CAMBRIDGE
UNIVERSITY PRESS

POSITION PAPER

Virtual laboratories: transforming research with AI

Arto Klami[1] (iD), Theo Damoulas[2] (iD), Ola Engkvist[3] (iD), Patrick Rinke[4] (iD) and Samuel Kaski[5,6]

[1]Finnish Center for Artificial Intelligence (FCAI) and Department of Computer Science, University of Helsinki, Helsinki, Finland
[2]Alan Turing Institute and Departments of Computer Science and Statistics, University of Warwick, Warwick, UK
[3]Molecular AI, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden
[4]Finnish Center for Artificial Intelligence (FCAI) and Department of Applied Physics, Aalto University, Espoo, Finland
[5]Finnish Center for Artificial Intelligence (FCAI) and Department of Computer Science, Aalto University, Espoo, Finland
[6]Department of Computer Science, University of Manchester, Manchester, UK
Corresponding author: Arto Klami; Email: arto.klami@helsinki.fi

Abstract

New scientific knowledge is needed more urgently than ever, to address global challenges such as climate change, sustainability, health, and societal well-being. Could artificial intelligence (AI) accelerate science to meet these global challenges in time? AI is already revolutionizing individual scientific disciplines, but we argue here that it could be more holistic and encompassing. We introduce the concept of *virtual laboratories* as a new perspective on scientific knowledge generation and a means to incentivize new AI research and development. Despite the often perceived domain-specific research practices and inherent tacit knowledge, we argue that many elements of the research process recur across scientific domains and that even common software platforms for serving different domains may be possible. We outline how virtual laboratories will make it easier for AI researchers to contribute to a broad range of scientific domains, and highlight the mutual benefits virtual laboratories offer to both AI and domain scientists.

Impact Statement

Artificial intelligence (AI) is transforming research across natural and engineering sciences, often in the form of field-specific tools. We argue that by developing AI techniques for supporting the scientific discovery process itself we can take the next step and enable efficient human-AI collaboration for seeking new knowledge. In many cases, the AI methods can be applied across several scientific disciplines, but to encourage researchers to work on general methods we need environments that make it easy and attractive. We introduce the concept of virtual laboratories to facilitate research on domain-agnostic AI assistance for research and development and to make it easy for other scientific disciplines to use these tools.

1. Introduction

Merriam-Webster defines a *laboratory* as *"a place equipped for experimental study in a science or for testing and analysis"* or more broadly as *"a place providing opportunity for experimentation,*
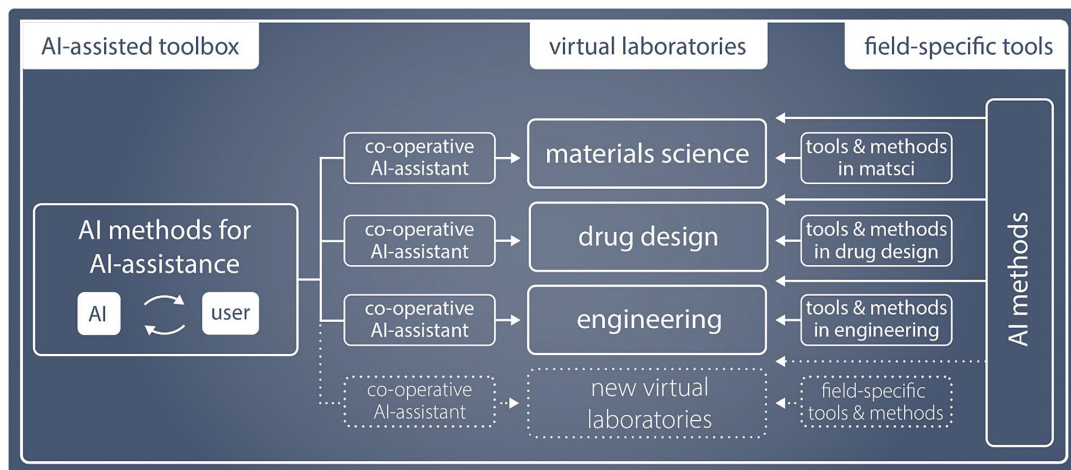
Figure 1. *AI methods enable generalizing across field-specific virtual laboratories, each using a mixture of field-specific and general methods.*

observation, or practice in a field of study."[1] The definition refers to a physical environment that exists for the purpose of making new discoveries. While laboratory tasks are now frequently carried out on computers, or more and more automated synthesis and measurement devices, the laboratory itself remains surprisingly similar to its 19th century form. In our increasingly digital world, we think it is time for a paradigm shift to *virtual laboratories* (VLs), following a loose conceptual analogy to *virtual machines* in computer systems. Continuing with the analogies, a VL is then a digital environment that enables users to carry out research tasks and that encompasses the elements of a classical laboratory as its components.

The starting point for a virtual laboratory is the computational methods and tools that are already an integral part of modern scientific practices. These include computational simulations, digital twins of various instruments, robotic measurement devices, and methods for experimental design, data analysis, and statistical estimation. In most scientific disciplines, physical laboratories already heavily use these computational tools, and research combines computation and real-world experiments. The new digital technologies already provide scale advantages and improve reproducibility and reliability.

In this point of view, we argue, however, that the current tools are not yet sufficient for building virtual laboratories, and two aspects need to be addressed. First, the current toolkit needs to be expanded. The current models and tools are typically field-specific and designed to address specific narrowly defined tasks. What is missing today are tools that would help more efficiently, as well as serve the scientific research process itself, by combining the isolated field-specific models to form a laboratory environment that supports researchers in using the digital tools to address the needs of increasing workflow, tool, and research complexity.

The second step required for reaching the full potential of virtual laboratories is to consider what can be done differently now that the computational backbone exists in different disciplines. Could we develop new types of tools (Figure 1) by thinking across laboratories, and in particular, could we benefit from advances in AI methodology? If the tools were not developed independently in each field but would instead pool the creativity, ingenuity, and resources from a variety of fields, progress would be faster and VLs could become a reality sooner. Such general applicability and acceleration is precisely the promise AI-based tools offer.

In this paper, we present a vision for AI-assisted virtual laboratories: Digitalization of research and development will move from isolated digital twins to AI-assisted support of the scientific innovation

---

[1] https://www.merriam-webster.com/dictionary/laboratory (accessed 11 December 2023).

process. In the future, new innovations are made in virtual laboratories, where researchers seamlessly operate with physical and virtual measurements in close collaboration with AI, accelerating the pace and improving the quality of research. The researchers would remain in charge and would continue to develop their research competences, but now assisted by personalized AIs that continuously learn to provide context-dependent support. The virtual laboratories are ideally supported by a common software library that combines field-specific digital twins with domain-agnostic tools for formalizing and assisting the research process.

Virtual laboratories provide a conceptual frame of reference, and in this paper, we outline practical directions for the transition from real to virtual laboratories. This paper is a call for both AI researchers and domain scientists to join forces. Section 2 introduces the main concept of virtual laboratory and outlines the high-level goals and challenges. In Section 3, we present the main actions we think should be taken by different parties. Lastly, Section 4 takes a close look at the current status, by discussing how soon we expect various disciplines to benefit from the efforts and by reviewing the state of emerging VLs in three different fields, outlining for instance how drug design is already largely done in a virtual realm but using field-specific tools.

2. Virtual laboratory concept

2.1. Virtual laboratory

Following our laboratory definition in the introduction, a *virtual laboratory* (VL) is the in silico equivalent of a physical laboratory. A VL exists primarily in a virtual space, or at least mediates the interaction of stakeholders with the VL remotely through a digital user interface. In practical terms, a VL is a collection of interconnected *digital twins* and a digital user interface (see Figure 2), operated by a scientist that remains in control but is assisted by AI. In our opinion, AI assistance is a critical VL element, because it facilitates the navigation of the complex VL environment and enhances the research process.

Digital twins are faithful computational representations of real-world entities or processes (Wright and Davidson, 2020; Blair, 2021; Niederer et al., 2021). We here consider a wider definition of digital twins
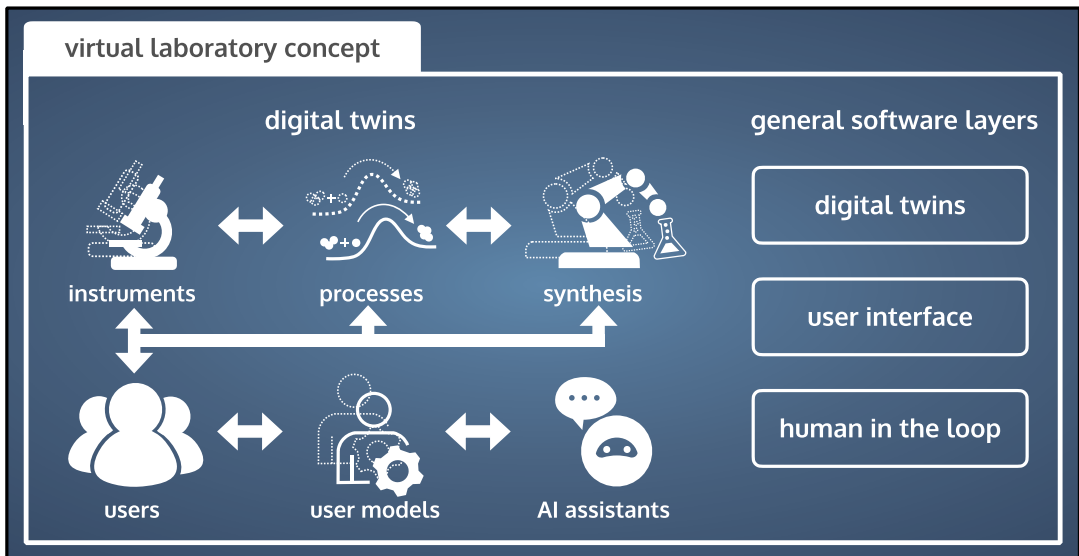

Figure 2. *In a virtual laboratory researchers or users perform virtual experiments using digital twins of assets or processes, where the experiments generally refer to any manipulation of the digital twin that may involve also other functions of the rest of the laboratory. They are assisted by AIs, which use digital twins of the researchers (user models) for interactive assistance. The process is grounded in the real world, through use of the physical instruments or assets (called for on-demand by the AI assistants).*

than usual and distinguish between three types: (a) assets, (b) processes, and (c) human interactions. In (a), physical assets is an umbrella term for scientific instruments, measuring devices, and equipment that manufacture goods, fabricate materials, and synthesize substances. In (b), computational models and simulators aim to capture physical or chemical processes. In (c), we refer to *user models* of human behavior and human-machine interactions. Combined, these three types of digital twins transfer real-world data into the virtual realm, where it is processed by simulators and AI methods.

