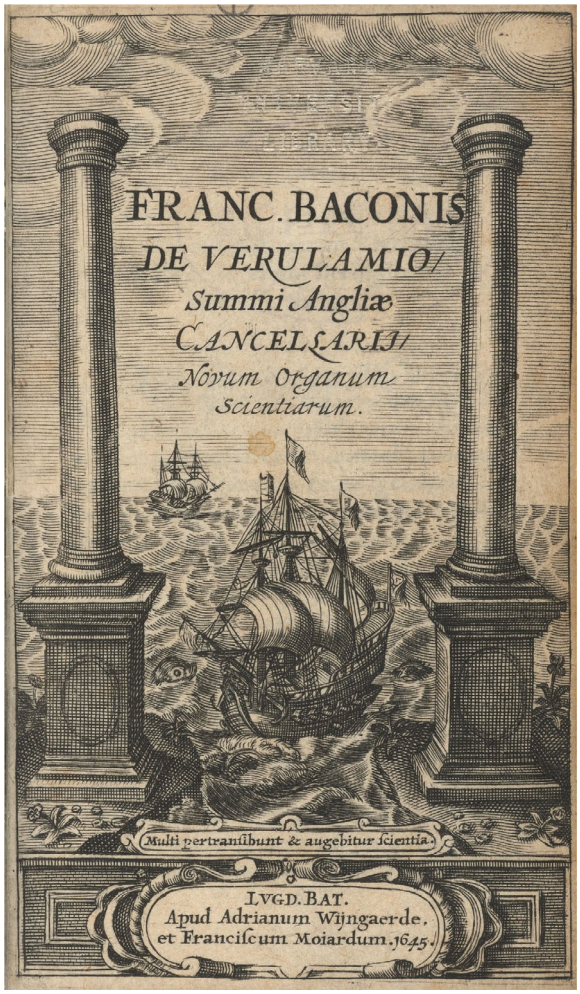

FIGURE 1 Title page for Novum organum scientiarum (second edition, 1645) by the English philosopher and statesman Francis Bacon who is often considered as one of the founders of modern science. In this and other work he outlines what he believed is needed to reveal and understand nature: to not only compile as many documented facts from literature and systematic observations as possible but, as a main element, to retrieve new knowledge from nature by experiments. Intervention in nature by experiments would reveal the secrets of nature better than observing how they "do in their usual course"[43]. The title page shows a ship passing through the Pillars of Hercules, which symbolized the end of the known world. Bacon repeatedly used this motif, drawing analogies between the exploration voyages and a need to go beyond the boundaries of traditional knowledge. Image: EC.B1328.620ib, Houghton Library, Harvard University

foundation of experimentation-based conceptions of causation in molecular genetics and biomedical research appear to be in stark contrast to present-day ML methods.

Deep learning models to predict causal genetic variants for complex traits and diseases

However, especially deep learning approaches have been suggested to be able to reveal genomic differences, such as inherited genetic variants, that are causally linked to traits or diseases. For these approaches can integrate big biological data sets to predict changes entailed by genomic differences in the complex cellular processes ("intermediate phenotypes") that lie in between genotype and phenotype. This feature is especially important for predicting effects of variants in non-coding (often regulatory) genome regions, which represent most genetic variants linked to common, multigenic diseases.[5,6] Causal genetic variants influence a molecular or cellular process to affect a disease, as opposed to variants that may be only statistically associated with a disease via genome-wide association studies (GWAS).[53] In fact various recent studies suggest that deep learning models can rapidly predict cell-type specific intermediate phenotypes, such as changes in transcription or pre-mRNA splicing, for any DNA sequence difference (including all variants linked to traits by GWAS). These models can so pinpoint putative causal variants for various conditions, like cardiological, neurological and immune-related disorders.[54–58] In some of the studies, intermediate phenotypes were further computationally integrated in multilevel models with gene and protein interaction networks of physiological processes in cells or organs, pointing to possible mechanistic pathways from genetic alterations to disease.[55,58]

Predicted causal variants and the difficulty of experimental proof

Yet despite efforts to make such models interpretable (e.g., allowing conclusions on how certain inputs may be linked to outputs),[58] rigorous proof that a genetic variant or a pathway is in fact "causal" has widely been suggested to require functional verification by experimental interventions. Such experiments would in particular include introducing corresponding variants or inactivating target genes in human cell-based models or in model organisms relevant for a disease.[38,39,59] This kind of proof is hard to achieve though, especially when it comes to the myriad of common variants that have been associated with various common disorders by GWAS.[53,60] In addition to the minor effect sizes these variants are often supposed to have individually, the sheer number of variants makes such functional verification challenging.[38,59] In keeping with these challenges, recent studies using deep learning to computationally predict putative causal variants or mechanistic pathways included no functional experiments.[54,58] Or, they did not link variants and their predicted effects directly back to the disease phenotype by introducing or reverting variants in the genome of disease-relevant cell or animal models. Rather, the forecasted functional effects (such as on direction and degree of changes in gene expression) were confirmed by artificial reporter gene constructs introduced to cell lines.[55–57] Furthermore, in all studies, the value of the models to pinpoint causal variants was supported by showing that regulatory changes predicted for curated pathogenic variants or variants found in patients differed significantly from changes forecasted for variants in unaffected people.

Thus, the deep learning approaches appear to not directly provide "rigorous" causal knowledge, but rather point to putative causal relations that may be further tested by experimental interventions.

THE UNCERTAIN FUTURE OF EXPERIMENTALLY VALIDATED KNOWLEDGE

If and to what extent such rigorous experimental verification can play a role to prove computationally predicted causality of genetic variants in complex diseases seems uncertain, however.

