Edited by
Thomas Walker · Stefan Wendt · Sherif Goubran · Tyler Schwartz

Artificial Intelligence for Sustainability

Innovations in Business and Financial Services

palgrave macmillan

Artificial Intelligence for Sustainability

Thomas Walker • Stefan Wendt
Sherif Goubran • Tyler Schwartz
Editors

Artificial Intelligence for Sustainability

Innovations in Business and Financial Services

palgrave
macmillan

Editors
Thomas Walker
Department of Finance, John Molson School of Business
Concordia University
Montréal, QC, Canada

Sherif Goubran
Department of Architecture, School of Sciences and Engineering
The American University in Cairo
New Cairo, Egypt

Stefan Wendt
Department of Business
Bifröst University
Borgarnes, Iceland

Tyler Schwartz
Department of Finance, John Molson School of Business
Concordia University
Montréal, QC, Canada

ISBN 978-3-031-49978-4   ISBN 978-3-031-49979-1 (eBook)
https://doi.org/10.1007/978-3-031-49979-1

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

Paper in this product is recyclable.

Preface

The current global climate and sustainability crisis has pushed enterprises to reduce their harm and increase their positive impacts on the environment and society. Technology is seen as a critical facilitator for this transition. Researchers, practitioners, and politicians are increasingly recognizing the potential of artificial intelligence (AI) as a vital tool in this transition due to its ability to expand the analytical capacity for data – including sustainability-focused data. In past years, deep learning has proven effective in extracting patterns from big data by using neural networks. This technology has substantially impacted our abilities to forecast, measure, and manage risks and has resulted in a significant increase in the uptake of AI technologies in industry, government, and research.

In business, AI has been employed to improve operational efficiency and profitability through automation, sales forecasting, inventory and supply chain management, and the prediction of maintenance tasks. While the business interest in AI often focuses on its ability to increase profitability, AI also has the potential to help mitigate climate change by enhancing the ability of businesses and the public sector to behave more sustainably through the optimization of material and transport cycles in manufacturing and operations, as well as increased adaptation capacity through forecasting and simulations, among other areas.

Artificial Intelligence for Sustainability – Applications, Innovations, and Policies for Business is an edited collection that explores the theory and mechanisms behind sustainable AI innovations and their applications in business, investment, and regulatory processes and activities. While artificial intelligence has been the subject of many discussions, this book focuses on its application in a sustainability context, contributing to a significant gap in the literature. The editors have selected chapters that provide readers with an inclusive review and a critical analysis of new sustainability developments and innovations at the intersection between artificial intelligence and sustainable business practices and public administration. These contributions also investigate the potential negative consequences associated with AI, including the need for regulatory oversight, the reduction of human consciousness in decision-making, and the potentially negative externalities related to AI.

Acknowledgements

We acknowledge the financial support provided through the L. Jacques Ménard – BMO Centre for Capital Markets. In addition, we appreciate the excellent copy-editing and editorial assistance we received from Gabrielle Machnik-Kekesi, Victoria Kelly, Miles Murphy, Meaghan Landrigan-Buttle, Maya Michaeli, Sara Amin, and Rita Gabriel. Finally, we feel greatly indebted to Patricia Leduc, Kathleen Boies, Joseph Capano, Anne-Marie Croteau, and Norma Paradis (all at Concordia University) who in various ways supported this project.

Contents

About the Editors and Contributors

About the Editors

Thomas Walker is a full professor of finance, the director of the Jacques Ménard – BMO Center for Capital Markets, and the Concordia University Research Chair in Emerging Risk Management (Tier 1) at Concordia University in Montreal, Canada. He previously served as an associate dean, department chair, and as director of Concordia's David O'Brien Centre for Sustainable Enterprise. Prior to his academic career, he worked for Mercedes Benz, KPMG, and Utility Consultants International. He has published over 80 journal articles and books.

Stefan Wendt is a full professor and the dean of the Department of Business at Bifröst University, Iceland. Previously, he was associate professor at Reykjavik University, Iceland, and worked as a research and teaching assistant at Bamberg University, Germany. His research covers corporate finance, corporate governance, financial markets, sustainability and sustainable finance, digital finance, and behavioral finance. His research has been published in a number of journals including the *Journal of Financial Services Research*, the *Journal of Multinational Financial Management*, the *International Review of Financial Analysis, the Journal of Behavioral Finance*, and *Finance Research Letters*.

Sherif Goubran is an assistant professor of sustainable design in the Department of Architecture in the School of Sciences and Engineering at the American

University in Cairo, Egypt. Sherif's research focus includes building sustainability and sustainability assessment, sustainability in architectural design, and design for sustainable development.

Tyler Schwartz is a data scientist at Theory and Practice. He received his MSc in data science and business analytics from HEC Montreal. Tyler completed his undergraduate degree at the John Molson School of Business, receiving an honors in finance. His research interests include sustainability, predictive modelling, bio-statistics, and natural language processing.

Notes on Contributors

Djordje Batic received his B.Sc. and M.Sc. from the University of Novi Sad, Serbia, and is currently an MSCA-ITN fellow and a Ph.D. candidate at the University of Strathclyde, Scotland. His research interests include machine learning, deep learning, signal processing, and data analysis within energy-related applications.

Natalia Berg is Associate Professor of Accounting at the School of Business and Economics, Linnaeus University, Sweden. She previously held positions at Umeå University, Sweden. She received her Dr.Sc. degree in Accounting from Åbo Akademi University, Finland. Her research interests include sustainability accounting, performance measurement, corporate valuation, active ownership, and stakeholder governance. She conducts ESG research of companies, industries, and public authorities. Her research has been published in the *Journal of Applied Accounting Research*, the *Journal of Business Ethics*, *Nordic Journal of Business*, and *Sustainable Development*. Her research has received awards from the Sustainable Investment Forum in Finland.

Mostafa Gad-El-Rab is at the Capabilities and Insights Division of McKinsey & Co and has professional experience at Arab Banking Corporation and Falak Startups. He holds a BA in Economics (highest honors) with a minor in Computer Science from the American University in Cairo, Egypt, including a one-year visiting exchange at the University of Chicago, USA.

