Advances in Neuroethics

Series Editors: V. Dubljević · F. Jotterand · R.J. Jox · E. Racine

Fabrice Jotterand
Marcello Ienca *Editors*

Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues

Springer

Advances in Neuroethics

Advances in neuroscience research are bringing to the forefront major benefits and ethical challenges for medicine and society. The ethical concerns related to patients with mental health and neurological conditions, as well as emerging social and philosophical problems created by advances in neuroscience, neurology and neurotechnology are addressed by a specialized and interdisciplinary field called neuroethics.

As neuroscience rapidly evolves, there is a need to define how society ought to move forward with respect to an ever growing range of issues. The ethical, legal and social ramifications of neuroscience, neurotechnology and neurology for research, patient care, and public health are diverse and far-reaching — and are only beginning to be understood.

In this context, the book series "Advances in Neuroethics" addresses how advances in brain sciences can be attended to for the benefit of patients and society at large.

More information about this series at http://www.springer.com/series/14360

Fabrice Jotterand • Marcello Ienca
Editors

Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues

Springer

Editors
Fabrice Jotterand
Medical College of Wisconsin
Center for Bioethics and Medical Humanities
Milwaukee, WI
USA

Institute for Biomedical Ethics
University of Basel
Basel
Switzerland

Marcello Ienca
Department of Health Sciences and Technology
ETH Zurich
Zürich, Switzerland

ISSN 2522-5677 ISSN 2522-5685 (electronic)
Advances in Neuroethics
ISBN 978-3-030-74187-7 ISBN 978-3-030-74188-4 (eBook)
https://doi.org/10.1007/978-3-030-74188-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Acknowledgments

The editors would like to thank the many people and institutions that have contributed to the realization of this edited volume. First of all, we are indebted to Prof. Dr. Mingyu Liang, Professor and Kohler Co. Chair in Physiology and Director of the Center of Systems Molecular Medicine (CoSMM) at the Medical College of Wisconsin, for his constant support in the conceptualization and preparation of this volume. We would also like to thank the CoSMM and the Swiss Academy of Medical Sciences for supporting the editorial work of this book. Further thanks go to the Center for Bioethics and Medical Humanities at the Medical College of Wisconsin and the Health Ethics and Policy Lab, Department of Medical Science and Technology, ETH Zurich, for providing us with an optimal academic environment during this editorial journey. A special mention goes to Sylvana Freyberg, and Deepika Devan and her team at Springer Nature for their outstanding editorial help. Last but not least, we would like to thank all authors for their excellent chapters, and for the constructive patience they exercised to complete their contributions during the worst pandemic in recent history.

Contents

About the Authors

Mark V. Albert is the director of the Biomedical AI Lab at the University of North Texas and holds a dual appointment in the Department of Computer Science and Engineering and the Department of Biomedical Engineering. He leverages machine learning to automate the collection and inference of clinically useful health information to improve clinical research. His projects in wearable sensor analytics have improved the measurement of health outcomes for individuals with Parkinson's disease, stroke, transfemoral amputations, and cerebral palsy. Current projects include video-based activity tracking and mobile robotic platforms, all to improve measures of clinical outcomes to support therapeutic interventions.

Julia Amann is a postdoctoral researcher at the Health Ethics and Policy Lab at the Swiss Federal Institute of Technology (ETH Zurich). She holds a PhD in Health Sciences and Health Policy from the University of Lucerne, Switzerland, and is currently co-leading the Technologies for Public Health special interest group of Public Health Schweiz. Her research focuses on the impact of digital technologies on the doctor-patient relationship and healthcare, more generally. As part of her work on the Horizon2020 project PRECISE4Q, Julia is investigating the opportunities and challenges of machine learning in stroke medicine with a particular focus on the ethical implications for research and clinical practice. Julia previously held a postdoctoral appointment at Swiss Paraplegic Research. She is also a member of the research committee of the International Association for Communication in Healthcare (EACH).

Emily E. Anderson is associate professor of bioethics and medical education at Loyola University Chicago's Stritch School of Medicine. She teaches courses in research ethics and responsible conduct of research to graduate and medical students. Her areas of interest and expertise include researcher and physician professionalism and misconduct, ethics and community engagement, research with vulnerable populations, informed consent, and institutional review board (IRB) policy. Dr. Anderson serves as associate editor for *Narrative Inquiry in Bioethics* and *Progress in Community Health Partnerships* and is coauthor of *100 Questions and Answers About Research Ethics* (2018, SAGE) with Amy Corneli, PhD.

Timothy Brown is currently a postdoctoral scholar working primarily on a National Institutes of Health–funded project on the effect of neurotechnologies on user agency. Further, he is a long-time contributor to the Center for Neurotechnology's (CNT) Neuroethics Thrust—where he supports efforts to teach neuroethics to young investigators, catalyze ethics investigations through interdisciplinary collaborations, and promote the field of neuroethics through public outreach. More generally, Tim's work lies at the intersection of biomedical ethics, philosophy of technology, (black/latinx/queer) feminism, and aesthetics.

Ishan Dasgupta is a postdoctoral scholar in the Department of Philosophy and the Center of Neurotechnology at the University of Washington. He works at the intersection of law, ethics, and public health policy as it relates to emerging technology. His past work has focused on ethical issues surrounding induced pluripotent stem cells, inclusion of pregnant women in biomedical research, and the use of tissue samples in genetics research. At UW, Ishan focuses on issues of agency in relation to neurotechnology.