There are mainly two reasons for this uncertainty. First, the number of variants with small effect sizes will further increase by ever larger GWAS,[60,61] likely paralleled by an increase in computationally predicted causal variants. Second, the complex genetic, physiological or developmental processes that generate phenotypes and common diseases are highly dynamic and driven by regulatory feedback and hierarchical interactions (including cell- and tissue-level interactions or environmental cues).[62,63] They may therefore be better represented and understandable by systems genetics and network approaches.[61,63,64] Based on multilevel and integrative modeling such approaches try to analyze biological systems as a whole, focusing on the relevant interactions in networks of genes or proteins that occur in cell types or tissues.[63,64] However it remains unclear which experimental and analytical approaches may allow to more fully recapitulate and validate true network behavior across time.[65] Due to the limited accessibility of most living human tissues to direct experimental assays, these networks often need to be inferred from large omics data by ML and statistical methods.[64]

A future, in which "rigorous" proof of causal relations between genotype and (disease) phenotype by experimental intervention will prevail, should thus not be taken for granted. Instead, linking AI-predicted candidates for causal variants and pathways to disease by integrative computational models involving tissue specific gene and protein networks[55,58,66]—which themselves may have to largely be inferred in silico—could become more widespread.

Distinct implications for research, translation and clinical practice

Such a potential shift toward scientific explanation derived from AI predictions would challenge the (self-)conceptions of scientific understanding and "quality" of knowledge based on experimental interventions, that still appear to prevail in important areas of basic biomedical research (such as molecular genetics or physiology).[37,38,59] However, new techniques and approaches may be used to combine and integrate rigorous molecular interventions with dynamic network models toward exploring and distinguishing between possible causal mechanisms, to understand how and why a process occurs in a certain manner over time.[45,62] Thus approaches that combine human GWAS data with experimentation-derived tissue- and cell-type-specific networks from suitable model organisms (such as *Caenorhabditiselegans* or mouse) may help to experimentally test actionable network elements and look more at systems behavior.[66,67] Furthermore, genome-wide and combinatorial functional screening by CRISPR/Cas-mediated methods in tissue cultures derived from human induced pluripotent stem cells (hiPSCs) may contribute to probe disease-relevant network models.[38,59,68,69]

Likewise such a shift would raise questions about implications for translation of knowledge for drug development. Thus retrospective studies on drug approvals suggest that genetic support linking drug targets to disease significantly increases the likelihood for successful drug development. This appears to be in particular true if there is clear causal genetic evidence (e.g., when causal genes were identified in severe genetic disorders, as opposed to mere statistical associations of common genetic variants by GWAS).[47,70] Similarly, such genetic evidence for effects of variants on phenotypes in tissues or organs can be used to predict safety issues linked to drug targets.[71] Does this mean that less "well-founded" causal knowledge derived from AI/ML-based approaches would impair drug development? Not necessarily, for such knowledge might still help solving the pharmaceutical industry's research and development (R&D) productivity challenge: to increase the number and quality of cost-effective new drugs, without incurring unsustainable R&D costs.[72] Computational models to (more) rapidly pinpoint "reasonably good bets" (e.g., putative causal variants) for drug targets may be combined with AI-based, automated approaches for identifying, designing, synthesizing or repurposing drug candidates in shorter time,[10,73,74] and with more relevant target validation by new cellular or animal disease models (i.e., models with higher predictive validity).[75] Combining these approaches might increase (overall) quality in selecting promising targets and shift project closures to early stages, as well as reduce development cycle times and cost. All these factors have been linked to enhanced R&D productivity.[72,75,76]

Possible effects by AI/ML-based approaches on the quality of causal knowledge about genetic variants and mechanisms may even be less clear when it comes to clinical trials and clinical medical practice (i.e., to infer diagnoses or to reach decisions on treatments). This is because the role and value of basic science and mechanistic knowledge in these areas—and especially related to EBM, which has become their dominant concept—is contentious among practitioners and in philosophy of medicine.[14,77–79] In EBM, RCTs and systematic reviews of such studies are widely seen as the "gold standard" for judging diagnostic tests and/or treatments, and making recommendations on them. Proving causality in RCTs relies on showing that a treatment makes a difference for the probability of patient outcomes. Mechanistic reasoning, that is, inferring patient outcomes following interventions in the pathophysiological mechanisms, is generally ranked as evidence of low quality by EBM proponents.[77] This has been ascribed to the challenges and failures in making such inferences due to confounding factors (like interactions between various mechanisms) linked to the complexity of common diseases.[78] However, mechanistic knowledge and reasoning may play a role in interpreting trial results, for instance to successfully transfer recommendations from the test population to a different (target) population.[46,77,78]

Similarly, the role of basic science and mechanistic knowledge for diagnosis is far from clear. Expert clinicians appear to rarely use basic science or causal pathophysiological knowledge, which rather gets "encapsulated" in diagnostic labels or high-level, simplified models.[80,81] Yet mobilization of such knowledge can become beneficial when cases are rare or complex.[80]

DIVISIONS IN CAUSE AND EXPLANATION: BLACK-BOX CAUSAL CLAIMS VERSUS EXPLAINING WHY AN EFFECT OCCURS

The apparent divide in the importance of causal knowledge on physiological mechanisms between biomedical research and clinical trials and medical practice is not created because causal claims do not play a role in inferring diagnoses or in evaluating the efficacy of treatments by EBM. But rather because the foundations and purposes of causal claims in these application areas are usually different from the mechanistic theories of causality in biomedical research.