Camilla Della Giovampaola is a Ph.D. candidate in International History and Politics at the Geneva Graduate Institute (IHEID) in Switzerland. Her doctoral research focuses on regional philanthropic practices, with a particular focus on the contemporary development of institutional philanthropic organizations in the United Arab Emirates (UAE). Adopting an interdisciplinary approach, she explores how traditional practices of giving are shaped by institutionalized forms of giving, and vice versa. Her past experience as junior consultant for a Swiss-based philanthropic advisory firm also informs her research. Recently, she has been exploring the digitalization of the philanthropic sector, placing particular attention on how the philanthropic sector can benefit from artificial intelligence (AI) whilst simultaneously contributing to the development of Ethical and Inclusive AI (EIAI). She is also a Research Assistant for the AI & Philanthropy project initiated by the Geneva Centre for Philanthropy.

Lucia Gomez is a postdoctoral researcher specializing in the development of data-driven insights and solutions as part of the AI and Philanthropy research project initiated by the Geneva Centre for Philanthropy. Holding a bachelor's degree in Psychology and a Ph.D. in Neurosciences, her expertise spans multiple domains such as philanthropy, sustainability, behavior, cognition, brain development, and evolution. With a passion for leveraging AI in interdisciplinary research, she integrates cutting edge technologies to uncover novel insights and address complex challenges. Her current research focuses on promoting transparency, digitalization, and partnership formation in philanthropy. By mapping and reporting the current philanthropic landscape, she aims to drive meaningful transformations and empower the philanthropy of the future.

Hubert Halopé is a Ph.D. candidate in Management at the University of Geneva, Switzerland. His doctoral research focuses on responsible AI value creation across the private, public, and philanthropic sector. In addition, he supports the research team for the AI and Philanthropy project. He is also Lead for Artificial Intelligence & Machine Learning at the World Economic Forum, where he works on projects spanning emerging industry use cases for AI and technology policy and leads AI projects across the global network of Centres for the Fourth Industrial Revolution. He has experience in technology transformation, change management, and strategy consulting. He holds a dual Master's degree in International Management from the UCD Michael Smurfit Graduate Business School, Ireland, and CEMS, and a Bachelor in Economics from Maastricht University, Netherlands.

Andreas Kaplan is the president of Kuehne Logistics University, in Hamburg, Germany. His research addresses society's digital transformation, particularly the effects of artificial intelligence, social media, and virtual reality. A high-profile PLoS study counts Dr. Kaplan within the group of most cited and impactful scientists worldwide. His work has been published by Cambridge, Harvard, and

Oxford University Presses. Kaplan did his Habilitation at Panthéon-Sorbonne University and his Ph.D. at the University of Cologne in a joint program with HEC Paris. He holds an MPA from École Nationale d'Administration (ENA/INSP), an MBA certificate from The Power Business School, an MSc. from ESCP Business School, and a BSc from Ludwig Maximilian University of Munich. An internationally renowned higher education expert, he has extensive leadership experience in the academic world. Prior to becoming Kühne Logistics University (KLU)'s President and Director General, Kaplan served as Dean and Rector at ESCP Business School, Sorbonne Alliance.

Steve Lee is an architect and the founder of Aprilli Design Studio, a Toronto-based design consulting firm that focuses on research, design, and the implementation of sustainable strategies to enhance human life, society, and the surrounding environments. As an architect and designer, he is deeply interested in exploring the potential social transformations driven by future technologies, including Artificial Intelligence, Autonomous Driving Technology, Robotics, and Mixed Reality. His progressive design projects, such as Urban Skyfarm and Autonomous Travel Suite, have garnered significant attention from international media and received numerous international design awards, including the Radical Innovation Awards and A+ Design Awards. He earned a Bachelor of Science in Architecture from Yonsei University, South Korea, in 2002 and a Master of Architecture from Yale University, USA, in 2007. He previously instructed architectural design studios at various universities, including UCLA's March II program "Suprastudio" and California Polytechnic State University, San Luis Obispo.

Presha Neidermeyer is Professor of Accounting at Chambers College, West Virginia University, USA. She has a broad interest in issues in accounting and auditing and substantial knowledge relating to the potential of using new technology to solve fundamental problems in these fields. She has contributed substantially to the literature on auditor independence, sustainability accounting, and has done important work to improve women's rights in the USA and elsewhere. She has a long list of publications in highly reputable journals including, e.g., *Accounting and Business Research*, *The British Accounting Review*, *Accounting, Auditing & Accountability Journal*.

Peter Öhman is Professor of Accounting at The Mid Sweden University and director of the Center for Research on Economic Relations. He has a broad background of conducting research on fundamental issues in accounting and auditing and is one of the most published accounting researchers in Sweden, with publications in *Contemporary Accounting Research*, the *British Accounting Review*, and the *Accounting, Auditing & Accountability Journal*.

Won-Yong Oh is Lee Professor of Strategy and an associate professor at the Lee Business School, University of Nevada, USA. Previously, he was an associate professor at the Haskayne School of Business, Canada, and a visiting professor at the Faculty of Management, University of Warsaw in Poland. Before venturing into academia, he worked as a management consultant for A.T. Kearney and Deloitte Consulting in Korea. His research interests include corporate social responsibility, corporate governance, and strategic leadership. Currently, he is an associate editor of the *Journal of Business Research*.

Mark Perry has been Professor, Faculty of Law, University of New England, Armidale, Australia, since 2012. He is an academic leader, legal professional and computer scientist (Professor Emeritus) with strong governance, management, coordination, and negotiation skills. He was a member of the Academic Board of the University of New England from 2013 to 2021 and Council Member and Chair of the Board for three years. He is an experienced author, presenter, and author of over 200 publications in both law and technology. He is also an engaging keynote speaker at international conferences regarding technology and law, coordinator and organizer of global conferences in Canada, China, and Australia, advisor for international symposia, and serves as an advisor to governments on law and technology.

Tarek Rana is Associate Professor of Accounting at the Royal Melbourne Institute of Technology (RMIT) in Melbourne, Australia. Tarek investigates some of the most fundamental problems in accounting, is a guest editor of leading journals such as *Accounting, Auditing & Accountability* and the *British Accounting Review*, has received Australian awards for distinguished research, and is one of the most frequently published Australian researchers in accounting. Tarek is currently a research fellow a Gies College of Business, University of Illinois, USA, Urbana-Campaign.