As implied by the word *virtual*, one purpose of VLs is to transfer the experimentation and discovery process from the real into the virtual realm. In this *virtual mode*, users interact with the digital twins instead of their real-world manifestations to derive new knowledge, educate themselves, or receive assistance in complex decision-making. This usually offers significant time and resource savings compared to directly operating in a physical laboratory. The digital twins interact with the real world, when necessary, to stay up-to-date and react to changing conditions. In the *real mode*, the VL has a direct physical outcome, for example, a material or drug. The VL facilitates, accelerates, or even enables the design and development of the physical outcome.

While digital twins for assets and processes are still largely specific to a field, or even a laboratory, user models and human-machine interactions are already somewhat task-agnostic and more widely deployable. We argue that a field-agnostic virtual laboratory concept should start from these user-related components and develop generalized tools that can be quickly deployed across fields. For example, the chemical models in digital twins for drug design differ from those in construction engineering, but the cognitive models of human sequential decision-making and human biases, employed by collaborative AI assistants, are the same. These models will be useful already in cases where some of the components are not yet digitalized but involve physical measurements or manual work, often resulting in time or cost savings. Similarly, much of the backend machinery (e.g., data passing, digital twin maintenance) is general-purpose. A VL helps in delineating the field-specific components from those that generalize across fields, and thus allows efficient and parallel tool development.

2.2. Elements of digital twins

Although each digital twin serves a specific purpose, current realizations of digital twins share several aspects: (a) live coupling between the physical asset and its digital twin, via multiple streaming data sources originating from live sensing of the physical process, (b) access to additional information about the modeled process, such as geometry, topology, physical laws or 3-D characteristics for physical assets, (c) AI models utilizing the aforementioned data sources and prior knowledge to accurately predict or simulate future states of the physical twin from these, (d) some ability to perform what-if scenarios and counterfactual reasoning over the process, and (e) a decision-making mechanism (typically with a human-in-the-loop) for acting on the asset/process given the model and any what-if reasoning abilities therein.

A digital twin can also be composed of several sub-twins. These sub-twins could be in different physical locations, for example, different real-world laboratories. The VL would integrate (or aggregate) all these sub-twins into one digital twin as illustrated in Figure 3. We are now moving to the realm of decentralized/distributed inference and borrowing statistical and causal strength across experiments. Methodological frameworks such as multioutput/multitask learning (Caruana, 1997), transportability and data fusion (Pearl and Bareinboim, 2014; Bareinboim and Pearl, 2016), federated learning (Yang et al., 2019), physics-informed ML (Mao et al., 2020), and semi-parametric statistics bridging to traditional numerical methods (Girolami et al., 2021) are central for such Causal DTs and interconnected VLs.

These interconnections are graphically depicted in Figure 3 using the gemini symbol (Ⅱ) as a playful abstraction of a digital "twin." On the bottom hull, the true *data-generating process* of the real-world component is generating data that is noisily sensed via multiple sensor networks with potentially different characteristics. These are streamed upwards to the digital twin for model-based inference. The model is by construction an approximation of reality, but potential error or uncertainties can be characterized and accounted for, for example, Kleijn and Van der Vaart (2012) and Dellaporta et al. (2022). In turn, the posterior beliefs of the digital twin over the model and/or the model parameters are utilized to compute
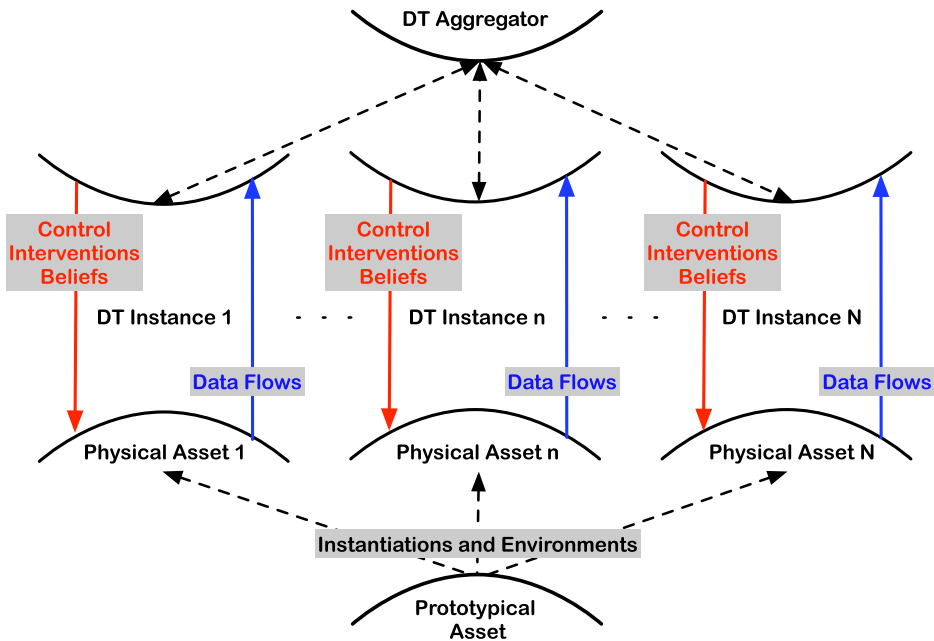
Figure 3. *Schematic of digital twins depicting the key information flow and quantities of interest. Several DTs could be aggregated in a VL. Each of the instances could be different instruments combined into one digital twin or different realizations of a device in different labs around the world.*

expectations of functions of interest with respect to these posteriors, and subsequent decision-making or simulation of what-if scenarios is performed leading to actionable interventions back on the physical asset.

2.3. A virtuous cycle of virtual laboratory development

A *virtuous cycle* is a recurring chain of events with a positive outcome. If we can start such a virtuous cycle in VL development, in which advances in AI and domain-specific knowledge benefit each other, we can increase the pace, quality, and cost-efficiency of scientific research. In this cycle, the VLs are the catalysts for facilitating the interaction and the research environment.

Once kick-started, the virtuous cycle will produce the first success stories. These will trigger an increased interest from both AI researchers and VL users, which, as the cycle progresses, should then result in a self-sustaining community effort.

2.4. Virtual laboratory library: a common software ecosystem

We argue that a key requirement for a successful, virtuous cycle is the increased use of common software components across fields. Such a necessary software ecosystem should be built so that AI advances can be developed independently in a modular fashion and taken immediately into use across many VLs with minimal effort, while making it easy for VLs to include their field-specific digital twins for assets and processes. A collection of modules and tools forms a *virtual laboratory library* (VLL), and open interfaces allow anyone to either contribute new modules or re-use the existing modules to facilitate concrete VLs in specific fields. The software ecosystem provides the technical realization of the VLs to produce scientific and societal impact. We believe that the development of VLLs is necessary not only for the permeation of VLs across fields, but also for incentivizing VL developers, VL users, and AI researchers to join forces.

3. Toward real virtual laboratories

In this section, we outline the main steps we consider necessary, from a technical perspective, to build VLs and to generalize the concept across disciplines.

3.1. Virtualization

Transferring key components of the scientific method, such as hypothesis generation, experimentation, confirmation, and discovery, from the physical to the virtual setting is the central objective of VLs that enables acceleration, reproducibility, and scalability of research. The primary vehicles for the virtualization of such components are digital twins of assets, processes, and human interactions that are interlinked inside the virtual lab and attempt to address the gap between simulation and real world through their live coupling to the physical twin and emerging frameworks for mispecification, calibration, and interventional consistency, for example, Kennedy and O'Hagan (2001) and Dellaporta et al. (2022). Significant resources are already being dedicated to improving the quality and versatility of digital twins as components of the VL and the transition will directly benefit from the results of these efforts, but dedicated research will be needed for virtualization of the research process and the human elements of that.

Many key AI technologies and research areas are necessary for the virtualization process itself: from simulators, emulators, artificial agents, and their data calibration and optimization, to reinforcement learning and robotics for automated measurement devices. Some of these, such as robotics, target the automation of physical measurements while other areas are necessary for exploring and optimizing virtual measurements and for counterfactual reasoning. We note that a large body of AI research including experimental design, Bayesian optimization, reinforcement learning, language models, generative models, synthetic data, causal inference, domain adaptation, transfer learning, probabilistic modeling, probabilistic numerics, uncertainty quantification, and physics-informed ML (Φ-ML) will be central in enabling full virtualization. Any methodological advances will then directly benefit VLs.