Sara Goering is Professor of Philosophy and the Program on Ethics and has affiliations with the Department of Bioethics and Humanities, and the Disability Studies Program. In addition, she currently leads the ethics thrust at the UW Center for Neurotechnology. She teaches courses in bioethics, ethics, philosophy of disability, feminist philosophy, and philosophy of medicine. She also spends time discussing philosophy with children in the Seattle public schools, through her role as the Program Director for the UW Center for Philosophy of Children.

Sarah Graham is a postdoctoral fellow in geriatric mental health in the Department of Psychiatry and the Stein Institute for Research on Aging at UC San Diego. She conducts research using biosensors and artificial intelligence to better understand aging from a multimodal perspective involving mental, cognitive, and physical health and well-being with the ultimate goal of enabling older adults to live independently longer with a high quality of life.

Pim Haselager is Associate Professor at the Donders Institute for Brain, Cognition and Behaviour and the Department of Artificial Intelligence at the Radboud University, Nijmegen. He focuses on the ethical and societal implications of cognitive neuroscience and AI. He publishes in journals such as *Nature: Biotechnology*, *Science and Engineering Ethics*, *American Journal of Bioethics*, *Neuroethics*, *Journal of Cognitive Neuroscience*, and *Journal of Social Robotics*.

José Hernández-Orallo is Professor at the Universitat Politècnica de València, Spain, and Senior Research Fellow at the Leverhulme Centre for the Future of Intelligence, University of Cambridge, UK. He received a BSc and a MSc in Computer Science from UPV, partly completed at the École Nationale Supérieure de l'Électronique et de ses Applications (France), and a PhD in Logic with a doctoral extraordinary prize from the University of Valencia. His academic and research activities have spanned several areas of artificial intelligence, machine learning, data science, and intelligence measurement. He has published five books and more than two hundred journal articles and conference papers on these topics. His research in the area of machine intelligence evaluation has been covered by several popular outlets, such as *The Economist*, *New Scientist*, or *Nature*. His most recent book addressed an integrated view of the evaluation of natural and artificial intelligence (Cambridge University Press, 2017, PROSE Award 2018).

Hudlicka is a member of the International Society for Research on Emotion, the Society for Affective Science, the Association for the Advancement of Artificial Intelligence, and the Coalition for Technology in Behavioral Science. She was a member of the National Research Council committee on "Behavioral Modeling and Simulation" and is an Associate Editor of the *International Journal of Synthetic Emotions* and a member of the Editorial Board of the *Journal of Cognitive Systems Research* and the *Oxford Series on Cognitive Models and Architectures*. She has authored over 50 journal and conference papers and numerous book chapters.

Dr. Hudlicka was born in Prague, Czech Republic, and received her BS in Biochemistry from Virginia Tech, MS from The Ohio State University in Computer Science, PhD in Computer Science from the University of Massachusetts-Amherst, and MSW from the Simmons School of Social Work.

Eva Hudlicka has a dual career combining research in affective computing and clinical work in psychotherapy. She is a Principal Scientist at Psychometrix Associates, which she founded in 1998 to pursue research in computational models of emotion, and since 2014 she is also a psychotherapist in private practice in Amherst, MA. During the 2017/2018 academic year she was a Fulbright Canada-Palix Foundation Distinguished Research Chair at the University of Alberta, Edmonton, and University of Lethbridge, Lethbridge, exploring the applications of affective computing in behavioral health technologies. Between 2013 and 2018 she was a Visiting Lecturer at the College of Information and Computer Sciences at the University of Massachusetts-Amherst, where she taught courses in affective computing and human-computer interaction. Prior to founding Psychometrix, she was a Senior Scientist at BBN in Cambridge, MA.

Marcello Ienca is a Senior Researcher in the Department of Health Sciences and Technology at ETH Zurich, Switzerland. His research focuses on the ethical, legal, and social implications of neurotechnology and artificial intelligence, with particular focus on big data trends in neuroscience and biomedicine, human-machine interaction, social robotics, digital health, and cognitive assistance for people with intellectual disabilities. He is interested in comparative approaches to the study of human and artificial cognition. Ienca is the Principal Investigator of multidisciplinary federal research projects and has received several awards for social responsibility in science and technology such as the *Prize Pato de Carvalho* (Portugal), the *Vontobel Award for Ageing Research* (Switzerland), and the *Paul Schotsmans Prize* from the *European Association of Centres of Medical Ethics* (EACME). He has authored one monograph, one edited volume (*Intelligent Assistive Technologies for Dementia*, *Oxford University Press*, 2019), +50 scientific articles in peer-reviewed journals, several book chapters, and is a frequent contributor to *Scientific American*. His research was featured in academic journals such as *Nature Medicine*, *Nature Biotechnology*, *Neuron*, *the Lancet Digital Health*, and the *American Journal of Bioethics* and media outlets such as *Nature, The New Yorker, The Guardian, The Times, Die Welt, The Independent, the Financial Times*, and others. Ienca is serving as appointed member or expert advisor in a number of national and international governance bodies including the Steering Group on *Neurotechnology and Society* of the Organisation for Economic Co-operation and Development (OECD) and the Council of Europe's Ad Hoc Committee on Artificial Intelligence. He is also a former Board Member of the International Neuroethics Society.