Isak Samsten is a Senior Lecturer in Computer and Systems Science at Stockholm University, Sweden, and one of the leading researchers in machine learning in Sweden. He has researched machine learning issues and applications in several fields, such as clinical treatment in medicine and is currently involved in fundamental machine learning research looking at algorithm development. He is a leading expert on tree-based models and has published in a large number of machine learning journals and has presented at many conferences.

Tarek H. Selim has 30 years of experience in the fields of economic policy, business strategy, investment analysis, and development impact. He has been professor and former chair of economics at the American University in Cairo (AUC), Egypt, and is on the global expert council of the World Economic Forum, Davos, Switzerland. He has been visiting research professor at MIT, USA, and received Lifetime Research Fellow Award from the Economic Research Forum. He holds a Ph.D. in Economics from George Washington University, USA, am MBA from Johns Hopkins University, USA, an MS in Economics from Iowa State University, USA, an MS in Industrial Engineering, and a BS in Mechanical Engineering from AUC.

Throstur Olaf Sigurjonsson is a Professor at the University of Iceland. His areas of research are corporate governance, business ethics, policymaking, and policy implementation. He has published more than 90 articles, books, book chapters and cases. He is a permanent academic at Copenhagen Business School, Denmark where he earned his PhD. He also holds an MBA from IESE, as well as a BA in Philosophy and a Cand. Oecon. degree in Business Administration, both from the University of Iceland.

Darryl Robert Smith is a Consultant Electrical Engineer with practical experience and theoretical knowledge. His areas of expertise include wireless networks, radio systems, energy conservation and management, software, Linux, hardware, firmware, project management, systems engineering, critical thinking, cost engineering, and engineering information management. As a graduate, he spent nine months as a plant owner at a large coal-fired power station working on DC systems, switchboards, battery rooms, and more. During this time, he received authorization from the organization to work on high voltage and supervise others doing this. He then spent several years working on drawing and information management, transitioning to new technology, and architecting new systems.

Robert Brian Smith graduated with a BSc in 1969 and has since undertaken postgraduate studies in Education, Engineering, and Law. He holds a Ph.D. in engineering and is currently a Ph.D. candidate in law at the University of New England, Australia. His current research is entitled "IP Frameworks and Innovation in the Agricultural Sector: A Comparative Study of Western Europe and Southeast Asia." His career began as a government officer for 20 years, followed by nearly 25 years as an engineering consultant. Since April 2012, he has been a full-time international development consultant working in Southeast and South Asia. He has been a team leader in Bangladesh, India, Myanmar, Nepal, and Sri Lanka. He has spent shorter periods on projects in Cambodia, Egypt, Indonesia, Papua New Guinea, the Philippines, Thailand, and Timor Leste. This has given him unique insights into the development needs of least developed and developing economies. He has published over 120 refereed papers in engineering, international development, and law, particularly on Thai energy law, cybercrime, and intellectual property.

Lina Stankovic received a B.Eng. (Hons.) in electronic communications engineering and her Ph.D. from Lancaster University, U.K., in 1999 and 2003, respectively. She is a Reader at the University of Strathclyde, Scotland. Her main research interests include smart monitoring from sensor network platforms and meaningful information extraction that is user-centric, focused on algorithmic analysis of energy data to understand how people use appliances in the home.

Vladimir Stankovic received the Dr. Ing. Ph.D. from the University of Leipzig, Germany in 2003. From 2003 to 2006, he was with Texas A&M University, College Station, USA, first as a Research Associate and then as a Research Assistant Professor. From 2006 to 2007, he was with Lancaster University, U.K.. Since 2007, he has been with the Department of Electronic and Electrical Engineering at University of Strathclyde, Scotland, where he is currently a Professor. He has co-authored 5 book chapters and over 200 peer-reviewed research papers, including over 80 journal publications. His research interests include signal and information processing and machine learning, including graphical representation approaches for heterogeneous and highly dimensional signals and explainable artificial intelligence solutions with applications to smart home energy management, smart buildings and cities, computer vision, seismic signal analysis, and personal healthcare.

Jan Svanberg is Associate Professor at the University of Gävle, Sweden, and now Fulbright Scholar and Research Fellow at Chambers College, West Virginia University, USA. His research interest focuses on how to use technology to solve research problems in accounting and auditing. He has published on a broad variety of topics in journals spanning from *Behavioral Research in Accounting* to the *British Accounting Review*. He has led several projects to solve important problems by developing machine learning applications in cooperation with large corporations.

Lukasz Swiatek lectures in the School of the Arts and Media within the Faculty of Arts, Design, and Architecture at the University of New South Wales (UNSW) in Sydney, Australia. He draws on a wealth of practical industry experience in both his teaching and research. Over the years, he has taught a range of undergraduate and postgraduate (junior and senior) courses in media studies, communication, and international and global studies at universities in Australia and Aotearoa New Zealand. He mainly undertakes research in media and communication, higher education, and cultural studies.

Maria Cristiana Tudor is pursuing a Master of Research in Neurosciences at the University of Geneva, Switzerland, focusing on computational modelling of the human brain's learning mechanisms using reinforcement learning algorithms. She holds an interdisciplinary bachelor's degree from University College London, U.K., where she integrated classes in Mathematics, Engineering, Entrepreneurship, and Management. Driven by perpetual curiosity and interdisciplinary endeavors, her research interests lie at the intersection of technology, human behavior and cognition, and sustainability. Sample projects include the applications of behavioral sciences in purpose-driven companies (e.g., B Corps) to foster sustainable consumer behaviors or applying AI tools to research and advance philanthropic and sustainable purposes. She is currently a Research Assistant in the Behavioral Philanthropy Lab at the Geneva Finance Research Institute.

Giuseppe Ugazio holds two doctorates, one in philosophy and the other in neuro-economics, and studies the neuropsychological mechanisms that influence complex human social behavior. Using interdisciplinary experimental methods, he explores the affective, cognitive, and behavioral control mechanisms underlying social decision-making. In particular, his research aims to understand how these decision mechanisms interact and compete to guide choices related to moral values and philanthropy. More recently, his research has focused on understanding the role of AI for Philanthropy, for instance developing a machine learning approach to disclose motivations to engage in prosocial endeavors. He is currently the Edmond de Rothschild Foundations Chaired Assistant Professor in Behavioral Philanthropy at the Geneva Finance Research Institute at the University of Geneva, Switzerland.