Making use of RCTs to assess and make recommendations on diagnostic methods and therapies, in particular, appears to relate to difference making probabilistic conceptions of causation. Under these conceptions, causation requires that a cause (e.g., a drug) makes a difference for the probability of its effect (patient outcome).[14,16] Causation may be seen as a form of correlation after all, under conditions where, ideally, all biases or confounding factors are controlled.[14,46] In RCTs randomization is used to control for putative other difference-making, confounding factors between treatment and control groups (such as age or comorbidities) that may affect the probability for a given outcome. Since they provide no mechanistic rationale for why an intervention entails a certain outcome, difference-making probabilistic approaches like RCTs can usually only provide black-box causal claims (Box 2) about the (statistical) effectiveness of interventions in a studied population.[14,15,46]

These differences in the foundations of causal claims thus appear to be linked to two kinds of use: an inferential use to infer causal relationships between interventions and outcomes (e.g., in RCTs) or to predict effects of interventions (e.g., by transferring RCT results to new target groups), and an explanatory use to tell why an effect occurred.[16] The difference-making conception of causation in EBM is suited for inferential use, but it does not suffice for explanatory use; that is, why the effect occurs. Such explanation needs knowledge on linking causal mechanisms.[16,17]

MISSING EXPLANATIONS AND THEIR SUPPOSED ROLE FOR POLICY

Present-day ML methods share some features with EBM's difference-making probabilistic conceptions for (black-box) causal claims. For these methods also use correlative associations to predict outcomes and can show marked black-box characteristics (Box 2). However, they are not designed to exclude or control for biases or confounding factors with respect to the found associations (e.g., between certain mutations or treatments and patient outcomes) in order to make any causal claims.

When it comes to policies on applying AI/ML-based systems for medical innovations, it appears to be not this fundamental, conceptual difference regarding causality and its implications that are discussed most, though. Instead, discussions and proposals often focus on the black-box properties or opacity of certain ML techniques such as deep learning (Box 2), for they may be linked to possible flaws and biases in data and models.[18,19,12] Furthermore, the (missing) "interpretability" or "explainability" of the inner workings of these black-boxes has widely been suggested to be key for trust in, and adoption of AI/ML-based applications or innovations.[18,21–23] Though no common definitions of "interpretability" or "explainability" exist, there are two more widely accepted dimensions of these terms: transparency of models and post-hoc interpretability.[12] Transparent models convey some degree of interpretability by themselves, for example, if a model is simple enough that a human can contemplate the entire model at once (simulatability) or if one can understand how it produces any given output from its input data (algorithmic transparency). Post-hoc interpretations do typically not explicate specifically how a model works, but provide explanations by examples (such as similar training examples) or text explanations for already made predictive outcomes.[12,13] The meaning and, in turn, the usefulness of "interpretability" or explanations on how and why AI/ML systems produce the output they do will thus differ between groups of people. For instance, detailed information on inputs or algorithms may be useful for software developers and to some extent also regulators, in order to test, evaluate and/or improve models. But such information might be less understandable and meaningful to (end-)users of the systems, like clinicians or patients. For them, post-hoc explanations may be more helpful. These differences might thus also affect the perception of benefits and possible risks (such as ethical and social issues linked to hidden biases in models) and therefore trust in applications.

Medical AI regulation and explainability

Despite the suggestions that the black-box character of certain advanced AI/ML-systems may affect their assessment of safety and effectiveness, current regulatory schemes (e.g., in Europe and the USA) that cover such systems, as so called Software as a Medical Device, lack clear standards on "explainability" or "interpretability".[82] Yet under the U.S. Food and Drug Administration's medical device regulations developers should provide information such as an "explanation of how the software works",[83] and the ability of clinicians to "understand" or "independently review" the basis of recommendations is important to initially decide whether to regulate a software.[83] It appears still unclear, however, to what the mainly technical information that developers must disclose (such as "the logic or rationale used by an algorithm")[84] had to amount to in practice. This issue may be particularly relevant for modern deep learning approaches since these lack this sort of algorithmic transparency, though upcoming "explainable AI" techniques may help to provide some relevant information[12] (Box 2).

Explainability as the key for trust and adoption: "pretty conjecture" or "demonstrable knowledge"?

Similarly, understanding how and why a prediction or recommendation was made has been argued to be crucial for trust in and adoption of

AI/ML applications by stakeholders at large, and by users in the medical domain in particular.[18,12,21,85] However, there exist considerable difficulties and variance in defining "explainability" or "trust", and in empirically assessing the role of explanations for trust in AI-based systems or recommendations.[22,13,86,87] Moreover, trust in such systems and their outcomes may depend on various factors, in general [86] and in the medical domain.[88–90]

For instance, several studies on early AI-based systems suggest that users appreciate explanatory features.[91] Yet empirical evidence that such features may actually increase trust or confidence in a system's recommendations is rather limited and mixed,[92–95] and it remains unclear to what extent these mostly laboratory studies (such as on hypothetical e-commerce websites) can be conferred to "real-life" settings or the medical domain. Also, the role of explanations for trust can depend on users' prior beliefs or expectations on outcomes.[96] Finally, empirical studies on AI systems for different tasks suggest that various other factors can affect trust, such as a system's reliability, the perceived level of machine intelligence (e.g., in form of personalized outputs or responses), or questions regarding who is intended to benefit.[86]

Some of these empirical findings also appear to resonate with observations on trust of physicians and patients in diagnoses or recommendations involving such systems. Thus the belief that AI does not take into account one's unique characteristics and circumstances ("uniqueness neglect") can be an important factor that impedes trust and use of AI systems.[97] Or, when a physician uses AI, trust of patients seems to depend on the physician and its confirmation of the AI system's recommendation, rather than on explanations on the system's performance or on how it works.[97] Likewise, trust by patients in physicians in general appears to be not strongly dependent on being involved in medical decision making, but is most closely linked to the personality and behavior of physicians.[88] Lastly, given the difficulty in keeping up with the rapidly expanding breadth and depth of medical knowledge, clinical practice guidelines (CPGs) play an important role in current clinical decision making.[98,99] CPGs on how to include the outputs of specific AI-based systems in decisions might play a key role for trust and adoption by physicians, not least because of unclear liability issues linked to the use of AI systems.[100,101] Similar to the evidence on the role of explanations for trust in AI systems by physicians or patients, various studies indicate that diffusion or uptake of medical innovations is a complex social process and depends on many factors and their interrelations, including marketing, data from clinical trials or regulatory environments.[102–104]

Thus a too strong focus on "interpretability" or "explainability" at the expense of other elements appears to be wrong-headed. In addition to raising ethical questions,[17] trading "nudged" trust and acceptance based on "plausible" explanations for potentially better patient outcomes or proof by clinical trials may be myopic, since such trust could be short-lived. Further evidence on what type of information on AI applications and provisions can sustainably generate trust in and adoption of medical AI systems will be needed to inform more differentiated policies.