3.2. Human in the loop

AI tools are predominantly used to automate tasks and supplement or replace human-derived insight with data-driven models. The evolution toward "robot scientists" (Burger et al., 2020) has been invoked, but in reality, human scientists remain involved, in two ways. They drive the scientific process, by instantiating, designing, and applying AI methods, and they provide knowledge. Through *human-in-the-loop* machine learning (Monarch, 2021), prior human knowledge could be directly integrated into VLs. Human-in-the-loop methods elicit knowledge from human users to maximally improve AI models with minimal user effort.

Current human-in-the-loop methods are not compatible with the other reason humans are involved—that they drive the research process. The current methods treat humans as passive data sources (e.g., in knowledge elicitation; Mikkola et al. 2023) instead of active agents. For VLs, we need to develop human-centric AIs and human-AI collaborations (Dafoe et al., 2021). Multi-agent modeling methods from human-robot interaction (Hadfield-Menell et al., 2016) are a start, but work is still needed for formulating assistants which are useful to human scientists while leaving them in full control (Celikok et al., 2022). For this, the assistants will need to infer their user's intentions, objectives, and goals and then recommend actions in a way they understand—in other words, they would need models of human users to efficiently collaborate with them. In Section 2, we referred to these models as *user models* or digital twins for human-machine interactions. With user models of scientists, AI assistants will be able to anticipate their actions and aid them in the scientific discovery process.

While dedicated interfaces, for example, for analyzing and interpreting scientific measurements remain the main form of interaction between humans and AIs, recent advances in large language models (LLM) and their dialogue capabilities have made it possible to consider also natural language as the interface. How LLMs can assist with the analysis of scientific literature and scientific writing remain hot topics, but some studies already go further by incorporating LLMs into the scientific discovery process. For instance, discipline-specific LLMs (Luo et al., 2022; Korolev and Protsenko, 2023) can, for example,

find relationships between molecules or predict material properties based on written descriptions, blurring the line between a model and communication interface. Research on how to best use these tools for scientific discovery is only starting now (Stokel-Walker and Van Noorden, 2023; Van Dis et al., 2023), and fundamental questions on their capabilities and limits in supporting complex and long-term tasks remain. VLs are ideal platforms for studying these questions, by providing data on long-term structured interaction between teams of highly skilled human experts and dialogue systems.

In summary, VLs will be a fitting environment for mixed human-AI research teams. Before fully-fledged AI assistants become available, AI tools that give better recommendations would already be beneficial. To reach this point, advances in both AI and human-computer interaction are required.

3.3. Software layer

Building a common software layer for VLs will be critical. We only benefit from up-scaling if multiple VLs can re-use the same underlying software modules, so that AI researchers can easily develop and evaluate their methods for multiple use cases and VL hosts can easily integrate new AI elements.

The VL software layer mediates the scientific process in the virtual realm and provides the link to the physical realm, but should not be specific to any particular laboratory type. The software layer needs to represent digital twins, moderate data flows between digital twins (as well as their physical counterparts), and enable human-AI collaboration. This requires a modular architecture that communicates with domain-specific databases and models, so that all elements of AI assistance and DT operation are provided as independent modules. Common interfaces will be critical, so that both open-source and commercial tools for operating VLs will efficiently re-use the components.

We are not aware of any general VL software development even though many libraries for the individual VL components and the automation of data analysis workflows (Singh, 2019; Mölder et al., 2021) already exist (see also Section 4.4). Besides a modular architecture, an emerging software layer should:

- allow re-use existing (or future) libraries and help avoid the need to replicate general-purpose algorithms (e.g., machine learning tools or language models).
- run on standard cloud architectures and databases.
- be free and open-source, but licensed to enable commercial support for broad use in industry and research.
- be designed from the start as shared community effort, and eventually establish new standards for information transfer between digital twins and laboratories.
- actively support the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles (Wilkinson et al., 2016), required for community development and reliable operation.
- support compartmentalization of public and private data and models, so that sensitive data and proprietary simulators and digital twins can be excluded from externalized AI development.
- make it easy to run virtual instances with no direct connection to the physical laboratory, for external AI research.

3.4. Enabling and encouraging VL research

VLs generate added value from the synergy between AI and research in other domains. To create this synergy, the barrier for contributions from AI researchers and VL domain scientists should be lowered. The common software layer is necessary but not yet sufficient for this. Numerous examples demonstrate a clear benefit from lowering the contribution barrier: ImageNet data (Deng et al., 2009) revolutionized computer vision and MuJoCo (Todorov et al., 2012) and OpenAI gym (Brockman et al., 2016) reinforcement learning. We need similar success stories for VLs.

A key difference between VLs and the above examples is that VLs are linked also with physical reality and many of the interesting research questions involve humans, as explained in Section 3.2. This

introduces additional challenges but we have not identified any immediate show-stoppers that could not be overcome by combining different approaches. Many AI elements can be developed in purely digital laboratories, using simulated human activity if needed. For example, `ChemGymRL` (ChemGymRL: Chemistry Laboratory Environment)[2] offers a reinforcement learning environment for a purely virtualized chemical laboratory, Trubucco et al. provide a virtual environment for design problems (Trabucco et al., 2021), and some cognitive models of researchers can be trained with non-experts in crowd-sourcing experiments, for instance, models of working memory and decision-making (Brown et al., 2014).

We believe the most important step toward realizing our vision will be the activation of the research community. Providing computational platforms, theoretical concepts, and individual AI modules is a community effort, both in terms of sufficient resourcing, but also to ensure open standards and broad applicability across different fields. This transformation is best driven by a global open initiative, and we will work toward establishing one. The initiative will bring AI researchers and domain scientists together to design and develop the software platform, to determine incentive structures and funding models for VL hosts, for example, by extending the practices currently in place for data releases, and to work toward key standards. It will also have an important role in increasing awareness of the concept, via a workshop series organized alongside one of the leading AI conferences and challenges designed for steering the efforts of AI researchers, motivated by, for example, the effect the Netflix prize had on recommendation engine research (Bennett et al., 2007). Besides accelerating the pace of scientific research, we also expect educational synergies, for instance in the form of masters programs centered around the key aspects of VLs, bringing together students working in different disciplines to share knowledge.

4. Current status and future outlook

No fully operational VLs exist today, but significant progress is being made. To make the concept more concrete and to highlight ongoing research and potential outcomes, we first inspect different scientific disciplines from the perspective of how easy the VL concept is to apply in supporting their progress and discuss potential barriers also from non-technical perspectives in Sections 4.1 and 4.2.

Section 4.3 outlines one concrete example of domain-agnostic tools we expect to become available first, and in Section 4.4, we discuss three examples of ongoing research toward VLs in three different scientific disciplines. Each one showcases progress in digital twins of assets and processes, validating that the tools for constructing VLs are starting to be in place, while also revealing that the work toward incorporating AI within the process is happening separately within each field. This is wasteful and limits opportunities for AI research contributions, and highlights how there are clear differences between disciplines in how effectively the digital tools are being integrated into the core discover process. For ease of communication, we use examples from the authors' research domains, but emphasize that the VL concept is applicable for considerably broader range of fields.

4.1. Which domains will seize the opportunity?

The VL concept aims for maximal generality, but naturally, disciplines will vary in how much they benefit from incorporating external AI development. Table 1 provides a high-level categorization of various domains into four groups in terms of their potential for adopting VLs. The easiest conceptual mapping from current practices to VLs lies within the natural sciences, where empirical or computational observations drive the scientific method. Here we find the *early adopters* that already use computational simulations and hence will likely drive the development of VLs in the early years (see Section 4.4 for two examples), as well as the *stable followers* that will be able to re-use many of the same tools, but for whom the transition will take longer due to digitalization being more laborious.

In the other two groups, we anticipate disciplines for which it is less clear how they can best benefit from VLs. These disciplines or deployment areas do not necessarily adhere to the natural science

---

[2] https://github.com/chemgymrl/chemgymrl (accessed 11 December 2023).

Table 1. *Categorization of VL application domains and their potential*

Early adopters with high conceptual and technical readiness

Disciplines with high digitalization rate, extensive use of simulations and computational models, or fields with automated and reliable but costly experimentation

State: Currently wasting resources in developing field–specific AI for common problems. Will drive the early development of VLs and domain–agnostic AI for scientific discovery

Actions: Use more unified interfaces, specify problems for AI community, and ensure data availability

Examples: Drug discovery, material science, computational biology

Stable followers with high conceptual readiness but technical challenges

Disciplines follow the scientific method but with challenges in collecting and/or using reliable data, for example, due to laborious human experimentation, sparse passive observations, or sensitive personal data that cannot be easily shared

State: Currently implementing digitalization. Will benefit from AI methods and AI–driven experiment design. Likely to have the largest economic benefits

Actions: Ensure comprehensive data collection, use data–efficient AI methods, focus on the role of human as information source, follow the best practices of early adopters

Examples: Physics, process engineering, ecology, neuroscience, industrial R&D

Future VLs with high technical readiness but less clear mapping to the VL concept

Disciplines that use digital twins, computational models, or machine learning techniques for analysis, but do not follow the natural science paradigm

State: Currently using computational models, for example, to answer qualitative research questions or for confirmative validation, but typically not for discovery. Will benefit from having data and models more tightly integrated into the research process

Actions: Formalize the use of computational models, ensure reliability and repeatability of computational experiments, consider human as information source, focus on design problems

Examples: DTs in engineering, cognitive science, digital humanities, e–commerce

Unexplored opportunities with both technical and conceptual open challenges

Disciplines that focus primarily on qualitative research or creation, not considering their source material (interviews, historical texts, etc.) primarily as data

State: Currently mostly not using computational models, but may already be exploring generative AI as means of creativity or collaborative assistance. Will benefit from general VL tools, but likely need new conceptualization

Actions: Follow the capabilities of more formally framed VLs, focus on human–AI collaboration

Examples: Architecture, social sciences, business strategy

paradigm, but will still benefit from the broad range of domain-agnostic AI tools once the conceptual mapping is clear. We foresee individual examples of *future virtual laboratories* emerging already in relatively near future; for instance, some formal workflows using computational models for behavioral modeling already exist (Oulasvirta et al., 2022), and Section 4.4 discusses digital twins in engineering from this perspective. The *unexplored opportunities* group will likely not benefit from a large proportion of the VL tools, but will in the future be using more AI technologies and is likely to skip the phase of developing field-specific VL tools.

