# 16TH IMG PHD CONFERENCE

Review

## Scientific discovery in the age of artificial intelligence

https://doi.org/10.1038/s41586-023-06221-2

Received: 30 March 2022

Accepted: 16 May 2023

Published online: 2 August 2023

Check for updates

Hanchen Wang[1,2,3,14,15], Tianfan Fu[3,15], Yuanqi Du[4,15], Wenhao Gao[5], Kexin Huang[6], Ziming Liu[7], Payal Chandak[8], Shengchao Liu[9,10], Peter Van Katwyk[11,12], Andreea Deac[9,10], Anima Anandkumar[3,13], Karianne Bergen[11,12], Carla P. Gomes[4], Shirley Ho[14,15,16,17], Pushmeet Kohli[18], Joan Lasenby[1], Jure Leskovec[8], Tie-Yan Liu[19], Arjun Manrai[20], Debora Marks[21,22], Bharath Ramsundar[23], Le Song[24,25], Jimeng Sun[26], Jian Tang[9,27,28], Petar Veličković[18,29], Max Welling[30,31], Linfeng Zhang[32,33], Connor W. Coley[5,34], Yoshua Bengio[9,35] & Marinka Zitnik[20,22,36] ✉

Artificial intelligence (AI) is being increasingly integrated into scientific discovery to augment and accelerate research, helping scientists to generate hypotheses, design experiments, collect and interpret large datasets, and gain insights that might not have been possible using traditional scientific methods alone. Here we examine breakthroughs over the past decade that include self-supervised learning, which allows models to be trained on vast amounts of unlabelled data, and geometric deep learning, which leverages knowledge about the structure of scientific data to enhance model accuracy and efficiency. Generative AI methods can create designs, such as small-molecule drugs and proteins, by analysing diverse data modalities, including images and sequences. We discuss how these methods can help scientists throughout the scientific process and the central issues that remain despite such advances. Both developers and users of AI tools need a better understanding of when such approaches need improvement, and challenges posed by poor data quality and stewardship remain. These issues cut across scientific disciplines and require developing foundational algorithmic approaches that can [illegible] scientific understanding or acquire it autonomously, making [illegible] for AI innovation.

The foundation for forming scientific insights and theori[illegible]ologies, from physical instruments such how data are collected, transformed and understood. T[illegible]hniques such as bootstrapping, have learning in the early 2010s has significantly expa[illegible]ns[1]. The introduction of digitization in ambition of these scientific discovery processe[illegible]general use of computing in scientific (AI) is increasingly used across scientific di[illegible]nce since the 2010s has enabled AI to sive datasets, refine measurements, gu[illegible] identifying scientifically relevant par- the space of theories compatible with[illegible] and reliable models integrated w[illegible]ctices and procedures vary across stages rnous discovery.

Data collection and analysi[illegible]he development of AI algorithms cuts across ing and discovery, ra[illegible] disciplines (Fig. 1). Such algorithms can enhance [illegible]ecution of scientific studies. They are becoming

[1]Department of Eng[illegible]es, California Institute of Technology, Pasadena, CA, USA. [2]Department of Co[illegible]Computer Science, Cornell University, Ithaca, NY, USA. [3]Department of Chemical Engin[illegible]anford University, Stanford, CA, USA. [4]Department of Physics, Massachusetts Ins[illegible]idge, MA, USA. [5]Mila—Quebec AI Institute, Montreal, Quebec, Canada. [6]Univers[illegible]n University, Providence, RI, USA. [7]Data Science Institute, Brown University, Provide[illegible]ork, NY, USA. [8]Department of Astrophysical Sciences, Princeton University, Princet[illegible]ndi Center for Data Science, New York University, New York, NY, USA. [9]Google Dee[illegible]dical School, Boston, MA, USA. [11]Department of Systems Biology, Harvard Medical Sc[illegible]Palo Alto, CA, USA. [13]BioMap, Beijing, China. [14]Mohamed bin Zayed University of Artific[illegible]m, IL, USA. [17]HEC Montréal, Montreal, Quebec, Canada. [18]CIFAR AI Chair, Toronto, Ont[illegible]K. [19]University of Amsterdam, Amsterdam, Netherlands. [20]Microsoft Research Amsterd[illegible]artment of Electrical Engineering and Computer Science, Massachusetts Inst[illegible]r Institute for the Study of Natural and Artificial Intelligence, Harvard University, [illegible]outh San Francisco, CA, USA. [36]Present address: Department of Computer Science, [illegible]ingi Du. ✉e-mail: marinka@hms.harvard.edu



## KEYNOTE SPEAKERS

# MUSHTAQ BILAL, PHD.

# EWA PLUCIENNICKA, PHD.

## OCTOBER 6TH

## MILAN HAŠEK AUDITORIUM

IMG

EASTPORT LIFESCIENCE

sotio

DIANA BIOTECHNOLOGIES

Bene Meat Technologies

Trigon plus®
...řešení pro vaši laboratoř

ThermoFisher SCIENTIFIC

exbio