WHY CAUSALITY MATTERS AND WHAT MAY DETERMINE FUTURE DEVELOPMENTS

A too strong or even single focus on "interpretability" or "explainability" of AI/ML systems may also be less rewarding than widely suggested when it comes to assessing and approving the efficacy and safety of such systems, given the only correlative knowledge current ML models can provide. Similarly, to what degree ML models will transform knowledge generation in genomic and biomedical research may depend on whether such systems can once provide causal inference, and on how people in these areas will rate such causal models and their underlying assumptions.

Mind the causality gap: Explainability should not become all-important for approval of biomedical AI systems

As regards evaluation and approval, we contend that rigorous testing with respect to patient outcomes of AI-based systems for diagnosis and treatment recommendations (e.g., by RCTs or prospective cohort studies[21,52,101]) will be required as the most important element. This is because of the conceptional issue that, even if made "explainable", current high-capacity AI methods can only account for how associations (and predictions based on them) have been drawn by the software.[4,105] These statistical modeling methods and interpretations of how they work cannot provide causal inferences (e.g., that a diagnosis and cognate intervention by a treatment will be effective), at least not without existing causal knowledge or making causal assumptions ("no causes in; no causes out").[105–108] Integrating counterfactual reasoning into current ML algorithms may however improve accuracy of associative diagnosis models, especially for rare and very rare diseases.[109] Yet, ultimately, any causal assumption may need experimental control and cannot be inferred from statistical associations alone.[107]

As long as this conceptional issue persists, it appears worth to reconsider making "explainability" or "interpretability" an all-important element for the approval of AI methods. Furthermore, giving priority to "explanations" and "understanding" of such methods over potentially higher performance of certain AI systems[12] and/or the best available evidence on patient outcomes, poses ethical as well as legal issues. These include important questions as to whether patients have the right to benefit from, and doctors the duty to use the most effective diagnosis systems or treatments.[17,100]

Technological and social or psychological factors will shape AI's future in biomedical research

To what extent AI systems in biomedical research can develop a role beyond a quicker or more comprehensive means for the generation of hypotheses (such as on causal genetic variants or drug

candidates) will depend on both technological and social or psychological factors.

Various efforts aim to develop AI methods to identify causal relations from observational and interventional data, by incorporating causal and counterfactual reasoning in suitable high-capacity ML systems, such as deep learning models. Proposed solutions include combining structural causal modeling and representation learning,[51] or neural computing frameworks to infer causality from time series (i.e., grounded on the assumptions of time-order).[110,111] However, in how far and at which level (e.g., single genes or complex networks) such new methods might once "establish" causal relationships in genomics and genomic medicine remains to be seen. An important issue may be to what degree the assumptions underlying such methods can be tested by (and stand up to) experimental approaches in cells or organisms.

Furthermore, the role of AI methods in research and their relation to causal knowledge will likely not only depend on new algorithms or the kind of observational or interventional data used. But these issues may also be affected by the extent to which causal relations (such as pathophysiological mechanisms), be they "established" by future "causal AI" systems or only predicted by current associative ones, will be accepted or perceived to require experimental validation. This may be determined by entrenched and newly upcoming beliefs or thinking about scientific methods (including new AI models and their underlying causal assumptions), results and theories among groups of scientists. Work from the history and philosophy of science and from cognitive sciences suggests an important role for such often "incommensurable" kinds of perception and thinking, that underlie concepts such as "thought styles",[112] "paradigms"[113] or "habits of minds",[114] in both continuous and radical conceptual ("revolutionary") scientific change. A putative change in perception and thinking regarding causality, and thus scientific self-conceptions and understanding, may not least be driven by new people from or close to the AI field, who enter biomedical research and education. In keeping with this, recent empirical data from biomedical research indicate that such change by "outsiders" can be fostered by the premature death of eminent scientists in a research area.[115]

Obviously such potential change by "new entrants" does not mean that AI talent or people from other areas will not be needed or should not move into genomic research and biomedicine, or that biomedical researchers and clinicians should not be educated in AI. Quite to the contrary, such influx of expertise will be required to further advance the use of AI in biomedicine, and its opportunities outlined above. Yet in addition to mutual learning between scientist from both fields, young scientists at the interface of the two fields might benefit from courses with input from biomedical and AI scientists with diverse thinking as well as from other relevant fields such as philosophy of science or cognitive sciences.

CONCLUSIONS

Current associative AI models are in stark contrast to the strong foundation of intervention-based conceptions of causation in genomic research and medicine. Given this fundamental conceptual difference, present discussions and policy proposals too narrowly focus on the black-box issue. Notions of making "explainability" or "interpretability" of AI models an all-important element for their assessment or for generating sustained trust need rethinking. Rather, the future development of knowledge from AI models for genomic research and medicine, their adoption and possible societal benefits will, for one thing, depend on whether such models can develop beyond hypotheses generators and association-based prediction tools, as generating rigorous causal knowledge by experimental intervention is laborious and costly. For another, in particular as regards direct medical applications like diagnostics, respective policies demanding clinical trials will be critical. Both the development of AI-based knowledge and policies on its use will not only hinge on technological progress on causal AI models and means to test them and their underlying assumptions experimentally. But they will also be driven by how such knowledge is perceived or judged by different actors.

Given the complexity and broad implications of issues ranging from scientific understanding to adoption of innovations, as well as the current scarcity of evidence on how to best govern these issues, the development of policies for different contexts may need engagement with people and perspectives from different disciplines and societal groups. Such inclusive approaches—which should not be expected to produce simple consensus but rather learn about and recognize different needs, preferences or (scientific) thinking—may reduce the risk that education, R&D and policy schemes to govern them succumb to one-dimensional concepts. These could narrow down, rather than broaden and leverage, the potential for societal benefits from using AI to understand genome function in biomedical research and to advance genomic medicine.