4.2. Challenges

We propose a vision of a feasible transformation and hence focus on the technological enablers and the need for a cultural transformation, and to raise awareness of the opportunities offered by VLs. Naturally, there will be significant challenges as well, spanning across different technical and societal dimensions. Some already identified challenges are briefly discussed below.

4.2.1. Sensitive and personal data

Many VLs need to handle sensitive data, either because of the research subject (e.g., medical research) or because data is collected on the behavior of researchers to improve AI assistance. Running the VL tools in standard secure computing environments is sufficient for ensuring data protection, but explicit effort will be required in ensuring ethical and legal use of the interaction data and the user models; see Sartor et al. (2020) for a discussion on the European legislation.

4.2.2. Legal aspects, IPR

The complex landscape of AI and intellectual property rights is currently being primarily studied from the perspective of copyright (Samuelson, 2023). VLs will need to adhere to the established regulation, but additionally, provide opportunities to study, for example, ownership of innovation. By making the roles of different entities (researchers, AI, and data) more systematic and transparent, VLs will play a role in developing responsible innovation practices (Tekic and Füller, 2023).

4.2.3. Computational and energy cost

Digital twins and computational models can save cost and human time in physical experimentation, but in some cases, the cost of simulations or AI model training is a major factor. However, ML proxies and AI tools for selecting maximally efficient experimental protocols (see Section 4.3) directly address this, and hence VLs are inherently designed to minimize also the energy cost.

4.2.4. Proprietary and legacy components, technical debt

One major technical challenge is support for software and measurement devices with prohibitive licenses. Many important devices, intentionally or unintentionally, only provide a fixed interface and lack support for efficient reading and writing of data. Disciplines, in which the central tools operate in this fashion, may drop down to the *stable followers* category (Table 1). Any use of AI in these disciplines will require case-specific workarounds or software robotics solutions, resulting in technical debt and high maintenance cost. This problem is particularly apparent for VL integration of such assets. Note, however, that dedicated AI methods, for example, quantifying the uncertainty of black-box models (Sudret et al., 2017) can help with integrating such legacy components into VLs.

4.2.5. Societal acceptance

Researchers are sometimes conservative and may raise objections on AI involvement, either because of focusing on qualitative conclusions that should be made by humans or because of hesitating on the use of approximations in place of solid evidence. More generally, we often set higher standards for autonomous systems than humans, best exemplified in the case of autonomous cars (Shariff et al., 2021). These aspects highlight the need for dedicated studies of *how* the human-AI collaboration is carried out, detailed in Section 3.2. The general public is right now rapidly being exposed to chat-based AI assistance in daily tasks with emerging research on how this transforms the society, and VLs will be an environment to study similar questions in the context of scientific research.

4.3. What generalizes across fields?

The added value of VLs hinges on a sufficiently broad set of components that are usable across fields, that is, can be generalized to new fields. DTs for assets and processes will to a large degree remain domain-specific and we should not expect them all to generalize, but we argue that almost everything else does. First, this holds for the tools needed for creating and maintaining the DTs which are general, as explained in Section 3.1, and it will be important to build on the common advances in these AI technologies. For instance, developing domain-specific solutions for addressing concept drift in measurement data would waste significant effort and the tools would soon get outdated. Second, the DTs for the researcher and the core tools for AI assistance are largely field-independent as explained in Section 3.2, and solutions for
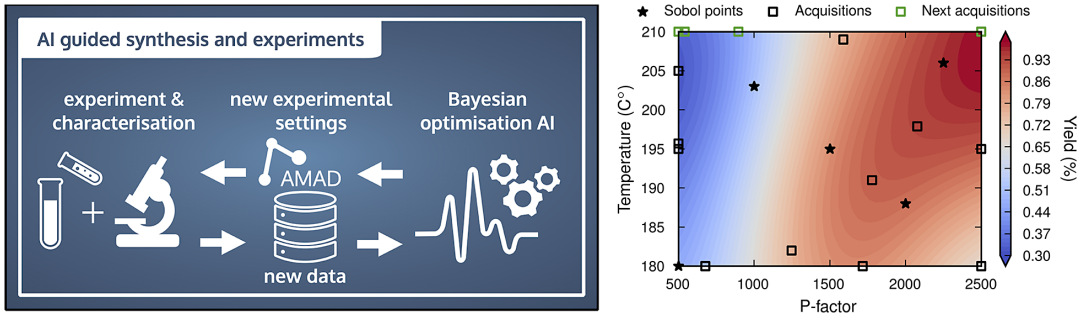
Figure 4. *Left: Conceptual illustration of AI-guided materials synthesis and characterization. The Aalto Materials Digitalization (AMAD) Platform (https://www.aalto.fi/en/services/aalto-materials-digitization-platform-amad) facilitates data transfer and collection. Right: Biomaterials example, in which AI guided the extraction and characterization of lignin from birch wood with Bayesian optimization. With very few data points (black and green squares and stars) lignin properties (here the yield) can be correlated to the experimental control variables (here temperature and reactor severity (P-factor).*

helping researchers to interact with the AI assistants would be better developed directly to be generalizable. The VL concept is a necessary enabler for their development and evaluation.

The first demonstrations of added value will come from the *early adopters* (Table 1) and already active research areas, best exemplified by AI assistance in experimental design that generalizes for all empirical sciences. A significant gap remains between the AI advances and the practical needs of a VL, with no general solutions, for example, for guiding how to best interleave approximate virtual experiments and (often high-cost) empirical experiments while ensuring the user remains in full control, and virtually no research on how to best utilize multiple alternative measurement technologies. Already rudimentary AI assistance for general empirical research setups would be a significant improvement. It would steer the AI development to directions that matter in practice, while opening opportunities for new types of AI, by providing the first systematic and field-agnostic source of data about human interaction in research, kick-starting the virtuous cycle.

4.3.1. Example: AI assistance in optimization

Consider the common task of a researcher iteratively optimizing the parameters of a process to maximize a task-specific quality measure. For instance, a chemist would seek reaction conditions to maximize the yield, or a material scientist would explore the design space to find a material with desired properties. In many cases—especially in the *stable follower* disciplines—the results of the experiments depend significantly on the competence and practices of the researcher, and irrespective of the field we can develop AI assistance for improving the reliability of experimentation, both on the level of individual measurements and the overall process.

In its simplest form, this task is described in a VL as an iteration that combines a human decision on the next configuration to try with a measurement device or its digital twin to quantify the quality of the configuration, continued until the human decides to terminate the optimization. A natural mode of AI assistance is to supplement the iteration with recommendations for the next configuration (De Peuter and Kaski, 2023), as illustrated also in Figure 4, and typically also for when to terminate. However, formulating the process in a VL enables a much broader form of assistance without notable additional effort from the VL host. For example, the following tasks could all be assisted by stand-alone AI functionality that generalizes for all VLs:

- *Education and training.* AI monitors the choices of the researcher and explains afterward which of the configurations they tried were not ideal, or creates personalized training tasks for a student-centered learning process.

- *Identification of human biases.* Humans search like search algorithms do (Borji and Itti, 2013; Sandholtz et al., 2023). Deviations from algorithmic recommendations can reveal two kinds of human biases: harmful biases preventing efficient work, and useful prior information the researcher uses to solve the problem more efficiently. Both can be used for person-specific assistance.
- *Evaluation and comparison.* Human choices can be analyzed to detect, for example, differences between laboratories or temporal patterns, as a basis for more unified work practices for the field. Data coming from different laboratories can be better combined for meta-analysis or future model training by accounting for the differences.
- *Team work.* AIs can recognize situations in which expertise of a single researcher is not sufficient for making an optimal decision, and recommend, for example, an additional evaluation of the current decision, or assist in combining the expertise of several researchers involved in the process (Song et al., 2019; Mikkola et al., 2023).
- *From repeated experimentation to DT.* Information accumulates throughout the optimization processes and AIs can effectively use this history (Swersky et al., 2013). Accumulated data is needed for building DTs and AI can assist by identifying a minimal set of required additional measurements.

For many of the tasks in this list, no tools currently exist and research into such tools is only just beginning. VLs would provide a unifying framework to join presently disjoint research and to combine independent data sources for developing methods that generalize across domains.