Fabrice Jotterand is Professor of Bioethics and Medical Humanities and Director of the Graduate Program in Bioethics at the Center for Bioethics and Medical Humanities, Medical College of Wisconsin. He holds a second appointment as Senior Researcher at the Institute for Biomedical Ethics at the University of Basel, Switzerland. His scholarship and research interests focus on issues including neuroethics including the ethics of AI in medicine, ethical issues in psychiatry and mental health, the use of neurotechnologies in psychiatry, medical professionalism, neurotechnologies and human identity, and moral and political philosophy. He has published more than 65 articles and book chapters as well as reviews and edited five books. His present research focuses on an examination of the ethical, regulatory, and social issues arising from the use of emerging neurotechnologies in psychiatry and neurology.

Eran Klein is a neurologist specializing in dementia at Oregon Health and Sciences University (OHSU) and the Portland VA Medical Center. He is part of the Neuroethics thrust at the NSF Center for Sensorimotor Neural Engineering (CSNE) at the University of Washington. He works at the intersection of neurology, neuroscience, and philosophy.

Karola Kreitmair is an assistant professor in bioethics at the University of Wisconsin—Madison. She completed a PhD in philosophy and a clinical ethics fellowship at Stanford University. Her research includes neuroethics, in particular issues surrounding consciousness, digital behavioral technology, and citizen science.

David D. Luxton is a nationally recognized expert and trainer in suicide prevention, telehealth, and innovative technologies in behavioral healthcare. He is Director of Research and Data Analytics, Washington State Department of Corrections, and Associate Professor in the Department of Psychiatry and Behavioral Sciences at the University of Washington School of Medicine in Seattle. Dr. Luxton previously served as a Research Health Scientist at the Naval Health Research Center in San Diego, CA, and Research Psychologist and Program Manager at the National Center for Telehealth and Technology (Defense Health Agency), Joint Base Lewis-McChord. A seasoned researcher, he has authored more than 100 scientific articles and book chapters and published three books: *Artificial Intelligence in Behavioral and Mental Health Care* (2015), *A Practitioner's Guide to Telemental Health* (2016), and *Behind the Machine* (2020). He has also helped to develop national guidelines for telemental health and clinical best practices in the use of technology in behavioral healthcare. In 2015 he was awarded the American Psychological Association Division 19 (Military Psychology) Arthur W. Melton Award for Early Career Achievement. He is a licensed clinical psychologist and served in the U.S. Air Force.

Gary Marchant teaches, researches, and speaks about the governance of a variety of emerging technologies, including biotechnology, genomics, neuroscience, nanotechnology, artificial intelligence, and blockchain. Prior to joining Arizona State University in 1999, he was a partner at the Washington, D.C., office of Kirkland & Ellis, where his practice focused on environmental and administrative law. He received his JD from Harvard Law School, where he was awarded the Fay Diploma (awarded to top graduating student at Harvard Law School). He also has a PhD in genetics from the University of British Columbia and a Master of Public Policy from the Kennedy School of Government. Professor Marchant frequently lectures about the intersection of law and science at national and international conferences, including speaking at over 75 judicial conferences. He has authored more than 200 articles and book chapters on various issues relating to emerging technologies. Among other activities, he has served on six National Academy of Sciences committees, has been the principal investigator on several major grants, and has organized over 50 academic conferences and workshops on law and science issues. He is an elected lifetime member of the American Law Institute and Fellow of the Association for the Advancement of Science.

Nicole Martinez-Martin is an assistant professor at Stanford University's Center for Biomedical Ethics, with a secondary appointment in the Department of Psychiatry. She is conducting research for an NIMH grant-funded study of the ethics of digital mental health technologies, and her areas of scholarship include neuroethics, digital health, and artificial intelligence in healthcare.

Giulio Mecacci is Assistant Professor at the Donders Institute for Brain, Cognition and Behaviour, and the Department of Artificial Intelligence at the Radboud University, Nijmegen. He investigates the impact of intelligent technologies and neurotechnologies on human values such as responsibility and privacy.

Camille Nebeker is Associate Professor of Behavioral Medicine in the Department of Family Medicine and Public Health, School of Medicine, UC San Diego. Her research and teaching focus on two intersecting areas: (1) research capacity building (e.g., participant-led, community-engaged research) and (2) digital health research ethics (e.g., consent, risk/benefit, data management). She directs the Research

Center for Optimal Digital Ethics (ReCODE.Health) and is affiliated faculty with the UCSD Design Lab, the Stein Institute for Research on Aging, and the Center for Wireless and Population Health Systems. Dr. Nebeker has received continuous support for her research from government, foundation, and industry sources since 2002.

Yair Neuman is a Full Professor at the Department of Cognitive and Brain Sciences and a member of the Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev. He received his BA in Psychology (Major) and Philosophy (Minor) and his PhD in Cognition (Hebrew Univ. 1999), and his expertise is in interdisciplinary research where he draws on diverse disciplines to address problems from an unusual perspective. More specifically, his expertise is in studying complex textual-symbolic, social, psychological, and cognitive systems, with a specific emphasis on the development of novel research methodologies and computational models. Prof. Neuman has published numerous papers and six academic books and was a visiting scholar/Prof. at M.I.T, University of Toronto, University of Oxford, and Weizmann Institute of Science. Beyond his purely academic work, he developed state-of-the-art algorithms for social and cognitive computing such as those he developed for the IARPA metaphor project (ADAMA group) and for his innovative work in computational personality analysis.