ACKNOWLEDGMENT

The work was financially supported by a project grant from the Bundesministerium für Bildung und Forschung, BMBF (16ITA201A).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Harald König developed the concept for the article and drafted the manuscript. Daniel Frank contributed to refine the article concept. All authors contributed expertise and edits to the contents of this manuscript.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Harald König https://orcid.org/0000-0002-0117-0939

REFERENCES

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

2. Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., Ravi, V., & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, *194*, a105596. https://doi.org/10.1016/j.knosys.2020.105596
3. Kersting, K. (2018). Machine learning and artificial intelligence: Two fellow travelers on the quest for intelligent behavior in machines. *Frontiers in Big Data*, *1*, a6. https://doi.org/10.3389/fdata.2018.00006
4. Schölkopf, B. (2019). Causality for machine learning. arXiv preprint arXiv:1911.10500.
5. Wainberg, M., Merico, D., Delong, A., & Frey, B. J. (2018). Deep learning in biomedicine. *Nature Biotechnology*, *36*(9), 829–838. https://doi.org/10.1038/nbt.4233
6. Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics*, *51*(1), 12–18. https://doi.org/10.1038/s41588-018-0295-5
7. Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, X., Bo, Y., & Xie, Z. (2018). Deep learning and its applications in biomedicine. *Genomics, Proteomics & Bioinformatics*, *16*(1), 17–32. https://doi.org/10.1016/j.gpb.2017.07.003
8. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
9. Dias, R., & Torkamani, A. (2019). Artificial intelligence in clinical and genomic diagnostics. *Genome Medicine*, *11*, a70. https://doi.org/10.1186/s13073-019-0689-8
10. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today*, *26*(1), 80–93. https://doi.org/10.1016/j.drudis.2020.10.010
11. Im, Y., Tsui, D., Diaz Jr., L., & Wan, J. (2021). Next-generation liquid biopsies: Embracing data science in oncology. *Trends in Cancer*, *7*(4), 283–292. https://doi.org/10.1016/j.trecan.2020.11.001
12. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garciag, S., Gil-Lopeza, S., Molinag, D., Benjaminsh, R., Chatila, R., & Herrerag, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
13. Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, *16*(3), 31–57. https://doi.org/10.1145/3236386.3241340
14. Reiss, J. & Ankeny, R. A. (2016). Philosophy of medicine. The Stanford Encyclopedia of Philosophy (Summer 2016 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2016/entries/medicine/
15. Cartwright, N. (2011). A philosopher's view of the long road from RCTs to effectiveness. *Lancet*, *377*(9775), 1400–1401. https://doi.org/10.1016/S0140-6736(11)60563-1
16. Williamson, J. (2009). Probabilistic theories. In Beebee, H., Hitchcock, C. & Menzies, P. (Eds.), *The Oxford Handbook of Causation* (pp. 185–212). Oxford: Oxford University Press.
17. London, A. J. (2019). Artificial intelligence and black box medical decisions: Accuracy versus explainability. *Hastings Center Report*, *49*(1), 15–21. https://doi.org/10.1002/hast.973
18. Holzinger, A., Biemann, C., Pattichis, C. S. & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923.
19. Samek, W., Wiegand, T. & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.
20. Babic, B., Gerke, S., Evgeniou, T., & Cohen, I. G. (2019). Algorithms on regulatory lockdown in medicine. *Science*, *366*(6470), 1202–1204. https://doi.org/10.1126/science.aay9547
21. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*, a195. https://doi.org/10.1186/s12916-019-1426-2
22. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. arXiv preprint arXiv:1806.00069.
23. Ribeiro, M. T., Singh, S. & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. KDD '16: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, August 2016, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778
24. Price, H., & Corry, R. (2007). *Causation, physics, and the constitution of reality: Russell's republic revisited*. Oxford: Oxford University Press.
25. White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin*, *108*(1), 3–18. https://doi.org/10.1037/0033-2909.108.1.3
26. Cartwright, N. (2004). Causation: One word, many things. *Philosophy of Science*, *71*(5), 805–820. https://doi.org/10.1086/426771
27. Danks, D. (2009). The psychology of causal perception and reasoning. In Beebee, H. & Menzies, P. (Eds.), *The Oxford Handbook of Causation* (pp. 447–470). Oxford: Oxford University Press.
28. Gärdenfors, P., & Lombard, M. (2018). Causal cognition, force dynamics and early hunting technologies. *Frontiers in Psychology*, *9*, a87. https://doi.org/10.3389/fpsyg.2018.00087
29. Derex, M., & Boyd, R. (2015). The foundations of the human cultural niche. *Nature Communications*, *6*, a8398. https://doi.org/10.1038/ncomms9398
30. Woodward, J. (2011). A philosopher looks at tool use and causal understanding. In McCormack, T., Hoerl, C., & Butterfill, S. (Eds.), *Tool use and causal cognition*. (pp. 18–50). Oxford: Oxford University Press.
31. De Regt, H. W. (2017). *Understanding scientific understanding*. Oxford: Oxford University Press.
32. Braillard, P.-A., & Malaterre, C. (2015). Explanation in biology: An introduction. In Braillard, P.-A., & Malaterre, C. (Eds.), *Explanation in Biology—An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. (pp. 1–28). Dordrecht: Springer.
33. Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is Mayr's proximate-ultimate dichotomy still useful? *Science*, *334*(6062), 1512–1516. https://doi.org/10.1126/science.1210879
34. Blanchard, T. (2016). Physics and causation. *Philosophy Compass*, *11*(5), 256–266. https://doi.org/10.1111/phc3.12319
35. Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
36. Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
37. MacArthur, D. G., Manolio, T. A., Dimmock, D. P., Rehm, H. L., Shendure, J., Abecasis, G. R., Adams, D. R., Altman, R. B., Antonarakis, S. E., Ashley, E. A., Barrett, J. C., Biesecker, L. G., Conrad, D. F., Cooper, G. M., Cox, N. J., Daly, M. J., Gerstein, M. B., Goldstein, D. B., Hirschhorn, J. N., … Gunter, C. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature*, *508*(7497), 469–476. https://doi.org/10.1038/nature13127
38. Soldner, F., & Jaenisch, R. (2018). Stem cells, genome editing, and the path to translational medicine. *Cell*, *175*(3), 615–632. https://doi.org/10.1016/j.cell.2018.09.010
39. Chakravarti, A., Clark, A. G., & Mootha, V. K. (2013). Distilling pathophysiology from complex disease genetics. *Cell*, *155*(1), 21–26. https://doi.org/10.1016/j.cell.2013.09.001
40. Menzies, P. & Beebee, H. (2020). Counterfactual theories of causation. The Stanford Encyclopedia of Philosophy (Winter 2020 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/