Stefan Wendt is a Professor and the Dean of the Department of Business at Bifröst University, Iceland. He was an Associate Professor at Reykjavik University, Iceland, from 2015 to 2021. Prior to that that, he was a Research and Teaching Assistant at the Department of Finance at Bamberg University, Germany. His fields of interest include corporate finance, corporate governance, financial markets, sustainability and sustainable finance, digital finance, and behavioral finance. His research has been published in a number of journals, including the *Journal of Financial Services Research*, the *Journal of Multinational Financial Management*, the *International Review of Financial Analysis*, the *Journal of*

Behavioral Finance, the *International Journal of Managerial Finance*, and *Finance Research Letters*.

Irene Yi is an Assistant Professor of Finance at the Rotman School of Management, University of Toronto, Canada, specializing in corporate finance, corporate governance, and labor and finance. She earned her Ph.D. in Finance and Business Economics from the University of Southern California, USA, and an A.M. in Statistics from Harvard University, USA. Before joining Rotman in 2018, she served as a Visiting Assistant Professor of Finance at Purdue University, USA. Her research includes a study on labor unions' strategic use of proposals and an analysis of the effects of SEC no-action-letter decisions, published in the *Review of Financial Studies* and the *Journal of Law and Economics*, respectively. She applies her research insights to teaching Corporate Finance and Corporate Governance at both the undergraduate and Ph.D. level.

Copy Editing Team

Victoria Kelly graduated from Concordia University (Montreal) with a BSc in Biology with an additional major in Irish Studies. She is currently a first-year master's student in History and Irish Studies where her research focuses on the 1832 cholera epidemic and its management on a social, urban, economic and medical level, drawing parallels with the recent COVID-19 epidemic.

Meaghan Landrigan-Buttle holds a master's degree in history and Irish Studies from Concordia University (Montreal) where she is currently a second-year Ph.D. student. She also holds a professional copyediting certification from the University of Waterloo, Canada. Her two main projects focus on Irish identity, masculinities, and violence in the First World War, as well as their representation in contemporary Irish mafia dark romance novels.

Gabrielle Machnik-Kekesi is a Research Associate at the Emerging Risks Information Centre at the John Molson School of Business at Concordia University, Canada, and is a PhD Candidate at the University of Galway, Ireland. She holds an Individualized Program master's degree from Concordia University, which was funded by both the Social Sciences and Humanities Research Council and the Fonds de recherche du Québec en Société et Culture, and a master's degree in Information Studies from McGill University.

List of Figures

List of Tables

1

Artificial Intelligence for Sustainability: An Overview

Thomas Walker, Stefan Wendt, Sherif Goubran, and Tyler Schwartz

1.1 Introduction

Businesses are expected to embrace sustainability to reduce their adverse effects on the environment and society while, at the same time, strengthening their organizations' positive impacts. This sustainability paradigm has pushed enterprises to investigate means to replace fossil fuels with cleaner energy sources, minimize their carbon footprint and waste,

T. Walker (✉) • T. Schwartz
Department of Finance, John Molson School of Business, Concordia University, Montréal, QC, Canada
e-mail: thomas.walker@concordia.ca; tyler.schwartz@mail.concordia.ca

S. Wendt
Department of Business, Bifröst University, Borgarnes, Iceland
e-mail: stefanwendt@bifrost.is

S. Goubran
Department of Architecture, School of Sciences and Engineering, The American University in Cairo, New Cairo, Egypt
e-mail: sherifg@aucegypt.edu

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_1

optimize their logistics, reduce the negative environmental and social externalities throughout their supply chains, and assess the impact of their sourcing and procurement practices (Hepburn & Bowen, 2013). Not sufficiently dealing with environmental and social or societal risks constitutes a substantial financial risk for corporations and investors (Ekins & Zenghelis, 2021), and research indicates that less sustainable companies are considered less creditworthy (Höck et al., 2020).

This edited volume explores how artificial intelligence (AI) can support or even drive the transition to more sustainable business practices. Even though developments in artificial intelligence have been ongoing for about seven decades, AI has attracted particular attention and investments in recent years. Total global corporate investment in artificial intelligence has increased more than sevenfold since 2015, reaching more than USD 90 billion annually in 2021 and 2022 (Statista, 2023). Based on sensory inputs, one of the main applications of artificial intelligence is prediction, for example, of human behavior. Judgment as an assessment of consequences is an area where artificial intelligence is advancing but where human judgment is still needed, e.g., when it comes to ethical judgment or emotional intelligence. Performing actions based on the corresponding decisions refers to robotics which has also seen huge advancement (Agrawal et al., 2017; for definitions of AI systems, see Russell & Norvig, 2016).

Potential applications of artificial intelligence are manifold. Based on the computerized ability to analyze, generalize, reason, and learn, artificial intelligence applications are used in automated decision-making, targeted advertising and recommendation systems, web search engines, virtual assistants and chatbots, supply chain management, the recognition of human speech and automatic language translation, image and facial recognition, medical diagnosis, and autonomous vehicles, to name a few (Agrawal et al., 2017; Brynjolfsson et al., 2019). Many of these artificial intelligence applications are based on machine learning, which uses statistical techniques to allow machines to "learn" with experience by correcting previous errors at each iteration. More recently, deep learning, a sub-domain of machine learning, has proven effective in extracting patterns from big data using neural networks inspired by the neurological basis of brain functioning. Primary applications of deep learning include image and voice recognition, text summarization and generation, and classification, amongst other functions (Dargan et al., 2020). More recently, large language models (LLMs), trained using large deep learning models, have also become very popular by introducing "intelligent" chatbots such as ChatGPT. While these LLMs have many ethical and practical implications in our modern society, they also have enormous environmental impacts resulting from the training and use of the models (Li et al., 2023; Patterson et al., 2022).