Emma M. Parrish is a graduate student in the San Diego State University/University of California San Diego Joint Doctoral Program in Clinical Psychology, and a T32 Predoctoral Fellow. Emma's research interests lie in real-time interventions and assessments through technology for people with serious mental illness, with a focus on suicide prevention, functioning, and cognition, as well as the ethics of digital mental health interventions.

Andreas Schönau is a postdoctoral scholar in the Department of Philosophy and Center for Neurotechnology at the University of Washington. His past research focused on the clarification of conceptual theories and empirical methods in philosophical and neuroscientific research, the interdisciplinary combination of their respective insights, and the generation of conclusions towards understanding the phenomenon of free will from an action-theoretical perspective. At UW, he continues working on agency-related issues in the intersection of Neuroscience and Philosophy.

Naveen Shamsudhin was born in Kerala, India. He received his BTech in Instrumentation and Control Engineering from the National Institute of Technology—Tiruchirappalli and his MSc in Micro and Nanosystems from ETH Zurich. He is a recipient of the NITT Alumni Award (2009) and the Swiss Government Scholarship (ESKAS 2009–2011) for excellence in undergraduate and graduate studies. He gained experience in the design, development, and quality control of MEMS technology through internships at the Indian Institute of Science (Bangalore) and at the Automotive Electronics division of Robert Bosch GmbH (Reutlingen). Working in close association with IBM Research Zurich and the

Institute of Plant Biology at the University of Zurich, he developed micro- and nanorobotic tools for single cell mechanics, completing his doctoral degree at the Multi-Scale Robotics Lab in January 2017. He was awarded the Prix OMEGA scientifique 2018 for his doctoral thesis. He joined the Multi-Scale Robotics Lab as a postdoctoral researcher in November 2017 and currently is a lecturer for two courses, Introduction to Robotics and Mechatronics (Spring Semester) and Microrobotics (Fall Semester). He is also the co-founder of The Origin AG. His current academic interests are the history and philosophy of technology in particular robotics and artificial intelligence (e.g., www.roboethics.ch), engineering for the developing world, and revitalizing inter-(trans-) disciplinarity in academic education and teaching (e.g., ETH Cortona Week, Roboethics Workshop, Kaleido).

Ryan Spellecy is the Ursula von der Ruhr Professor of Bioethics in the Center for Bioethics and Medical Humanities at the Medical College of Wisconsin, where he chairs one of the IRBs. His work focuses on research ethics, informed consent, ethical issues in psychiatry, and community involvement in research. He advised the Patient-Centered Outcomes Research Institute regarding engaging stakeholders in the peer review process, the Association of American Medical Colleges on IRBs, and community-based research. Recently, he was the co-PI for an NIH-funded national study evaluating a novel, easier to read consent form for blood and marrow transplant trials and currently leads a project to create a community-engaged research ethics training program as well as a study to evaluate the strengths and barriers regarding cancer clinical trial participation in African American churches.

Mónika Sziron is a Hungarian-American technology and humanities PhD candidate at Illinois Institute of Technology in Chicago, Illinois. Generally, her research focuses on the influence of various technologies on our daily lives today and throughout history. More recently, her PhD research focuses on how AI influences our daily lives, specifically considering moral status, ethics, and human rights issues and considerations in AI and robotics.

Lucille Nalbach Tournas researches and lectures in the intersection of law and emerging technologies, with an emphasis on artificial intelligence, big data, and neurotechnologies. She received her JD from the Sandra Day O'Connor College of Law at Arizona State University, where she was awarded the Strouse Prize for excellence in law, science, and technology. She is currently a PhD student in the School of Life Sciences at Arizona State University, where she is working on the global governance of neurotechnology and the data it provides.

Karina Vold is Assistant Professor at the Institute for the History and Philosophy of Science and Technology and a Faculty Affiliate at the Centre for Ethics, University of Toronto. She was previously a Research Fellow at the Leverhulme Centre for the Future of Intelligence, University of Cambridge, UK, and a Digital Charter Fellow at the Alan Turing Institute, the UK's national center for data science and artificial intelligence. Vold has been a visiting scholar at the University of Oslo, Ruhr

University, Duke University, and the Australian National University. She received an Honors BA in Philosophy and Political Science from the University of Toronto and a PhD in Philosophy from McGill University. Her current research spans across topics in philosophy of cognitive science, the ethics of artificial Intelligence, and the limits of machine learning.

Ting Xiao is a research assistant professor in the Department of Computer Science and Engineering at the University of North Texas. She was formally a postdoctoral researcher in experimental particle physics at Northwestern University applying statistical and computational tools to extract meaningful signals from large (~10 TB) data sets. In recent years she leveraged those skills to a variety of projects in computer science applications, generally applying signal processing and machine learning to video, audio, and wearable sensor data. She has published numerous papers with over 1800 citations. Her most cited individual effort involves the first observation of a new particle—Zc0 (3900).

Jie Yin is Associate Professor at the School of Philosophy and Center for Biomedical Ethics, Fudan University in China. She works on a wide range of topics in bioethics, philosophy of medicine, and Kant. She received training from medical school (B.M., Fudan University) as well as philosophy department (MPhil, Fudan University; PhD, SUNY Albany). Dr. Yin teaches undergraduate and graduate courses on Kant's *Critique of Practical Reason*, political philosophy, just health, bioethics, neuroethics, and nursing philosophy. Recent publications in Chinese include several articles on neuroethics and a textbook on philosophy of medicine.