41. Höfler, M. (2005). Causal inference based on counterfactuals. *BMC Medical Research Methodology*, *5*, a28. https://doi.org/10.1186/1471-2288-5-28
42. Bacon, F. (2018). Novum Organum (1620). *The new organon or true directions concerning the interpretation of nature*. Global Grey Ebooks.
43. Schwarz, A. (2012). The becoming of the experimental mode. *Scientiae Studia*, *10*(SPE), 65–83. https://doi.org/10.1590/S1678-31662012000500004
44. Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330
45. Bizzarri, M., Brash, D. E., Briscoe, J., Grieneisen, V. A., Stern, C. D., & Levin, M. (2019). A call for a better understanding of causation in cell biology. *Nature Reviews Molecular Cell Biology*, *20*(5), 261–262. https://doi.org/10.1038/s41580-019-0127-1
46. Aronson, J. K., La Caze, A., Kelly, M. P., Parkkinen, V. P., & Williamson, J. (2018). The use of mechanistic evidence in drug approval. *Journal of Evaluation in Clinical Practice*, *24*(5), 1166–1176. https://doi.org/10.1111/jep.12960
47. King, E. A., Davis, J. W., & Degner, J. F. (2019). Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genetics*, *15*(12), e1008489. https://doi.org/10.1371/journal.pgen.1008489
48. van Diepen, M., Ramspek, C. L., Jager, K. J., Zoccali, C., & Dekker, F. W. (2017). Prediction versus aetiology: Common pitfalls and how to avoid them. *Nephrology, Dialysis, Transplantation*, *32*(suppl_2), ii1–ii5. https://doi.org/10.1093/ndt/gfw459
49. Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine—Beyond the peak of inflated expectations. *The New England Journal of Medicine*, *376*(26), 2507–2509. https://doi.org/10.1056/NEJMp1702071
50. Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, *10*, a524. https://doi.org/10.3389/fgene.2019.00524
51. Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the Institute of Electrical and Electronics Engineers*, *109*(5), 612–634. https://doi.org/10.1109/JPROC.2021.3058954
52. Liu, M., Oxnard, G., Klein, E., Swanton, C., Seiden, M., & CCGA Consortium (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology*, *31*(6), 745–759. https://doi.org/10.1016/j.annonc.2020.02.011
53. Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467–484. https://doi.org/10.1038/s41576-019-0127-1
54. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J A., Arbelaez, J., Cui, W., Schwartz, G B., Chow, E D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S J., & Farh, K. K. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell*, *176*(3), 535–548.e24. https://doi.org/10.1016/j.cell.2018.12.015
55. Zhou, J., Park, C. Y., Theesfeld, C. L., Wong, A. K., Yuan, Y., Scheckel, C., Fak, J. J., Funk, J., Yao, K., Tajima, Y., Packer, A., Darnell, R. B., & Troyanskaya, O. G. (2019). Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. *Nature Genetics*, *51*(6), 973–980. https://doi.org/10.1038/s41588-019-0420-0
56. Zhou, J., Theesfeld, C. L., Yao, K., Chen, K. M., Wong, A. K., & Troyanskaya, O. G. (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature Genetics*, *50*(8), 1171–1179. https://doi.org/10.1038/s41588-018-0160-6
57. Richter, F., Morton, S. U., Kim, S. W., Kitaygorodsky, A., Wasson, L. K., Chen, K. M., Zhou, J., Qi, H., Patel, N., DePalma, S. R., Parfenov, M., Homsy, J., Gorham, J. M., Manheimer, K. B., Velinder, M., Farrell, A., Marth, G., Schadt, E. E., Kaltman, J. R., … Gelb, B. D. (2020). Genomic analyses implicate noncoding de novo variants in congenital heart disease. *Nature Genetics*, *52*(8), 769–777. https://doi.org/10.1038/s41588-020-0652-z
58. Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C., Clarke, D., Gu, M., Emani, P., Yang, Y. T., Xu, M., Gandal, M. J., Lou, S., Zhang, J., Park, J. J., Yan, C., Rhie, S. K., Manakongtreecheep, K., Zhou, H., … Gerstein, M. B. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science*, *362*(6420), eaat8464. https://doi.org/10.1126/science.aat8464
59. Musunuru, K., Bernstein, D., Cole, F. S., Khokha, M. K., Lee, F. S., Lin, S., McDonald, T. V., Moskowitz, I. P., Quertermous, T., Sankaran, V. G., Schwartz, D. A., Silverman, E K., Zhou, X., Hasan, A. A. K., & Luo, X.-J. (2018). Functional assays to screen and dissect genomic hits: Doubling down on the national investment in genomic research. *Circulation: Genomic and Precision Medicine*, *11*(4), e002178. https://doi.org/10.1161/CIRCGEN.118.002178
60. Zhang, Y., Qi, G., Park, J.-H., & Chatterjee, N. (2018). Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nature Genetics*, *50*(9), 1318–1326. https://doi.org/10.1038/s41588-018-0193-x
61. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., & Visscher, P. M. (2018). Common disease is more complex than implied by the core gene omnigenic model. *Cell*, *173*(7), 1573–1580. https://doi.org/10.1016/j.cell.2018.05.051
62. DiFrisco, J., & Jaeger, J. (2020). Genetic causation in complex regulatory systems: An integrative dynamic perspective. *Bioessays*, *42*(6), 1900226. https://doi.org/10.1002/bies.201900226
63. Civelek, M., & Lusis, A. J. (2014). Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*, *15*(1), 34–48. https://doi.org/10.1038/nrg3575
64. Yao, V., Wong, A. K., & Troyanskaya, O. G. (2018). Enabling precision medicine through integrative network models. *Journal of Molecular Biology*, *430*(18), 2913–2923. https://doi.org/10.1016/j.jmb.2018.07.004
65. Baliga, N. S., Björkegren, J. L., Boeke, J. D., Boutros, M., Crawford, N. P., Dudley, A. M., Farber, C. R., Jones, A., Levey, A. I., Lusis, A. J., Mak, H. C., Nadeau, J. H., Noyes, M. B., Petretto, E., Seyfried, N. T., Steinmetz, L. M., & Vonesch, S. C. (2017). The state of systems genetics in 2017. *Cell Systems*, *4*(1), 7–15. https://doi.org/10.1016/j.cels.2017.01.005
66. Roussarie, J.-P., Yao, V., Rodriguez-Rodriguez, P., Oughtred, R., Rust, J., Plautz, Z., Kasturia, S., Albornoz, C., Wang, W., Schmidt, E. F., Dannenfelser, R., Tadych, A., Brichta, L., Barnea-Cramer, A., Heintz, N., Hof, P. R., Heiman, M., Dolinski, K., Flajolet, M., … Greengard, P. (2020). Selective neuronal vulnerability in Alzheimer's disease: A network-based analysis. *Neuron*, *107*(5), 821–835.e12. https://doi.org/10.1016/j.neuron.2020.06.010
67. Yao, V., Kaletsky, R., Keyes, W., Mor, D. E., Wong, A. K., Sohrabi, S., Murphy, C. T., & Troyanskaya, O. G. (2018). An integrative tissue-network approach to identify and test human disease genes. *Nature Biotechnology*, *36*(11), 1091–1099. https://doi.org/10.1038/nbt.4246
68. Kampmann, M. (2018). CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. *ACS Chemical Biology*, *13*(2), 406–416. https://doi.org/10.1021/acschembio.7b00657
69. Fernando, M. B., Ahfeldt, T., & Brennand, K. J. (2020). Modeling the complex genetic architectures of brain disease. *Nature Genetics*, *52*(4), 363–369. https://doi.org/10.1038/s41588-020-0596-3
70. Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P. C., Li, M. J., Wang, J., Cardon, L. R., Whittaker, J. C., & Sanseau, P. (2015). The support of human genetic evidence for approved drug indications. *Nature Genetics*, *47*(8), 856–860. https://doi.org/10.1038/ng.3314
71. Nguyen, P. A., Born, D. A., Deaton, A. M., Nioi, P., & Ward, L. D. (2019). Phenotypes associated with genes encoding drug targets are