In the business sector, artificial intelligence has been successfully employed in many areas to improve the efficiency and profitability of operations, most notably through the automation of tasks that humans previously handled and by more accurately forecasting sales, e.g., via analyzing customer behavior, diagnosing when maintenance is needed for machinery, and managing inventory (Ahmed et al., 2022; Albayrak Ünal et al., 2023; Huang & Rust, 2021). A 2021 survey conducted by McKinsey shows that 50% more businesses adopted artificial intelligence in 2021 than in 2020 (McKinsey, 2021). Moreover, Purdy and Daugherty (2017) estimate that artificial intelligence has the potential to raise profitability by 38% by 2035 across a sample of 12 economies and 16 industries, which translates to about 14 trillion dollars added for these economies. Interestingly, though, Brynjolfsson et al. (2019) describe the paradox of potential productivity increase due to transformative new technologies on the one hand but reduced overall economy-wide productivity growth in recent decades on the other. This indicates that artificial intelligence has indeed a transformative – or even disruptive – character, which allows businesses that integrate these solutions to establish a competitive advantage while those that do not integrate artificial intelligence may disappear over the medium or longer run. Moreover, while implementing artificial intelligence requires large investments now, efficiency and profitability gains will only materialize in the years to come.

As the chapters in this edited volume show, artificial intelligence has an enormous potential to help mitigate climate change and further environmental and societal challenges by enhancing the ability of businesses to behave more sustainably. Using artificial intelligence can help reduce waste through more accurate demand prediction, improve the sustainability of manufacturing and operations through more precise monitoring, lower emissions across the supply chain by optimizing routes and transportation options, heighten transparency about operations through improved monitoring, and increase adaptation capacity through forecasting and simulations, among other areas (Walter, 2023).

Despite the benefits of adopting artificial intelligence business solutions to improve our environmental and social footprint, some downsides and challenges, including the ecological impacts of artificial intelligence itself, must be considered. Dauvergne (2022) points out that efficiency gains do not lead to a reduced use of resources. Instead, they result in more production and more natural resource extraction. Strubell et al. (2019) show that training a single large natural language processing (NLP) model has a carbon footprint equivalent to 125 round-trip flights from New York to Beijing. As artificial intelligence models continue to increase in complexity and scale, their carbon footprint will only continue to grow as long as cleaner energy sources are not used to a larger extent. Moreover, artificial intelligence might render some job profiles obsolete, which is a societal challenge that can increase the potential for unemployment in the short term, and, in the long term, would require significant changes in educational curricula to ensure the development of the required skill sets in employees entering the job market (Agrawal et al., 2017; Huang & Rust, 2018).

While the use of artificial intelligence has been the subject of many discussions, this book takes a more focused look at artificial intelligence applications in a sustainability context and thereby closes a significant gap in the literature. The contributions offer technical, practical, and policy insights into the different sustainable applications of artificial intelligence and highlight how these new developments may affect and contribute to our fight against climate change and further environmental and social challenges.

This edited book focuses on the positive opportunities of new artificial intelligence developments and critically explores and assesses the potential risks and challenges involved in their implementation. These include the ethics involved in collecting and storing data, the increased demand these technologies pose to our environment (land, natural resources, energy), and the large carbon footprint of these applications. Additionally, the socio-economic repercussions of these technologies are explored. The collection also explores the impact these artificial intelligence solutions have on addressing sustainable development and the 2030 Agenda with its Sustainable Development Goals (SDGs) (United Nations, 2015).

This collection contains contributions from an international community of scholars and practitioners at the interface of data science, artificial intelligence, sustainability sciences, supply chain management, public administration, business, and engineering. The contributions review and critically analyze new developments at the intersection between artificial intelligence and sustainable business practices and public administration, focusing on the sustainability benefits of applying artificial technologies to business, investment, and regulatory processes and activities. The book documents these developments, exposing the theory and mechanisms behind sustainable artificial intelligence innovations within business contexts.

1.2 Overview of Content

This book is composed of eleven chapters in total. After this introduction, the body of the book is composed of three parts – each with multiple chapters. The content of each of these parts is summarized below.

Part I: AI and Sustainable Industry Applications

The book's first part, *AI and Sustainable Industry Applications*, looks at applications of AI and sustainability in different industry contexts.

The part begins with Chap. 2, *Fast Fashion's Fate: Artificial Intelligence, Sustainability, and the Apparel Industry*, which discusses the potential of AI to promote sustainability in the fashion industry, one of the world's biggest polluters. **Kaplan** explains how AI can be applied to optimize supply chains and encourage environmentally friendly consumption, reducing the industry's carbon footprint. The chapter includes case studies of three fashion companies (H&M, Zara, and Farfetch) that have incorporated AI to enhance sustainability.

Chapter 3, *Artificial Intelligence and the Global Automotive Industry*, examines how the fourth industrial revolution and AI advancements will reshape the automotive industry's sustainability, especially regarding autonomous vehicles (AV). **Selim and Gad-El-Rab** discuss the role different key factors and considerations will play in this reshaping including AI-driven radar and laser tech, commercial risks, changing consumer behavior, city infrastructure adaptations, and societal response to technological shifts. Lastly, the chapter looks forward at what the most likely sustainable outcomes from the AV industry are.

Chapter 4, *The Emergence of the Nighttime Artificial Intelligence Robot-Driven Economy*, explores the rise of an AI and robot-driven autonomous economy. **Lee, Oh, and Yi** discuss that as these technologies take on roles from desk jobs to manual labor, businesses might lean towards self-sustaining systems, optimizing processes such as nighttime operations for energy savings and infrastructure ease which has many sustainability benefits. Lastly, the authors introduce the concept of a separate nighttime economy, powered solely by AI and robots, distinct from the traditional daytime human activities and how this change will play a role in the future of our society.

Part II: AI and Sustainable Business Operations

The book''s second part, *AI and Sustainable Business Operations,* looks at how AI is being leveraged in the operations of businesses to improve their sustainability footprint.

The part begins with Chap. 5, *Strengthening the Sustainability of Artificial Intelligence: Fostering Green Intelligence for a More Ethical Future*, which discusses the dual role of AI in the business sector—its utility in sustainability efforts and the ethical and environmental issues associated with its use, such as exploitative labor practices, discriminatory algorithms, and resource consumption. **Swiatek** advocates for the greening of bounding AI in the business sector to limit negative environmental impacts. This chapter warns that a failure to do so could lead to financial, legal, and reputational consequences. To illustrate the concept, a case study of NearMap, a company which has implemented and greened its bounded AI, is presented.