4.4. *Examples of progress toward VLs*

4.4.1. *Materials science*

While no fully-fledged virtual laboratories have emerged in materials science yet, the components are in place and the field is a clear *early adopter* example. The earliest databases date back to 1965. Their number has risen exponentially since the Materials Genome initiative[3] was launched in the United States in 2011 (Himanen et al., 2019). Databases evolved via data centers into materials discovery platforms by incorporating data analysis and machine learning tools. The Materials Project (Jain et al., 2013), the Novel Materials Discovery (NOMAD) laboratory (Ghiringhelli et al., 2017), and Citrine Informatics (O'Mara et al., 2016) are prominent examples of such materials discovery platforms and could be viewed as virtual laboratory incubators.

Digital twins are more common in engineering and industry (see Section 4.4.3). They are slowly emerging in materials science, too, with battery development leading the way (Ayerbe et al., 2021). Ngandjong et al. recently proposed a digital twin of a Li-ion battery manufacturing platform that combines modeling approaches at different scales (Ngandjong et al., 2021). Thomitzek et al. added a battery cell production digital twin based on digitalization and mechanistic modeling (Thomitzek et al., 2022). Regarding scientific instruments, Passananti et al. developed a digital twin of a chemical ionization atmospheric pressure interface time-of-flight mass spectrometry (CI-APi-TOF-MS) that facilitates the analysis of molecular cluster formation events in the atmosphere (Passananti et al., 2019).

In the Finnish Center for Artificial Intelligence (FCAI), we are developing AI-guided experimentation and synthesis techniques (Jin et al., 2022; Löfgren et al., 2022). An example is presented in Figure 4. A Bayesian optimization-based AI requests data from scientists who synthesize and characterize materials. The example shows the extraction of the biopolymer lignin from birch wood and the characterization of the structural properties with 2D nuclear magnetic resonance (NMR) spectroscopy. The data is returned to the AI, which updates its surrogate model of the process and subsequently issues new data requests. Data transfer is mediated by the Aalto Materials Digitalization

---

[3] https://www.mgi.gov/ (accessed 11 December 2023).

(AMAD) Platform[4] that acts as local data back-end for future virtual laboratories at Aalto University. The lower panel of Figure 4 demonstrates that with relatively few datapoints (i.e., time-consuming synthesis steps), the lignin yield can be maximized. In addition, the surrogate model provides an insightful visualization to the operating scientists of the relation between the extraction (or synthesis) conditions and the lignin (or materials) properties. Such AI-guidance tools are not only the first step toward autonomous experiments or fabrication, and thus the corresponding digital twins, but they also facilitate the collection of data that has traditionally been difficult to digitize due to its acquisition cost (e.g., human, process or computational time and instrument cost).

Bayesian optimization also forms an integral part of self-driving laboratories (SDLs). Such SDLs usually combine robotics, machine learning, and materials chemistry to autonomously synthesize compounds and make materials (Stein and Gregoire, 2019; MacLeod et al., 2020; Abdel-Latif et al., 2021; Rooney et al., 2022; Seifrid et al., 2022). SDLs accelerate materials design and exploration and are an important component of future virtual laboratories. SDL concepts and software are already deployed across different materials science domains (Stein and Gregoire, 2019; Stein et al., 2022), medicine, and drug design (Hickman et al., 2023; Nigam et al., 2022) and thus are an excellent example of the potential for generalization and a common virtual laboratory framework.

Increasingly, software for materials characterization and design is released as open source, such as ChemOS (Roch et al., 2020), ChemSpectra (Huang et al., 2021), LiberTem (Clausen et al., 2020), elabFTW (A Free and Open Source Electronic Lab Notebook),[5] FINALES (Vogler et al., 2023), or Chemotion (Electronic Laboratory Notebook & Repository for Research Data)[6]. These components have yet to be cast into a generalized software and ontology framework (Scheffler et al., 2022) for virtual laboratories as advocated in this article or by Brinson et al. in their recent call for *community action on FAIR data* (Brinson et al., 2023). In this context, Deagen et al. proposed a materials–information twin tetrahedra (MITT) (Deagen et al., 2022). The term "digital twin" is here used as an analogy between concepts in materials and information science and does not refer to a component of a virtual laboratory. With MITT they advocate a holistic, data-driven approach to materials science, which we believe could be further extended across scientific domains. In a similar vein, Suzuki et al. recently promoted a knowledge transfer from AI applications in pharmaceutical science to materials science through an automated machine learning framework (Suzuki et al., 2022).

4.4.2. Drug design

Applying AI to drug design has become very popular in the last 5 years, triggered by the innovations in AI (Chen et al., 2018). Common application areas are molecular de novo generation, synthetic route predictions, and molecular property predictions. In drug design a starting molecule with typically poor properties is iteratively optimized until a molecule with properties suitable to start clinical trials is identified. The iterative cycle is usually called the Design-Make-Test-Analyze (DMTA) cycle (Figure 5; Plowright et al., 2012). Due to both clearly defined iterative innovation processes and the growing use of computational experiments drug design is an *early adopter*.

The virtual drug design laboratory will consist of digital twins for the different components in the DMTA cycle. Several of the necessary digital twins are under development. Digital twins are developed for the design part through deep learning-based molecular generation, for the make part through designing synthetic routes by deep learning, and for the test part through developing digital twins for the assays that are used to test the molecules. These aspects are often studied in isolation, and advances in one task sometimes reveal limitations in others. For example, improved generative AI methods for molecule generation (Sadybekov and Katritch, 2023) call for better screening tools of the resulting, vast candidate sets (Gangwal et al., 2024). An outstanding important research task is to find

---

[4] https://www.aalto.fi/en/services/aalto-materials-digitization-platform-amad (accessed 11 December 2023).
[5] https://www.elabftw.net/ (accessed 11 December 2023).
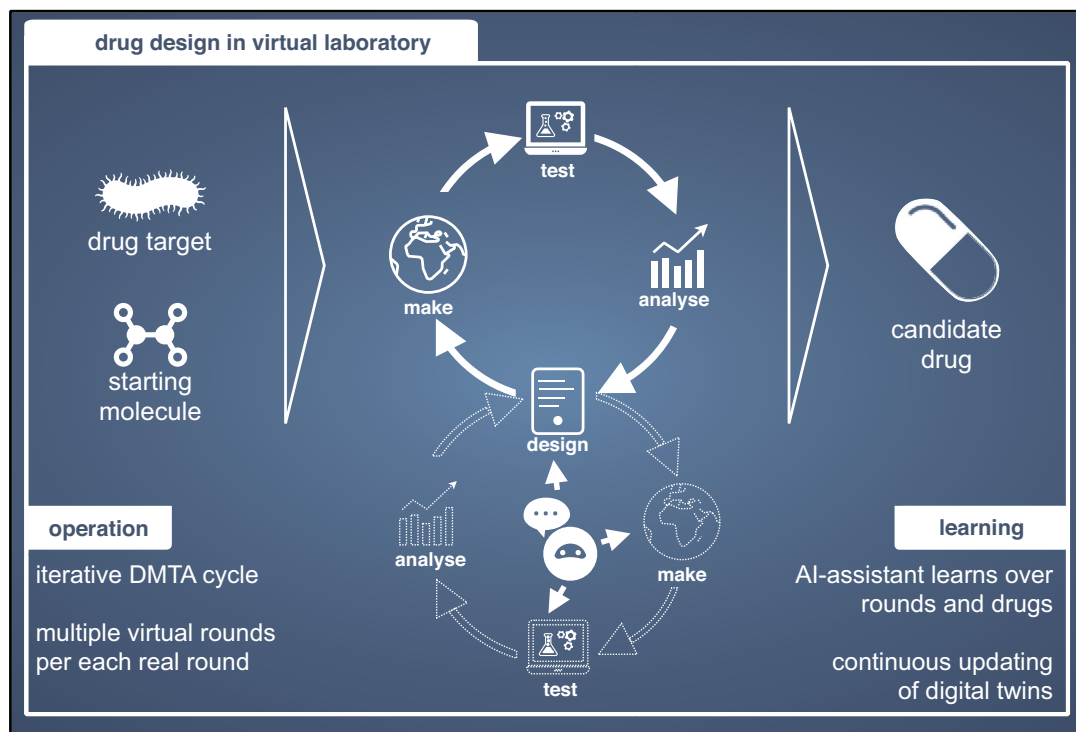[6] https://chemotion.net/ (accessed 11 December 2023).

Figure 5. *Drug design is based on iterative cycles of Design-Make-Test-Analyze (DMTA). Within each round, several iterative cycles can be performed in a virtual laboratory (bottom cycle).*

out how implicit knowledge residing with the scientists can be modeled through human-in-the-loop modeling, so that it can be included in the digital twins of the design and analysis steps. A recent concrete example (Sundin et al., 2022) infers implicit user goals from their interaction with the results of a drug synthesis model, providing means for learning the user model agnostic to the specific synthesis model used.

It is important to keep in mind that the virtual laboratory is an approximation of a real drug design laboratory. Virtual molecules are optimized in the virtual laboratory and then actually synthesized and tested in a real laboratory in an iterative manner. An optimal laboratory would combine a virtual laboratory with a fully automated real laboratory. There are several efforts ongoing to create autonomous automation systems for synthesizing and optimizing molecules (Coley et al., 2019) (see also Section 4.4.1). Thus for drug design, virtual and real laboratories need to exist in close collaboration, where as good compounds as possible are proposed by the virtual laboratory, the molecules are then synthesized as efficiently as possible in the laboratory, and the resulting data is fed back to the virtual laboratory. It should be noted that there are components describing the interactions between atoms that would be common for all VLs describing molecules like a material science VL. Examples are molecular dynamics software and atomic force field parameters.