predictive of clinical trial side effects. *Nature Communications*, *10*, a1579. https://doi.org/10.1038/s41467-019-09407-3

72. Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., & Schacht, A. L. (2010). How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, *9*(3), 203–214. https://doi.org/10.1038/nrd3078
73. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R. R., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B. A., L. H. Lee, Soll, R., Madge, D., … Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, *37*(9), 1038–1040. https://doi.org/10.1038/s41587-019-0224-x
74. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackerman, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. *Cell*, *180*(4), 688–702.e13. https://doi.org/10.1016/j.cell.2020.01.021
75. Morgan, P., Brown, D. G., Lennard, S., Anderton, M. J., Barrett, J. C., Eriksson, U., Fidock, M., Hamrén, B., Johnson, A., March, R. E., Matcham, J., Mettetal, J., Nicholls, D. J., Platz, S., Rees, S., Snowden, M. A., & Pangalos, M. N. (2018). Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nature Reviews Drug Discovery*, *17*(3), 167–181. https://doi.org/10.1038/nrd.2017.244
76. Scannell, J. W., & Bosley, J. (2016). When quality beats quantity: Decision theory, drug discovery, and the reproducibility crisis. *Plos One*, *11*(2), e0147215. https://doi.org/10.1371/journal.pone.0147215
77. La Caze, A. (2011). The role of basic science in evidence-based medicine. *Biology & Philosophy*, *26*(1), 81–98. https://doi.org/10.1007/s10539-010-9231-5
78. Andersen, H. (2012). Mechanisms: What are they evidence for in evidence-based medicine? *Journal of Evaluation in Clinical Practice*, *18*(5), 992–999. https://doi.org/10.1111/j.1365-2753.2012.01906.x
79. Pangaro, L. (2010). The role and value of the basic sciences in medical education: The perspective of clinical education—students' progress from understanding to action. *Medical Science Educator*, *20*(3), 307–313.
80. Brush Jr, J. E., Sherbino, J., & Norman, G. R. (2017). How expert clinicians intuitively recognize a medical diagnosis. *American Journal of Medicine*, *130*(6), 629–634. https://doi.org/10.1016/j.amjmed.2017.01.045
81. Norman, G. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education*, *39*(4), 418–427. https://doi.org/10.1111/j.1365-2929.2005.02127.x
82. Ordish, J., Hannah, M. & Allison, H. (2019). Algorithms as Medical Devices. PHG Foundation. https://www.phgfoundation.org/report/algorithms-as-medical-devices
83. FDA. (2019). Software Precertification. Program: Working Model—Version 1.0: U.S. Food & Drug Administration
84. FDA. (2019). Clinical decision support software: Draft guidance for Industry and Food and Drug Administration staff: U.S. Food & Drug Administration
85. Preece, A., Harborne, D., Braines, D., Tomsett, R. & Chakraborty, S. (2018). Stakeholders in explainable AI. arXiv preprint arXiv:1810.00184.
86. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057
87. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007
88. Hall, M. A., Dugan, E., Zheng, B., & Mishra, A. K. (2001). Trust in physicians and medical institutions: What is it, can it be measured, and does it matter? *The Milbank Quarterly*, *79*(4), 613–639. https://doi.org/10.1111/1468-0009.00223
89. Calnan, M., & Rowe, R. (2006). Researching trust relations in health care. *Journal of Health Organization and Management*, *20*(5), 349–358. https://doi.org/10.1108/14777260610701759
90. Wilk, A. S., & Platt, J. E. (2016). Measuring physicians' trust: A scoping review with implications for public policy. *Social Science & Medicine*, *165*, 75–81. https://doi.org/10.1016/j.socscimed.2016.07.039
91. Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *Management Information Systems Quarterly*, *23*(4), 497–530. https://doi.org/10.2307/249487
92. Ye, L. R., & Johnson, P. E. (1995). The impact of explanation facilities on user acceptance of expert systems advice. *Management Information Systems Quarterly*, *19*(2), 157–172. https://doi.org/10.2307/249686
93. Arnold, V., Clark, N., Collier, P. A., Leech, S. A., & Sutton, S. G. (2006). The differential use and effect of knowledge-based system explanations in novice and expert judgment decisions. *Management Information Systems Quarterly*, *30*(1), 79–97. https://doi.org/10.2307/25148718
94. Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, *23*(4), 217–246. https://doi.org/10.2753/MIS0742-1222230410
95. Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, *20*(6), 542–556. https://doi.org/10.1016/j.knosys.2007.04.004
96. Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, May 2016, pages 2390–2395. https://doi.org/10.1145/2858036.2858402
97. Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013
98. Djulbegovic, B., & Guyatt, G. H. (2017). Progress in evidence-based medicine: A quarter century on. *Lancet*, *390*(10092), 415–423. https://doi.org/10.1016/S0140-6736(16)31592-6
99. IOM. (2011). Clinical practice guidelines we can trust. Washington, DC: *Institute of Medicine, Committee on Standards for Developing Trustworthy Clinical Practice Guidelines*. National Academies Press.
100. Price, W. N., Gerke, S., & Cohen, I. G. (2019). Potential liability for physicians using artificial intelligence. *JAMA*, *322*(18), 1765–1766. https://doi.org/10.1001/jama.2019.15064
101. Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., Chan, A.-W., & The SPIRIT-AI and CONSORT-AI Working Group (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nature Medicine*, *26*(9), 1364–1374. https://doi.org/10.1038/s41591-020-1034-x
102. Azoulay, P. (2002). Do pharmaceutical sales respond to scientific evidence? *Journal of Economics & Management Strategy*, *11*(4), 551–594. https://doi.org/10.1111/j.1430-9134.2002.00551.x
103. Lublóy, Á. (2014). Factors affecting the uptake of new medicines: A systematic literature review. *BMC Health Services Research*, *14*, a469. https://doi.org/10.1186/1472-6963-14-469
104. Jia, J. & Wagman, L. (2020). The one-year impact of the General Data Protection Regulation (GDPR) on European ventures: Datacatalyst Technical Report, January 2020. https://datacatalyst.org/wp-content/uploads/2020/01/GDPR-report-2020.pdf
105. Lundberg, S., Dillon, E., LaRiviere, J., Roth, J. & Syrgkanis, V. (2021). Be careful when interpreting predictive models in search of causal insights. Towards Data Science. https://towardsdatascience.com/

be-careful-when-interpreting-predictive-models-in-search-of-causal-insights-e68626e664b6
106. Cartwright, N. (1995). Précis of nature's capacities and their measurement. *Philosophy and Phenomenological Research*, *55*(1), 153–156. https://doi.org/10.2307/2108313
107. Pearl, J. (2010). An introduction to causal inference. *The International Journal of Biostatistics*, *6*(2), a7. https://doi.org/10.2202/1557-4679.1203
108. Sgaier, S. K., Huang, V., & Charles, G. (2020). The case for causal AI. *Stanford Social Innovation Review*, *Summer 2020*, 50–55.
109. Richens, J. G., Lee, C. M., & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, *11*, a3923. https://doi.org/10.1038/s41467-020-17419-7
110. Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., & Zscheischler, J. (2019). Inferring causation from time series in Earth system sciences. *Nature Communications*, *10*, a2553. https://doi.org/10.1038/s41467-019-10105-3
111. Huang, Y., Fu, Z., & Franzke, C. L. (2020). Detecting causality from time series in a machine learning framework. *Chaos*, *30*, 063116. https://doi.org/10.1063/5.0007670
112. Fleck, L. (1979). *Genesis and Development of a Scientific Fact. Originally published as Entstehung und Entwicklung einer wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv (1935), Benno Schwabe und Co., Basel. Edited by T. J. Trenn & R. K. Merton. Translated by F. Bradley & T. J. Trenn. Foreword by T. S. Kuhn.* Chicago: University of Chicago Press.
113. Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
114. Margolis, H. (1993). *Paradigms and barriers: How habits of mind govern scientific beliefs*. Chicago: University of Chicago Press.
115. Azoulay, P., Fons-Rosen, C., & Graff Zivin, J. S. (2019). Does science advance one funeral at a time? *American Economic Review*, *109*(8), 2889–2920. https://doi.org/10.1257/aer.20161574

How to cite this article: König, H., Frank, D., Baumann, M., & Heil, R. (2021). AI models and the future of genomic research and medicine: True sons of knowledge? *BioEssays*, e2100025. https://doi.org/10.1002/bies.202100025