Chapter 6, *Predictive Machine Learning in Assessing Materiality: The Global Reporting Initiative Standard and Beyond*, delves into the challenges of sustainability reporting, specifically addressing the subjective nature of materiality assessments across various companies and sectors.

Svanberg, Öhman, Samsten, Neidermeyer, Rana, and Berg introduce a machine learning approach to standardize and predict materiality assessments, aligning with the GRI (Global Reporting Initiative) standards. Additionally, the chapter discusses the application of the SHAP (SHapley Additive exPlanations) method, offering an advanced perspective on estimating impact materiality using this tool.

Chapter 7, *Artificial Intelligence and the Food Value Chain*, looks at the potential for AI to enhance sustainability throughout the entire food value chain, from production to distribution. **Wendt and Sigurjonsson** highlight the use of AI for resource optimization, reducing fertilizer and drug use, enhancing efficiency in food processing and distribution, and minimizing food waste. They show that the benefits of AI mainly stem from continual monitoring and data analysis of crops, livestock, processed products, and storage and transportation conditions.

Part III: The Role of AI in Sustainable Development and the 2030 Agenda

The book's last part, *The Role of AI in Sustainable Development and the 2030 Agenda*, discusses how AI is being used in a business context to address issues illustrated in 2030 Sustainable Development Agenda.

The part begins with Chap. 8, *Analysis of Smart Meter Data for Energy Waste Management*, which discusses the application of AI in the analysis of smart meter data, a tool that promotes energy industry digitalization and lowers carbon emissions. **Batic, Stankovic, and Stankovic** discuss how smart meter data can facilitate the shift to renewable energy sources, optimize energy supply, and provide insights into home energy usage. However, the authors show how the fine-grained data analysis could reveal sensitive personal information, posing ethical challenges and potentially undermining public trust.

Chapter 9, *The Role of Artificial Intelligence in Uncovering Synergistic Partnerships between Philanthropic Actors Beneficial to the Achievement of the Sustainable Development Goals*, examines the potential role of AI, particularly NLP, in advancing the United Nations SDGs through enhancing partnerships among Philanthropic Organizations (POs). **Tudor, Gomez, Giovampaola, Halopé, and Ugazio** point out that a major obstacle to effective coordination between POs is insufficient knowledge and visibility of the activities aligned with the SDGs conducted by others. The proposed AI approach the chapter presents aims to map SDG coverage by POs in the Swiss ecosystem, thereby increasing visibility and facilitating the formation of partnerships, thus enhancing cooperation and resource pooling towards achieving the SDGs.

The book concludes with Chap. 10, *The Potential Role of Artificial Intelligence in the Commercialization of Traditional Medicines in Tropical Regions*, which explores how AI can revolutionize the process of identifying effective traditional and complementary medicines from marginalized communities in tropical regions for commercial uses**. Smith, Perry, and Smith** discuss how advancements made in AI, driven by enhanced code, investment, hardware, and accessible datasets, are harnessed to create a holistic approach to bring forward not only superior treatments rooted in traditional practices, but places focus on sustainability, fairness, and the conservation of biodiversity in vulnerable ecosystems.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2017, February 7). What to expect from artificial intelligence. *Sloan Management Review*. https://sloanreview.mit.edu/article/what-to-expect-from-artificial-intelligence/

Ahmed, I., Jeon, G., & Piccialli, F. (2022). From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where. *IEEE Transactions on Industrial Informatics, 18*(8), 5031–5042. https://doi.org/10.1109/TII.2022.3146552

Albayrak Ünal, Ö., Erkayman, B., & Usanmaz, B. (2023). Applications of artificial intelligence in inventory management: A systematic review of the literature. *Archives of Computational Methods in Engineering, 30*, 2605–2625. https://doi.org/10.1007/s11831-022-09879-5

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 23–57). University of Chicago Press.

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: A new paradigm to machine learning. *Archives of Computational Methods in Engineering, 27*(4), 1071–1092. https://doi.org/10.1007/s11831-019-09344-w

Dauvergne, P. (2022). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Review of International Political Economy, 29*(3), 696–718. https://doi.org/10.1080/09692290.2020.1814381

Ekins, P., & Zenghelis, D. (2021). The costs and benefits of environmental sustainability. *Sustainability Science, 16*(3), 949–965. https://doi.org/10.1007/s11625-021-00910-5

Hepburn, C., & Bowen, A. (2013). Prosperity with growth: Economic growth, climate change and environmental limits. In *Handbook on energy and climate change* (pp. 617–638). Edward Elgar Publishing. https://doi.org/10.4337/9780857933683.00041

Höck, A., Klein, C., Landau, A., & Zwergel, B. (2020). The effect of environmental sustainability on credit risk. *Journal of Asset Management, 21*, 85–93. https://doi.org/10.1057/s41260-020-00155-4

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*, 30–50. https://doi.org/10.1007/s11747-020-00749-9

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI less "thirsty": Uncovering and addressing the secret water footprint of AI models. arXiv preprint arXiv:2304.03271.

McKinsey. (2021). The State of AI in 2021. https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021

Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L. M., & Dean, J. (2022). The carbon footprint of machine learning training will plateau, then shrink. *Computer, 55*(7), 18–28.

Purdy, M., & Daugherty, P. (2017). How AI boosts industry profits and innovation. Accenture. https://www.accenture.com/fr-fr/_acnmedia/36dc7f76eab444cab6a7f44017cc3997.pdf

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.

Statista. (2023). Global Total Corporate Artificial Intelligence (AI) Investment from 2015 to 2022. https://www.statista.com/statistics/941137/ai-investment-and-funding-worldwide/

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of the 57th annual meeting of the Association for Computational Linguistics, pp. 3645–3650

United Nations. (2015). The 17 Goals. United Nations Department of Economic and Social Affairs. https://sdgs.un.org/goals

Walter, S. (2023). AI impacts on supply chain performance: A manufacturing use case study. *Discover Artificial Intelligence, 3*(1), 18. https://doi.org/10.1007/s44163-023-00061-9

Part I

AI and Sustainable Industry Applications

2

Fast Fashion's Fate: Artificial Intelligence, Sustainability, and the Apparel Industry

Andreas Kaplan

2.1 Silhouette: Chapter's Contour

The apparel industry is one of the world's biggest polluters. According to various reports, it is responsible for 10 percent of global carbon emissions; an additional 20 percent of global wastewater comes from textile dyeing, while cotton farming accounts for almost 25 percent of insecticide use and more than 10 percent of pesticides (Ellen MacArthur Foundation, 2017; United Nations Environment Programme, 2019). An examination of clothing consumption patterns further points to the unsustainability of mass-produced garments. Between 2000 and 2015, clothing sales doubled and reached 200 billion products annually, with the average number of times an item was worn decreased by nearly 40 percent overall (Ellen MacArthur Foundation, 2017).