4.4.3. Data-centric engineering

Engineering has recently witnessed a proliferation in data-centric techniques and digital twin development. In many cases constructing the DT is a significant investment and has been a central research focus, whereas the question on how they are best used for scientific or R&D advances remains more open, making the field a *future VL* example (Table 1). Limited forms of interaction and underuse of AI technologies in DTs have been identified as challenges (Tao et al., 2024), but the necessary components

for considering engineering DTs as part of a VL are starting to emerge. For example, recent open software tools for DTs (Tao et al., 2024) pave way for domain-agnostic DT operations and human-in-the-loop models are already considered in design use cases (Batty, 2024).

VLs that enable rapid deployment of DTs will offer tools that generalize across domain-specific instances. The main generalization areas lie at three boundary interfaces: (a) between sensing systems and the DT (e.g., optimal sensor placement and dataset shift tools), (b) between the DT and machine-assisted human decision-making (e.g., AI-assistant interfaces and sequential decision-making tools, and (c) between DT instances (e.g., federated learning and evidence synthesis tools) in a VL that integrates an ecosystem of DTs. Progress in common tools for these aspects will form the basis of VLs that efficiently leverage engineering DTs.
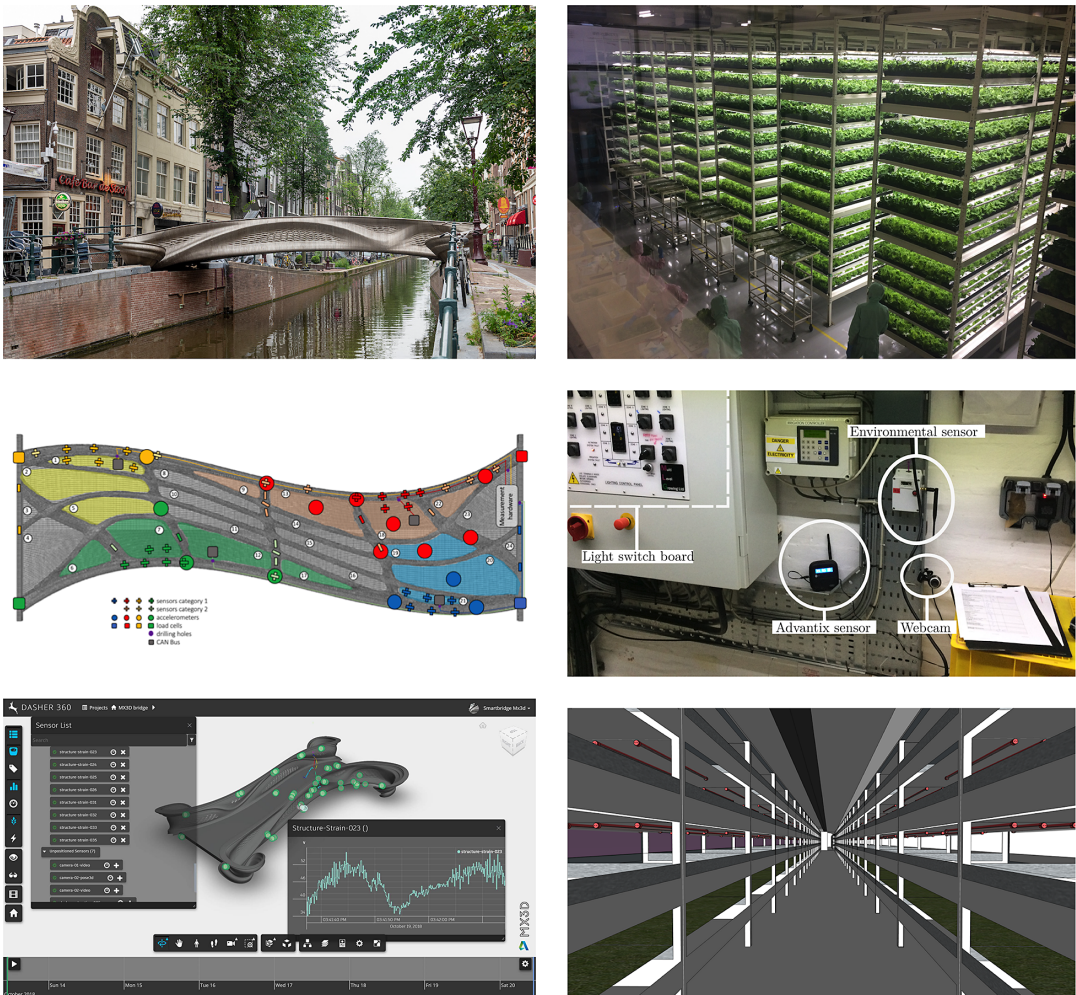

Figure 6. *Left: The 3D-printed steel bridge currently installed in Amsterdam, Netherlands, and its multiple sensing arrays that are streaming live data into the corresponding digital twin in the Turing, UK. Images by Joris Laarman Labs, Thea van den Heuvel, MX3D, and AutoDesk Research. Right: The underground farm in Clapham, London, and its multiple sensing arrays that are streaming live data flows into the corresponding digital twin in The Turing and the University of Cambridge, UK. Images by Rebecca Ward, Flora Roumpani, and Zero Carbon Farms Ltd.*

We showcase here two examples of recent engineering DTs developed at the Alan Turing Institute with academic and industrial partners, demonstrating the maturity of the technical enablers and outlining the opportunities a VL would open. The first one is the world's first 3-D printed steel bridge (MX3D bridge) depicted in Figure 6 (left) and currently situated in Amsterdam, Netherlands. Various sensor networks on the bridge, such as cameras, accelerometers, and load cells, stream live data to its digital twin at the Turing Institute in the UK. The underlying DT model has been developed based on the StatFEM methodology that was recently introduced (Girolami et al., 2021) to formally synthesize observational data and numerical models of its structure, and could be further developed to become a common tool for engineering DTs. Such DTs when embedded in the envisioned VL framework would benefit from evidence integration across physical structures; in this particular case rapidly improving our understanding of 3-D printed steel structures and their stress–strain response under load across environmental and loading conditions.

The second example, from the CROP project,[7] is a digital twin of an underground farm in a tunnel situated in Clapham, London, UK, depicted in Figure 6 (right). This is a hydroponics system with 2 aisles running in parallel in 23 zones and 2 meters long. The underlying DT model utilizes particle filtering for model calibration (Ward et al., 2021) and data synthesis, while various environmental measurements and camera footage are live-streamed from sensor networks to monitor crop health, environmental conditions, and yields. The DT already enables the study of a broad range of research questions, from forecasting yields and future conditions to optimizing the placement of crops and the environmental conditions. Within a VL it could be considered more generally as a digital experimentation tool that integrates into the overall research process, likely bringing scientific advantages beyond speed gains such as scenario testing. However, compared to standard measurement devices a DT is considerably more complex in its offering, making the best way of embedding it into a VL an open research question.

5. Conclusion

We introduced the virtual laboratory concept to amalgamate scientific research and R&D in industry with AI technology and AI assistance. We highlighted the benefits of VLs for both research laboratories and AI researchers and outlined key requirements of a common software layer and various research directions to proceed toward VLs. In our opinion, VLs are a community effort and we seek to bring AI researchers and scientists of other domains together to raise awareness for the VL concept and to work together toward realizing VLs.

The goal of transforming research with AI is ambitious and the transformation will not happen fast. The domain scientists are already working toward this direction, as highlighted by the examples in this paper, and hence we conclude our work with words of encouragement for the AI researchers. In short, VLs provide AI researchers means and incentives to contribute to the scientific efforts for solving the grand challenges we are facing. It is hard to think of a subarea of AI that would not be useful for VLs, and hence VLs will provide unique opportunities and cross-fertilization already within AI itself. In many areas, from reinforcement learning to constrained optimization and probabilistic modeling, the current techniques are already clearly sufficient for becoming core elements of VLs. In others, such as causal inference, probabilistic numerics, and domain-agnostic models of researchers, the VLs will provide concrete cases for testing the current solutions and identifying future research directions.

---

[7] https://github.com/alan-turing-institute/CROP.

References

Abdel-Latif K, **Epps RW**, **Bateni F**, **Han S**, **Reyes KG and Abolhasani M** (2021) Self-driven multistep quantum dot synthesis enabled by autonomous robotic experimentation in flow. *Advanced Intelligent Systems 3*(2), 2000245. https://doi.org/10.1002/aisy.202000245.

Ayerbe E, **Berecibar M**, **Clark S**, **Franco AA and Ruhland J** (2021) Digitalization of battery manufacturing: Current status, challenges, and opportunities. *Advanced Energy Materials 12*, 2102696. https://doi.org/10.1002/aenm.202102696.

Bareinboim E and Pearl J (2016) Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences 113*(27), 7345–7352.

Batty M (2024) Digital twins in city planning. *Nature Computational Science 4*(3), 192–199.

Bennett J, **Lanning S and Netflix N** (2007) The netflix prize. In *KDD Cup and Workshop in Conjunction with KDD*. San Jose, CA: ACM.

Blair GS (2021) Digital twins of the natural environment. *Patterns 2*(10), 100359.

Borji A and Itti L (2013) Bayesian optimization explains human active search. *Advances in Neural Information Processing Systems 26*, 1–9.

Brinson LC, **Bartolo LM**, **Blaiszik B**, **Elbert D**, **Foster I**, **Strachan A and Voorhees PW** (2023) Community action on fair data will fuel a revolution in materials research. *MRS Bulletin 49*, 12–16. https://doi.org/10.1557/s43577-023-00498-4.