1 Introduction

Fabrice Jotterand and Marcello Ienca

Artificial intelligence (AI) has the potential to transform the delivery and management of health care and improve biomedical research. Brain and mental health could significantly benefit from this technological transformation. Some of the most promising applications of AI in brain and mental health include the use of deep learning algorithms for early detection and diagnosis, as well as automated learning and the infusion of AI capabilities in everyday technologies such as smartphones, assistive social robots, and intelligent assistive technologies for continuous health monitoring and screening (e.g., Alzheimer's disease and schizophrenia) or for the assistance of psychogeriatric and neurorehabilitation patients. In addition, machine learning (ML) can also be used to improve existing neuropsychiatric therapies and allow new indications for existing drugs and tailor them to the individual patient through precision medicine approaches. For example, Watson, an AI-driven question-answering computing system developed by IBM, has proven to make similar treatment recommendations as human experts in 99% of the cases, and in 30% of the cases, Watson found treatment options missed by human physicians [1]. In addition, Watson can perform tasks such as data integration and aggregation, assessment of patients' risk to develop a particular disease or to require high cost treatment [2].

Further, big data analytics can be helpful to improve the epistemic power of neuropsychological explanations and unlock the etiology of brain and mental disorders by revealing relevant patterns across big and heterogeneous data volumes. In particular, multidimensional models integrating multiple biomarker data—for

F. Jotterand (✉)
Medical College of Wisconsin, Milwaukee, WI, USA

Institute for Biomedical Ethics, University of Basel, Basel, Switzerland
e-mail: fjotterand@mcw.edu

M. Ienca
Department of Health Sciences and Technology, ETH Zurich, Zürich, Switzerland
e-mail: marcello.ienca@hest.ethz.ch

F. Jotterand, M. Ienca (eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*, Advances in Neuroethics,
https://doi.org/10.1007/978-3-030-74188-4_1

example, neuroimaging biomarkers and digital phenotyping data—could help scientists overcome current reductionist approaches based on single explanatory neurobiological hypotheses. The automation of healthcare management processes via intelligent software to optimize healthcare delivery and reduce administrative cost is another promising implementation of AI technology.

The transformative potential of AI in brain and mental health does not limit to transforming the mode of generating scientific knowledge or assisting medical decision-making. In addition to that, it also portends to transform social and professional practices. For example, AI could redefine the therapeutic relationship. A study performed by researchers from the Dartmouth-Hitchcock health system, the American Medical Association (AMA), Sharp End Advisory, and the Australian Institute of Health Innovation revealed that physicians spend on overage 27% of their total time on direct clinical face time and 49.2% of their time on administrative work and Electronic Health Records (EHRs) [3]. The incorporation of AI in medical practice could help clinicians spend more time with patients and make health care more personal, albeit using more technology [4].

Such promissory outlook, however, has not materialized yet, at least, not entirely. The deployment of AI in neurology, psychiatry, neuropsychology, and brain research is still limited to sparse domains of application, often with suboptimal outcomes. Whether AI will re-humanize or de-humanize health care remains an open question as it is too early to understand the real impact long term of AI on clinical practice [5]. It is therefore paramount to cast light on emerging AI approaches in brain and mental health and provide an anticipatory impact assessment, with special focus on the assessment of emerging technical, scientific, ethical, and regulatory challenges. Such assessment is needed not only to chart the route ahead for scientific innovation in this domain but also to appraise such innovative dynamics within its broader socio-cultural and regulatory context. A broad spectrum of philosophical, ethical, regulatory, and social implications is rapidly emerging at the cross-section of AI and brain and mental health. Many of these implications have not been assessed in a comprehensive and systemic way. To this end, this unique volume provides an interdisciplinary collection of essays from leaders in various fields to address the current and future challenges arising from the implementation of AI in brain and mental health.

The volume is structured according to three main sections, each of them focusing on different types of AI technologies. Part I, *Big Data and Automated Learning: Scientific and Ethical Considerations*, specifically addresses issues arising from the use of AI software, especially machine learning, in the clinical context or for therapeutic applications. In Chap. 2 ("Big Data in Medical AI: How Larger Datasets Lead to Robust, Automated Learning for Medicine"), Ting Xiao and Mark V. Albert review the implications of the use of vast data sets in the context of medical research and clinical practice. They show how machine learning strategies can assist clinicians in various ways such as helping in the process of automatizing data selection for better diagnosis, improving the predictive power of statistical models tailored to specific hospitals or patient groups, or establishing the factor(s) that explains symptoms. However, Xiao and Albert point out that the collection of massive data sets is not without challenges such as data security, the interpretation and validation of data, and the accuracy of automated decision-making. In Chap. 3 ("Automatic Diagnosis and Screening of Personality Dimensions and Mental Health Problems"), Yair Neuman likewise addresses issues related to automatic diagnosis and screening but in the context of personality research. Computational Personality Analysis, as Neuman puts it, refers to the use of machine learning algorithms to measure variables in personality dimensions and disorders. As one can expect, such approach for the diagnosis of mental disorders or antisocial behaviors must be scientifically valid, ethically safe, and pragmatically relevant. So while "the promise of computational personality analysis is huge," Neuman concludes that the implementations of such technologies must be sensitive and critical to some of its challenges such as a good understanding of the complexity of human personality in light of the fact that automatic analysis of personality relies on "low-level features" in its categorization of personality. The other challenge is the fact that personality is a cluster of dynamic phenomena difficult to capture without a clear sense of the trajectory of the mental state captured. In Chap. 4 ("Intelligent Virtual Agents in Behavioral and Mental Healthcare: Ethics and Application Considerations"), David Luxton and Eva Hudlicka provide an overview of embodied Intelligent Virtual Agents (IVAs) and non-embodied conversational agents and examine the implications of their use in the context of behavior and mental health care. In particular, their analysis focuses on concerns about risks associated with the breach of privacy, the safety of individuals interacting with IVAs, and the ethical issues arising from artificial relationships. In Chap. 5 ("Machine Learning in Stroke Medicine: Opportunities and Challenges for Risk Prediction and Prevention"), Julia Amman examines issues related to the use of risk prediction and prevention tools such as novel machine learning-driven methods to reduce the global burden of stroke (incidence and mortality rates). There are many advantages for physicians and researchers to use such approaches as the increased accuracy of their predictions allow them to suggest interventions tailored to the specific needs of patients predisposed to strokes. But the implementation of such technology is not without challenges and limitations. These include issues of data sourcing, application development, and implementation in clinical setting, which, in Amman's estimation, should be fully recognized and addressed in order to benefit maximally from ML approaches to stroke predication and prevention. In the final chapter of the first section (Chap. 6, "Respect for Persons and Artificial Intelligence in the Age of Big Data"), Ryan Spellecy and Emily E. Anderson explore the extent to which traditional ways to honor respect for persons (in particular, informed consent) are challenged by AI and big data. In particular, they point out that in big data models where consent is not practicable due to the high data volume and velocity, waiving consent can be tempting for researchers for practical reasons but is ethically inadequate. They therefore argue that alternative approaches should be explored to hold the ethical standard of respect for persons. According to Spellecy and Anderson, "in discussions of ethics of AI and big data health research," there should be "less focus on the technical aspects of informed consent and more imagination regarding ways to demonstrate respect for persons" (p. 10 manuscript).