Conscious of an increasing number of eco-responsible consumers who are aligning their clothing purchases with their commitment to

A. Kaplan (✉)
KLU - Kühne Logistics University, Hamburg, Germany
e-mail: andreas.kaplan@klu.org

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_2

sustainability, several brands are trying to become greener. For some, this amounts to mere greenwashing, whereas others have implemented meaningful changes. Due to a shift in societal preferences toward more sustainable production methods and consumption patterns, it is hardly imaginable and probably unrealistic to believe that the apparel industry could continue mass-producing cheap clothes, which are often produced under dubious labor conditions. Significant change is required to transform such a problematic industry, which currently is far away from compliance with the United Nation's Sustainable Development Goals (SDGs). Artificial intelligence's (AI) application to fast fashion may comprise a pathway towards greening mass clothing production, notably regarding SDG 12, which calls for Responsible Consumption and Production.

AI has made significant process since its inception roughly 70 years ago, and seems to have reached a point in its development where it can have concrete applications in business (Haenlein & Kaplan, 2019). Artificial intelligence has become operational on a broad scale and is – along with sustainability – certainly one of the hot topics of today's societal discourse. The COVID-19 pandemic accelerated and accentuated society's digitalization and gave AI a further push (Kaplan, 2022): Whereas, e.g., the Metaverse and virtual worlds (Kaplan & Haenlein 2009a, b) were targeted toward a niche market before the pandemic, the advantages of working remotely and organizing business meetings in AI-powered virtual environments became much more apparent during the various lockdowns. Consequently, Internet giants such as Facebook have massively invested in virtual reality technology and AI since. Fashion, by definition, is at the forefront of innovation; embracing AI in such a trend-driven industry should, therefore, come as no surprise.

This chapter explains how AI can improve the sustainable production and consumption of apparel items. To do so, it first gives an overview of AI, analyzes and decrypts its potential in addition to associated challenges and risks, and provides several examples from the retailing and clothing industry. Then, it shows how AI can be applied to the fashion sector to reduce its carbon footprint significantly. Furthermore, this chapter presents three case studies from two fast-fashion brands (H&M and Zara) and one luxury fashion retail platform (Farfetch) – all of which have begun applying artificial intelligence to their processes to increase sustainability. Finally, this chapter concludes with recommendations regarding the fate of fast fashion.

2.2 Pattern: AI and Algorithms

Artificial Intelligence is defined as "a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation" (Kaplan & Haenlein, 2019, p. 17). While AI is not a new technology, it remains a subject of great interest in the fashion and financial sectors. It provides many possibilities and considerable challenges to overcome within and beyond the apparel industry (Kaplan, 2020, 2021, 2023).

(a) AI: Novel, but Not New

The term "Artificial Intelligence" was first coined in 1956, nearly seven decades ago, during an eight-week-long workshop at the Dartmouth Summer Research Project on Artificial Intelligence. This event marked the inception of AI, or what came to be known as the "AI Spring" (Haenlein & Kaplan, 2019).

What followed was an iteration of AI "summers" and "winters," i.e., ups and downs in AI's evolutionary trajectory. The first summer lasted nearly twenty years and attracted significant research funding, accompanied by general hype and enthusiasm. In 1970, Marvin Minsky, an American computer scientist considered one of the founders of the field of artificial intelligence, is said to have claimed that within less than eight years, a machine could attain average human intelligence (Darrach, 1970). As we know, this did not happen, and enthusiasm for AI dropped. Due to computational power not growing at the expected speed, AI-driven applications were not in line with expectations. In consequence, funding became rare; AI winter had arrived. In the 1980s, a second AI summer began, triggered by investments from the Japanese government and, for the second time, the US DARPA (Defense Advanced Research Projects Agency), followed by another cold winter shortly after that, since expectations once again were not met.

The main reason for AI's not taking off until recently was insufficient computational strength. As illustrated by Moravec (1976), computers would have to be millions of times stronger to display actual AI, necessitating computational power at the same degree as an aircraft that requires horsepower. New approaches, such as deep learning and artificial neural networks, have become a reality, enabling substantial advances in processing power. A new era began in 2015, when AlphaGo, an AI-driven computer program created at Google, beat the world champion in the highly complex board game Go, which had been widely considered impossible until then. This novel period embodies the harvest of past AI advances and progress (AI's fall period) and completes the four seasons of artificial intelligence (Kaplan, 2022).

(b) AI: Plenty of Possibilities

AI can be classified into analytical, human-inspired, and humanized versions depending upon the types of intelligence it exhibits: cognitive (e.g., competencies related to pattern recognition and systematic thinking); emotional intelligence (e.g., adaptability, self-confidence, emotional self-awareness, achievement orientation); and social intelligence (e.g., empathy, teamwork, inspirational leadership) (Kaplan & Haenlein, 2019). Most AI systems that fashion retailers use today fall into analytical AI. For example, Amazon, one of the world's biggest apparel sellers, uses analytical AI in its Amazon Go Store, recommendation engine, and Alexa. A shopper could potentially purchase clothes at the Go Store, get customized fashion tips from Alexa, and finally receive a personalized suggestion for further items from Amazon's recommendation engine, all based on their past purchases, browsing history, and behavior (Kaplan, 2020).

Human-inspired AI can understand human emotions and cognition. Advanced vision software can recognize emotions such as anger, happiness, or surprise at the same level of precision (or better) as humans. Walmart, known for its affordable apparel, is currently developing facial recognition software to identify in-store customers' emotions, for which it has already filed a patent (Nassauer, 2017). Via video cameras monitoring customer movements and facial expressions, this system could detect unhappy or dissatisfied shoppers, applying a scale to the degree of their dissatisfaction. If the system detects, e.g., several unhappy customers in the checkout area, more employees would automatically be sent to open new checkout lanes. Similarly, a sales associate could be dispatched for assistance if the system spots a frustrated client in the clothing department (Nassauer, 2017).