Brockman G, **Cheung V**, **Pettersson L**, **Schneider J**, **Schulman J**, **Tang J and Zaremba W** (2016) OpenAI gym. arXiv preprint arXiv:1606.01540.

Brown HR, **Zeidman P**, **Smittenaar P**, **Adams RA**, **McNab F**, **Rutledge RB and Dolan RJ** (2014) Crowdsourcing for cognitive science – The utility of smartphones. *PLoS One 9*(7), e100662.

Burger B, **Maffettone PM**, **Gusev VV**, **Aitchison CM**, **Bai Y**, **Wang X**, **Li X**, **Alston BM**, **Li B**, **Clowes R**, **Rankin N**, **Harris B**, **Sprick RS and Cooper AI** (2020) A mobile robotic chemist. *Nature 583*(7815), 237–241. https://doi.org/10.1038/s41586-020-2442-2.

Caruana R (1997) Multitask learning. *Machine Learning 28*(1), 41–75.

Celikok MM, **Oliehoek FA and Kaski S** (2022) Best-response Bayesian reinforcement learning with Bayes-adaptive POMDPs for centraurs. In *AAMAS 2022, Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Chen H, **Engkvist O**, **Wang Y**, **Olivecrona M and Blaschke T** (2018) The rise of deep learning in drug discovery. *Drug Discovery Today 23*(6), 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039.

Clausen A, **Weber D**, **Ruzaeva K**, **Migunov V**, **Baburajan A**, **Bahuleyan A**, **Caron J**, **Chandra R**, **Halder S**, **Nord M**, **Müller-Caspary K and Dunin-Borkowski RE** (2020) Libertem: Software platform for scalable multidimensional data processing in transmission electron microscopy. *Journal of Open Source Software 5*(50), 2006. https://doi.org/10.21105/joss.02006.

Coley CW, **Thomas DA**, **Lummiss JAM**, **Jaworski JN**, **Breen CP**, **Schultz V**, **Hart T**, **Fishman JS**, **Rogers L**, **Gao H**, **Hicklin RW**, **Plehiers PP**, **Byington J**, **Piotti JS**, **Green WH**, **Hart AJ**, **Jamison TF and Jensen KF** (2019) A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science 365*(6453), eaax1566. https://doi.org/10.1126/science.aax1566.

Dafoe A, **Bachrach Y**, **Hadfield G**, **Horvitz E**, **Larson K and Graepel T** (2021) Cooperative AI: Machines must learn to find common ground. *Nature 593*(7857), 33–36. https://doi.org/10.1038/d41586-021-01170-0.

De Peuter S and Kaski S (2023) Zero-shot assistance in sequential decision problems. In *Proceedings of AAAI-23, the Thirty-Seventh AAAI Conference on Artificial Intelligence*. Washington, DC: AAAI, pp. 11551–11559. https://doi.org/10.1609/aaai.v37i10.26365.

Deagen ME, **Brinson LC**, **Vaia RA and Schadler LS** (2022) The materials tetrahedron has a "digital twin". *MRS Bulletin 47*, 379. https://doi.org/10.1557/s43577-021-00214-0.

Dellaporta C, **Knoblauch J**, **Damoulas T and Briol F-X** (2022) Robust Bayesian inference for simulator-based models via the mmd posterior bootstrap. In *International Conference on Artificial Intelligence and Statistics*. Valencia: PMLR, pp. 943–970.

Deng J, **Dong W**, **Socher R**, **Li L-J**, **Li K, and Fei-Fei L** (2009) Imagenet: A large-scale image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami,FL: IEEE, pp. 248–255.

Gangwal A, **Ansari A**, **Ahmad I**, **Azad AK**, **Kumarasamy V**, **Subramaniyan V and Wong LS** (2024) Generative artificial intelligence in drug discovery: Basic framework, recent advances, challenges, and opportunities. *Frontiers in Pharmacology 15*, 1331062.

Ghiringhelli LM, **Carbogno C**, **Levchenko S**, **Mohamed F**, **Huhs G**, **Lüders M**, **Oliveira M and Scheffler M** (2017) Towards efficient data exchange and sharing for big-data driven materials science: Metadata and data formats. *NPJ Computational Materials 3*, 46. https://doi.org/10.1038/s41524-017-0048-5.

Girolami M, **Febrianto E**, **Yin G and Cirak F** (2021) The statistical finite element method (statFEM) for coherent synthesis of observation data and model predictions. *Computer Methods in Applied Mechanics and Engineering 375*, 113533.

Hadfield-Menell D, **Russell SJ**, **Abbeel P and Dragan A** (2016) Cooperative inverse reinforcement learning. In *Advances in Neural Information Processing Systems*, vol. 29. Barcelona, Spain: Curran Associates, Inc.

Hickman RJ, **Bannigan P**, **Bao Z**, **Aspuru-Guzik A and Allen C** (2023) Self-driving laboratories: A paradigm shift in nanomedicine development. *Matter 6*(4), 1071–1081.

Himanen L, **Geurts A**, **Foster AS and Rinke P** (2019) Data-driven materials science: Status, challenges, and perspectives. *Advancement of Science 6*(21), 1900808.

Huang Y-C, **Tremouilhac P**, **Nguyen A**, **Jung N and Bräse S** (2021) Chemspectra: A web-based spectra editor for analytical data. *Journal of Cheminformatics 13*(1), 8. https://doi.org/10.1186/s13321-020-00481-0.

Jain A, **Ong SP**, **Hautier G**, **Chen W**, **Richards WD**, **Dacek S**, **Cholia S**, **Gunter D**, **Skinner D**, **Ceder G and Persson K** (2013) The materials project: A materials genome approach to accelerating materials innovation. *APL Materials 1*(1), 011002. https://doi.org/10.1063/1.4812323.

Jin S-A, **Kämäräinen T**, **Rinke P**, **Rojas OJ and Todorović M** (2022) Machine learning as a tool to engineer microstructures: Morphological prediction of tannin-based colloids using Bayesian surrogate models. *MRS Bulletin 47*, 29.

Kennedy MC and O'Hagan A (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(3), 425–464.

Kleijn BJ and Van der Vaart AW (2012) The bernstein-von-mises theorem under misspecification.

Korolev V and Protsenko P (2023) Accurate, interpretable predictions of materials properties within transformer language models. *Patterns 4*(10), 100803.

Löfgren J, **Tarasov D**, **Koitto T**, **Rinke P**, **Balakshin M and Todorović M** (2022) *Lignin biorefinery optimization through machine learning. ChemRxiv.* Available at https://chemrxiv.org/engage/chemrxiv-article-details/61370b74b817b46c9a14e7cd (accessed 11 December 2023).

Luo R, **Sun L**, **Xia Y**, **Qin T**, **Zhang S**, **Poon H and Liu T-Y** (2022) BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics 23*(6), bbac409. https://doi.org/10.1093/bib/bbac409.

MacLeod BP, **Parlane FGL**, **Morrissey TD**, **Häse F**, **Roch LM**, **Dettelbach KE**, **Moreira R**, **Yunker LPE**, **Rooney MB**, **Deeth JR**, **Lai V**, **Ng GJ**, **Situ H**, **Zhang RH**, **Elliott MS**, **Haley TH**, **Dvorak DJ**, **Aspuru-Guzik A**, **Hein JE and Berlinguette CP** (2020) Self-driving laboratory for accelerated discovery of thin-film materials. *Science Advances 6*(20), eaaz8867. https://doi.org/10.1126/sciadv.aaz8867.

Mao Z, **Jagtap AD and Karniadakis GE** (2020) Physics-informed neural networks for high-speed flows. *Computer Methods in Applied Mechanics and Engineering 360*, 112789.

Mikkola P, **Martin OA**, **Chandramouli S**, **Hartmann M**, **Pla OA**, **Thomas O**, **Pesonen H**, **Corander J**, **Vehtari A**, **Kaski S**, **Bürkner P-C and Klami A** (2023) Prior knowledge elicitation: The past, present, and future. *Bayesian Analysis* 1–23. https://doi.org/10.1214/23-BA1381.

Mikkola P, **Martinelli J**, **Filstroff L and Kaski S** (2023) Multi-fidelity Bayesian optimization with unreliable information sources. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023*, vol. 206. Valencia: PMLR, pp. 7425–7454.

Mölder F, **Jablonski KP**, **Letcher B**, **Hall MB**, **Tomkins-Tinch CH**, **Sochat V**, **Forster J**, **Lee S**, **Twardziok SO**, **Kanitz A**, **Wilm A**, **Holtgrewe M**, **Rahmann S**, **Nahnsen S and Köster J** (2021) Sustainable data analysis with Snakemake. *F1000Research 10*, 33.

Monarch R (2021) *Human-in-the-Loop Machine Learning*. New York: Manning.

Ngandjong AC, **Lombardo T**, **Primo EN**, **Chouchane M**, **Shodiev A**, **Arcelus O and Franco AA** (2021) Investigating electrode calendering and its impact on electrochemical performance by means of a new discrete element method model: Towards a digital twin of Li-ion battery manufacturing. *Journal of Power Sources 485*, 229320. https://doi.org/10.1016/j.jpowsour.2020.229320.

Niederer SA, **Sacks MS**, **Girolami M and Willcox K** (2021) Scaling digital twins from the artisanal to the industrial. *Nature Computational Science 1*(5), 313–320.

Nigam AK, **Pollice R**, **Tom G**, **Jorner K**, **Thiede LA**, **Kundaje A and Aspuru-Guzik A** (2022) Tartarus: A benchmarking platform for realistic and practical inverse molecular design, arXiv:2209.12487. Available at https://arxiv.org/abs/2209.12487.