Part II, *AI for Digital Mental Health and Assistive Robotics: Philosophical and Regulatory Challenges*, examines philosophical, ethical, and regulatory issues arising from the use of an array of technologies beyond the clinical context. In Chap. 7 ("Social Robots and Dark Patterns: Where Does Persuasion End and Deception Begin?"), Naveen Shamsudhin and Fabrice Jotterand look at some of the challenges associated with the deployment of social robots for applications in areas such as entertainment, companionship, mental health, and well-being. The anthropomorphic design of these robots takes advantage of insights gained through human and social psychology, communication, and behavior which makes human beings vulnerable to manipulation and deception. Using digital media and web technologies, *dark patterns* are developed to deceive people to behave certain ways leading to addictive demeanor, hence undermining the autonomy of the users. The authors conclude that advances in robotics (i.e., social robots) should move forward but without the use of dark patterns. Nicole Martinez-Martin in Chap. 8 ("Minding the AI: Ethical Challenges and Practice for AI Mental Health Tools") directs her attention to fundamental questions of privacy, bias, and the potential impact of AI in the therapeutic relationship within the context of mental health. She contends that biases (i.e., systematic errors in a computer system that can cause unfair outcomes) may occur in the process of gathering data and health information and/or may depend on how algorithms are configured. These biases can cause inequities in the delivery of or access to mental health services. However, she also points out that the use of AI can be designed to address injustices. Martinez-Martin also examines how the implications of AI tools might affect the clinical encounter and provide recommendations for best practices. The use of digital behavioral technology (DBT) in combination with deep learning is the focus of Chap. 9 ("Digital Behavioral Technology, Deep Learning, and Self-Optimization"), authored by Karola Kreitmair. In her analysis, she considers technologies such as wearables, mobile health technologies, various smartphone apps, and noninvasive neurodevices that collect a large amount of data about individuals including brain activity, bodily functions, and behavioral patterns. Her analysis shows how the preferred way to process the data and make it relevant and useful for self-optimization (for instance, change of behavior through neurostimulation) is through an approach to AI known as deep learning. However, such technology presents many ethical challenges that are evaluated carefully by Kreitmair. In the next contribution, (Chap. 10, "Mental Health Chatbots, Moral Bio-enhancement and the Paradox of Weak Moral AI"), Jie Yin provides a philosophical exploration of the implications of the potential use of chatbots to enhance behavior in mental health. Hypothetically, her idea would be to use "a weak moral artificial intelligence" to enhance cognitive capacities, in particular moral deliberation. In principle, if such technology would be available, be safe, and respect human agency, it could be used for therapeutic purposes, although Yin argues, such approach would undermine essential elements of morality (such as motivation). However, she notes that mere philosophical argumentation is not sufficient for a final assessment of a weak moral artificial intelligence. Only once empirical evidence is available, we will be able to determine whether this type of technology ought to be implemented. In Chap. 11 ("The AI-Powered Digital Health

Sector: Ethical and Regulatory Considerations When Developing Digital Mental Health Tools for the Older Adult Demographic"), Camille Nebeker, Emma Parrish, and Sarah Graham examine the social benefits but also the potential ethical and regulatory pitfalls and risks associated with a widespread implementation of AI in day-to-day living, including "airline reservation systems, loan eligibility programs, college admissions, transportations systems, judicial decisions, and healthcare." In their analysis, they specifically focus on questions associated with the development of tools to help elderly people suffering from dementia which raise ethical questions regarding informed consent and agency. As more AI tools find their way into the marketplace and more data is collected, Nebeker et al. argue that new approaches to the governance of these technologies are needed in order to optimize their responsible implementation in the social context. Extra layers of protection should be put in place, particularly when dealing with vulnerable population such as elderly people with dementia. In Chap. 12 ("AI Extenders and the Ethics of Mental Health"), Karina Vold and José Hernandez-Orallo consider the extended mind thesis in the context of mental health and in light of AI technology. They examine the use of what they call "AI extenders" which is, in their view, different from previous cognitive extension based on simple technologies like a notebook or a smartphone. As they note, the "increased use of machine learning, and other functionalities brought by artificial intelligence, is importantly different from the kinds of cognitive extension that preceded it in many ways: these system can perceive, navigate, make complex decisions, understand and produce language, plan, understand emotions, etc., all in complex and changing situation". When applied to mental health to better diagnose and treat mental disorders, these technologies offer many opportunities to improve care but also raise many ethical challenges carefully outlined by Vold and Hernandez-Orallo.

In the final section of the volume, Part III entitled *AI in Neuroscience and Neurotechnology: Ethical, Social and Policy Issues*, contributions examine some of the implications of AI in neuroscience and neurotechnology and the regulatory gaps or ambiguities that could potentially hamper the responsible development and implementation of AI solutions in brain and mental health. The first contribution of this section by Pim Haslager and Giulio Mecacci (Chap. 13, "The Importance of Expiry Dates: Evaluating the Societal Impact of AI-Based Neuroimaging") analyzes the ethical and societal implications emerging from AI-powered neuroimaging. Such technology increases our ability to make predictive inferences about mental information and to recognize behavioral dispositions based on brain activity. However, Haselager and Mecacci argue that as more advances in AI-powered neuroimaging occur, further analysis must take place concerning the future implications of technologies for brain reading and the evaluative framework used in computational processing regarding neuroimaging. To this end, their contribution offers some fundamental recommendations for the regulation of the technology with a specific caveat: expiry dates for informed consent, data storage, and data analysis. In the next contribution, (Chap. 14, "Does Closed-Loop Deep Brain Stimulation for Treatment of Psychiatric Disorders Raise Salient Authenticity Concerns?"), Ishan Dasgupta, Andreas Schoenau, Tim Brown, Eran Klein, and Sara

Goering investigate issues associated with the new generation of deep brain stimulation (DBS) technology for the treatment of psychiatric disorders that employs artificial intelligence technologies as a means to "facilitate closed-loop implants that are adaptive and continuously modified by neural feedback". One major issue they examine is the impact of closed-loop DBS on authenticity. This chapter provides a salient empirical and philosophical analysis of the phenomenological implications of closed-loop neurostmulation for neuropsychiatric patients. Next, in Chap. 15 ("Matter Over Mind: Liability Considerations Surrounding Artificial Intelligence in Neuroscience"), Lucy Tournas and Gary Marchant address issues of liability. They recognize the benefits of the implementation of AI in the clinical setting for diagnostic and therapeutic purposes, but they also point out that there are risks and potential harms associated with the collection of neurological health data and an eagerness to deploy the technology without a careful consideration of liability concerns. They suggest building a "liability framework" that reconsiders informed consent in light of AI technology, increased education of physicians about AI, and an update of FDA regulations to include AI technology. In the last contribution of the volume (Chap. 16, "A Common Ground for Human Rights, AI and Brain and Mental Health"), Monika Sziron explores international regulations of AI in the context of health care and how human rights may be integrated in regulatory frameworks. The integration of human rights in international guidelines, however, is confronted to an important challenge: There are no agreed-upon international standards that regulate health care and AI. As she points out, "as philosophical and ethical environments vary across nations, subsequent policies reflect varying conceptions and fulfillments of human rights". She argues that despite this challenge, the development of ethical guidelines that encompass human rights may be possible at an international level if variations in their application and understanding are carefully acknowledged, which provide the common ground necessary to adapt policies and regulations. Finally, *the epilogue* ("Brains, Minds, and Machines: Brain and Mental Health in the Era of Artificial Intelligence") by Marcello Ienca concludes the volume by taking stock retrospectively of the work contained in this book and outlining the open challenges for future research in this field.

In light of its comprehensiveness and multidisciplinary character, this book marks an important milestone in the public understanding of the ethics of AI in brain and mental health and provides a useful resource for any future investigation in this crucial and rapidly evolving area of AI application.

References

1. Lohr S. IBM is counting on its Bet on Watson, and paying big money for it. The New York Times, October 17, 2016.
2. IBM Watson Health. Available online: https://www.ibm.com/watson/health/value-based-care/.
3. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med. 2016;165:753–60. https://doi.org/10.7326/M16-0961.

4. Vize R. Technology could redefine the doctor-patient relationship. The Guardian. March 11, 2017.
5. Jotterand F, Bosco C. Keeping the 'human in the loop' in the age of artificial intelligence: accompanying commentary for "correcting the brain?" by Rainey and Erden, Sci Eng Ethics. 2020. https://doi.org/10.1007/s11948-020-00241-1.

Part I

Big Data and Automated Learning: Scientific and Ethical Considerations

Big Data in Medical AI: How Larger Data Sets Lead to Robust, Automated Learning for Medicine

Ting Xiao and Mark V. Albert

2.1 Why the Big Data Revolution?

Machine learning is having a dramatic impact on the way we leverage information to make decisions [1, 2]. The success has been obvious in commercial business settings where data from advertising [3], supply logistics [4], and even social media [5, 6] is collected and processed in real time, enabling decisions at speeds and scales that would be impossible for hired employees. Medical applications present unique challenges due to risks but also provide satisfying targets due to the potential for improving health outcomes [7–10].