O'Mara J, **Meredig B and Michel K** (2016) Materials data infrastructure: A case study of the citrination platform to examine data import, storage, and access. *JOM 68*(8), 2031–2034. https://doi.org/10.1007/s11837-016-1984-0.

Oulasvirta A, **Jokinen JPP and Howes A** (2022) Computational rationality as a theory of interaction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22*. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3491102.3517739.

Passananti M, **Zapadinsky E**, **Zanca T**, **Kangasluoma J**, **Myllys N**, **Rissanen MP**, **Kurtén T**, **Ehn M**, **Attoui M and Vehkamäki H** (2019) How well can we predict cluster fragmentation inside a mass spectrometer? *Chemical Communications 55*, 5946–5949. https://doi.org/10.1039/C9CC02896J.

Pearl J and Bareinboim E (2014) External validity: From do-calculus to transportability across populations. *Statistical Science 29*(4), 579–595.

Plowright AT, **Johnstone C**, **Kihlberg J**, **Pettersson J**, **Robb G and Thompson RA** (2012) Hypothesis driven drug design: Improving quality and effectiveness of the design-make-test-analyse cycle. *Drug Discovery Today 17*(1), 56–62. https://doi. org/10.1016/j.drudis.2011.09.012.

Roch LM, **Häse F**, **Kreisbeck C**, **Tamayo-Mendoza T**, **Yunker LPE**, **Hein JE and Aspuru-Guzik A** (2020) Chemos: An orchestration software to democratize autonomous discovery. *PLoS One 15*(4), 1–18. https://doi.org/10.1371/journal. pone.0229862.

Rooney MB, **MacLeod BP**, **Oldford R**, **Thompson ZJ**, **White KL**, **Tungjunyatham J**, **Stankiewicz BJ and Berlinguette CP** (2022) A self-driving laboratory designed to accelerate the discovery of adhesive materials. *Digital Discovery 1*, 382–389. https:// doi.org/10.1039/D2DD00029F.

Sadybekov AV and Katritch V (2023) Computational approaches streamlining drug discovery. *Nature 616*(7958), 673–685.

Samuelson P (2023) Generative ai meets copyright. *Science 381*(6654), 158–161.

Sandholtz N, **Miyamoto Y**, **Bornn L and Smith MA** (2023) Inverse Bayesian optimization: Learning human acquisition functions in an exploration vs exploitation search task. *Bayesian Analysis 18*(1), 1–24.

Sartor G and Lagioia F (2020) The impact of the general data protection regulation (GDPR) on artificial intelligence.

Scheffler M, **Aeschlimann M**, **Albrecht M**, **Bereau T**, **Bungartz H-J**, **Felser C**, **Greiner M**, **Groß A**, **Koch CT**, **Kremer K**, **Nagel WE**, **Scheidgen M**, **Wöll C and Draxl C** (2022) Fair data enabling new horizons for materials research. *Nature 604* (7907), 635–642. https://doi.org/10.1038/s41586-022-04501-x.

Seifrid M, **Pollice R**, **Aguilar-Granda A**, **Chan ZM**, **Hotta K**, **Ser CT**, **Vestfrid J**, **Wu TC and Aspuru-Guzik A** (2022) Autonomous chemical experiments: Challenges and perspectives on establishing a self-driving lab. *Accounts of Chemical Research 55*(17), 2454–2466. https://doi.org/10.1021/acs.accounts.2c00220.

Shariff A, **Bonnefon J-F and Rahwan I** (2021) How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. *Transportation Research Part C: Emerging Technologies 126*, 103069. https://doi.org/10.1016/j. trc.2021.103069.

Singh P (2019) Airflow. In *Learn PySpark*. Cham: Springer, pp. 67–84.

Song J, **Chen Y and Yue Y** (2019) A general framework for multi-fidelity Bayesian optimization with Gaussian processes. In *The 22nd International Conference on Artificial Intelligence and Statistics*. Naha, Japan: PMLR, pp. 3158–3167.

Stein HS and Gregoire JM (2019) Progress and prospects for accelerating materials science with automated and autonomous workflows. *Chemical Science 10*, 9640–9649. https://doi.org/10.1039/C9SC03766G.

Stein HS, **Sanin A**, **Rahmanian F**, **Zhang B**, **Vogler M**, **Flowers JK**, **Fischer L**, **Fuchs S**, **Choudhary N and Schroeder L** (2022) From materials discovery to system optimization by integrating combinatorial electrochemistry and data science. *Current Opinion in Electrochemistry 35*, 101053. https://doi.org/10.1016/j.coelec.2022.101053.

Stokel-Walker C and Van Noorden R (2023) What ChatGPT and generative AI mean for science. *Nature 614*(7947), 214–216.

Sudret B, **Marelli S and Wiart J** (2017) Surrogate models for uncertainty quantification: An overview. In *2017 11th European Conference on Antennas and Propagation (EUCAP)*. Paris: IEEE, pp. 793–797.

Sundin I, **Voronov A**, **Xiao H**, **Papadopoulos K**, **Bjerrum EJ**, **Heinonen M**, **Patronov A**, **Kaski S and Engkvist O** (2022) Human-in-the-loop assisted de novo molecular design. *Journal of Cheminformatics 14*(1), 1–16.

Suzuki H, **Kurosawa S**, **Marcella S**, **Kanba M**, **Koretaka Y**, **Tsuji A and Okumura T** (2022) How AI application in pharmaceutical industries is beneficial to materials science. *Journal of Physics D: Applied Physics 55* (24), 243002. https:// doi.org/10.1088/1361-6463/ac3a48.

Swersky K, **Snoek J and Adams RP** (2013) Multi-task Bayesian optimization. *Advances in Neural Information Processing Systems 26*, 2004–2012.

Tao F, **Sun X**, **Cheng J**, **Zhu Y**, **Liu W**, **Wang Y**, **Xu H**, **Hu T**, **Liu X**, **Liu T**, **Sun Z**, **Xu J**, **Bao J**, **Xiang F and Jin X** (2024) makeTwin: A reference architecture for digital twin software platform. *Chinese Journal of Aeronautics 37*(1), 1–18. https://doi. org/10.1016/j.cja.2023.05.002.

Tao F, **Zhang H and Zhang C** (2024) Advancements and challenges of digital twins in industry. *Nature Computational Science 4* (3), 169–177.

Tekic Z and Füller J (2023) Managing innovation in the era of AI. *Technology in Society 73*, 102254. https://doi.org/10.1016/j. techsoc.2023.102254.

Thomitzek M, **Schmidt O**, **Silva GV**, **Karaki H**, **Lippke M**, **Krewer U**, **Schröder D**, **Kwade A and Herrmann C** (2022) Digitalization platform for mechanistic modeling of battery cell production. *Sustainability 14*(3), 1530. https://doi.org/10.3390/ su14031530.

Todorov E, **Erez T and Tassa Y** (2012) Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vilamoura-Algarve, Portugal: IEEE, pp. 5026–5033.

Trabucco B, **Kumar A**, **Geng X and Levine S** (2021) Design-Bench: Benchmarks for data-driven offline model-based optimization. Available at https://openreview.net/forum?id=cQzf26aA3vM (accessed 11 December 2023).

Van Dis EA, **Bollen J**, **Zuidema W**, **van Rooij R and Bockting CL** (2023) ChatGPT: Five priorities for research. *Nature 614* (7947), 224–226.

Vogler M, **Busk J**, **Hajiyani H**, **Jørgensen PB**, **Safaei N**, **Castelli IE**, **Ramirez FF**, **Carlsson J**, **Pizzi G**, **Clark S**, **Hanke F**, **Bhowmik A and Stein HS** (2023) Brokering between tenants for an international materials acceleration platform. *Matter 6*(9), 2647–2665. https://doi.org/10.1016/j.matt.2023.07.016.

Ward R, **Choudhary R**, **Gregory A**, **Jans-Singh M and Girolami M** (2021) Continuous calibration of a digital twin: Comparison of particle filter and Bayesian calibration approaches. *Data-Centric Engineering 2*, e15.

Wilkinson MD, **Dumontier M**, **Aalbersberg IJ**, **Appleton G**, **Axton M**, **Baak A**, **Blomberg N**, **Boiten J-W**, **da Silva Santos LB**, **Bourne PE**, **Bouwman J**, **Brookes AJ**, **Clark T**, **Crosas M**, **Dillo I**, **Dumon O**, **Edmunds S**, **Evelo CT**, **Finkers R**, **Gonzalez-Beltran A**, **Gray AJ**, **Groth P**, **Goble C**, **Grethe JS**, **Heringa J**, **'t Hoen PA**, **Hooft R**, **Kuhn T**, **Kok R**, **Kok J**, **Lusher SJ**, **Martone ME**, **Mons A**, **Packer AL**, **Persson B**, **Rocca-Serra P**, **Roos M**, **van Schaik R**, **Sansone SA**, **Schultes E**, **Sengstag T**, **Slater T**, **Strawn G**, **Swertz MA**, **Thompson M**, **van der Lei J**, **van Mulligen E**, **Velterop J**, **Waagmeester A**, **Wittenburg P**, **Wolstencroft K**, **Zhao J and Mons B** (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data 3*(1), 1–9.

Wright L and Davidson S (2020) How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences 7*(1), 13. https://doi.org/10.1186/s40323-020-00147-4.

Yang Q, **Liu Y**, **Cheng Y**, **Kang Y**, **Chen T and Yu H** (2019) Federated learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning 13*(3), 1–207.