Many steps of the medical decision-making process can benefit from the tools of machine learning (Table 2.1). For example, we can consider a common sequence of choices made during the course of a medical treatment.

1. The clinician is tasked with collecting the relevant information.
2. A judgement about the cause is made based on the information available.
3. A treatment is proposed when possible.
4. Response to treatment is periodically evaluated and altered when needed.

T. Xiao
Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA

Department of Information Science, University of North Texas, Denton, TX, USA
e-mail: ting.xiao@unt.edu

M. V. Albert (✉)
Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA

Department of Biomedical Engineering, University of North Texas, Denton, TX, USA

Department of Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine, Evanston, IL, USA
e-mail: mark.albert@unt.edu

F. Jotterand, M. Ienca (eds.), *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*, Advances in Neuroethics,
https://doi.org/10.1007/978-3-030-74188-4_2

Table 2.1 Definitions

Artificial intelligence (AI)	The development of computer systems performing tasks commonly associated with intelligent beings, either through explicit programming or by learning from data
Machine learning	A large subset of AI which makes data-driven inferences. Notably, this is the area in which the vast majority of AI advances are made
Big data	A term to describe the tools and techniques of inference that are particular to large data sets, which enable more robust, automated learning
Deep learning	Machine learning using multilayer ("deep") neural networks. Currently the state of the art in solving challenging inference problems with large data sets by learning intermediate features directly from raw data
TensorFlow, PyTorch	The two dominant deep learning frameworks
GPU	Graphics Processing Unit. A processor designed to handle graphics operations that can be used to dramatically speed up neural network training due to the similarly simple, distributed processing needs

Medical professionals are trained to perform each of these steps taking into account what they observe directly or measure, and they then relate that information to their own personal experience and the medical research. However, it is worth noting that each of these steps can loosely be associated with a related approach used in machine learning techniques which are particularly valuable for large data sets and suggest recommendations for complex decision-making problems. For example, here we can list four machine learning strategies that can be directly mapped to the four steps above to assist the clinician in certain cases:

1. *Feature selection*: With enough data, the process of determining which information is more or less important can be automated. If the data is difficult or invasive to collect, a ranking of the importance can be provided to help the clinician choose the best measures to collect for a diagnosis [11].
2. *Factor analysis*: Notwithstanding the philosophical arguments of truly establishing cause and effect relationships, much of approach to understand a collection of symptoms is finding the underlying factor or factors explaining the symptoms presented. This goes well beyond disease diagnosis. Underlying factors may be more fine-grained than disease states, or emerge from comorbid diseases—a factor analysis would be able to identify groups of common concern in an automated way to allow patients with similar conditions to be grouped and treated more effectively [12].
3. *Predictive modeling*: The choice of treatment relies on the belief of which option is expected to lead to the greatest improvement, while weighing appropriate risks. Clinical researchers use statistical models to evaluate the superiority of one treatment over another, and in ambiguous cases, medical practitioners also use internal estimates of future improvement through their years of medical experience. However, with larger data sets, such predictions can be explicit and even tailored to the particular hospital, patient group, clinician, surgical technique using available data on past outcomes to provide an additional point of reference to help make a treatment recommendation [13].

4. *Automated outcome data collection and synthesis*: For long-term treatments, follow-up is necessary to judge compliance, efficacy, and make adjustments as needed. However, visits to the clinic are costly in terms of clinician time and associated financial costs. Questions regarding symptoms in a clinical visit can be subjective or incomplete, and physical measures may differ based on a variety of factors. Sensor technologies exist now which enable convenient, continuous, and objective measures of a variety of symptoms, with associated analytics to distill the measures to clinically relevant information [14].

In short, machine learning, and the associated use of large data sets to improve the process of learning, can augment the process of clinical decision-making. Such analytics provide a unique perspective for each decision. Notably, such tools perform a similar function to a secondary consult or collective review among clinicians, without the associated time, costs, or overhead—enabling rapid, often automated assistance to inform medical care.

2.1.1 More Samples, More Features

One of the reasons for the explosion of machine learning is the availability of data for training decision-making systems. The amount of data varies along two dimensions that are particularly relevant to learning systems—additional samples and additional features. Samples generally represent more examples or cases. Features, on the other hand, are new types of information that can be collected for each sample. Modern technology has made it possible to dramatically increase both dimensions of data to build learning models. More data enable systems to be more capable of automated decision-making.

To understand why this is the case, let us begin with a common rule of thumb for collected data to train many standard machine learning prediction models.

$$n_{\text{samples}} \gg \left(n_{\text{features}}\right)^2$$

That is, the number of samples collected should be substantially greater than the square of the number of features. Double the number of features, and so the number of samples has to be quadruped, etc. Note this is only a rough "rule of thumb" with many exceptions. This is not as critical for some simpler prediction algorithms (such as Naive Bayes), but it is reasonably accurate for a number of common machine learning models which are sufficiently flexible and powerful to learn for a wider variety of prediction problems. Why is this true? That is beyond the scope of this chapter, but some motivation is provided in the footnote.[1]

[1] Succinctly, the goal of machine learning is roughly stated as the ability to group similar sample points together in a n_{features} dimensional space. Most ways of flexibly grouping points in a n-dimensional space require more than n^2 parameters (groups of planes, multidimensional ellipses, etc.), and a well-known fact of estimation is that you generally need more data points than you