Finally, humanized AI systems would ideally be self-aware and mindful in their interactions with human beings (or other AI systems). For example, a humanized AI robot helping a couple looking at various home furnishings would realize whenever they want private time to discuss a particular item, or on the contrary, when the AI-powered robot's presence is a welcome counterbalance to tension between the couple. Such systems are close to replacing humans. One might even ask oneself if, in the future, humanized AI robots might become clients of the fashion industry themselves (Kaplan, 2022).

(c) AI: Chock-a-Block with Challenges

Contemplating the various types of AI, one can imagine its ethical challenges and dilemmas, as detailed in Kaplan and Haenlein's (2020) six-dimension framework. When considering a potential customer asking Alexa to order apparel, the question is whether Alexa works for Amazon, the end user, or a third-party retailer paying to be Alexa's "favorite" brand. Might Walmart abuse its facial recognition software to detect vulnerable customers, selling them costlier products, with sales associates briefed to be extra welcoming and friendly to those shoppers? Moreover, humanized AI robots might lead to even more significant societal challenges with the potential unemployment of some or all human sales associates.

Moreover, several AI-related technological issues remain unresolved. For instance, developers can only create accurate and efficient AI systems if they have appropriate data to train AI. If the data collected is itself influenced by gender or race bias, this will necessarily be reproduced by the AI-driven system. Pertinent to that, a study by Raji and Buolamwini (2019) shows that Amazon's facial analysis technology revealed strong gender and racial bias. While the system incorrectly identified darker-complected women as men in more than 30 percent of the cases, it did so only in less than 10 percent of the cases for lighter-complected women. For men, no incorrect identification at all was detected. Most likely, the system was trained with data mainly from Caucasian men.

Politicians and regulators must intervene to ensure ethical behavior in an AI-driven fashion sector and beyond (Haenlein et al., 2022). Liability is one area that needs consideration, along with data protection and privacy laws. Imagine an AI-driven fashion recommendation system selling data about body size and weight to an online supermarket, using this data to target overweight persons with ads for various junk foods. Imagine the same system, enabled via facial expression software, to detect a person's general state of mind, trying to convince an apparently unhappy person to engage in some comfort shopping. These examples make it clear that regulation is needed.

Conversely, overly strict regulation might become a barrier to innovation, as investors and inventors might feel compelled to move their activities to less regulated parts of the world (Libai et al., 2020). For example, China is reluctant to constrain AI progress, and Chinese citizens are seemingly more open and flexible regarding privacy and data protection (Kopalle et al., 2022). This may be part of why Chinese-based fashion retailer Shein currently has a significant advantage over its competitors, Forever 21, Primark, and Boohoo, based in the US, Ireland, and the UK, respectively.

2.3 Ensemble: Apparel, Sustainability, and AI

Despite these problematic areas, environmentally speaking, AI creates many possibilities for fast fashion to become more sustainable, such as encouraging more eco-responsible consumption patterns and improving supply chain and inventory management. However, to fully tap into AI's potential, fast fashion needs to rethink its business model altogether. Whether fashion brands will go in this direction or stick to their old habits represents an ethical juncture for the industry.

(a) Enhanced Eco-Responsibility via AI-Driven, Enriched Customer Experience

The first application of artificial intelligence in fashion is to use it as a recommendation system. AI-driven assistants suggest items to customers based on their body measurements, purchasing history, and personal style. Via smart mirrors, customers can try on apparel virtually without needing dressing rooms. The mirror gives customers the best possible information by showing side-by-side comparisons of several fashion items and alternating colors to detect which shade suits a client's skin tone better. Such systems augment customer satisfaction with chosen items and result in fewer returns. This is of particular interest in the online sector, as up to 40 percent of online purchases are ultimately returned, leading to a high carbon footprint, as often, these items are not resold but rather destroyed (Byers, 2020). This consumption pattern is especially observable in China, where shipping is highly efficient and low-cost, and returning purchased products has become the preferred consumer choice (Chen, 2021).

AI even enables complete customization (Kaplan, 2006) and tailor-made clothing. With the application of three-dimensional modeling technologies, virtual fitting rooms precisely measure a client's body and adjust the digital models accordingly. Such personalization is possible by simply using one's smartphone camera during online shopping. For instance, Japanese fashion designer Yuima Nakazato uses this technology for his fashion creations. First, Nakazato takes the client's measurements using a 3D scanning device before sending the data to an AI-driven machine that directly cuts the fabric to assemble the complete garment. AI additionally limits material waste to a minimum by calculating the perfect cut, while ideal fit reduces the return of items (Mollard, 2020).

Furthermore, custom-tailored items increase a product's emotional value, leading to customers keeping such things longer and wearing them more often (Li et al., 2012), and a memorable purchasing and consumption experience augments this emotional attachment. This might also lead consumers to customize and buy the same sweater as their friends, diversifying the possibilities for clothing's affective potential. Such a sweater will most likely have high emotional value to the group of friends.

However, it is unclear how such an approach aligns with fast fashion's business model, incentivizing overconsumption and short product life cycles. This topic is discussed below.

(b) Strengthened Sustainability via Smart Supply Chains and Stock Supervision

Artificial intelligence and advanced data analytics also lead to optimized supply chain management, significantly improving the forecasting of sales numbers, fashion trends, and customer behavior. Much scientific research has been dedicated to it, leading to cost savings and simultaneously augmenting sustainability (Bendig et al., 2023; Benzidia et al., 2021; Naz et al., 2022). According to a Capgemini survey, AI could save retailers more than €300 billion annually by enabling efficiency in several supply chain processes and operations (Jacobs et al., 2018). In particular, the apparel industry has battled inventory distortion for a long time, with frequent overstocks leading to overproduction, high energy consumption, and enormous waste, i.e., high carbon footprints. Artificial intelligence can help overcome these practices.

Supply chain efficiency is another consequence of the higher usage of robotics and automation. According to NextMSC (2021), robotics automation's value in retailing was estimated at approximately €10 billion in 2020 and is expected to attain just under €30 billion in 2026. Robots will likely replace the human workforce in various sectors, such as production and transportation, marketing and sales, and warehouse sorting and packing. Automation leads to cost reductions and allows for the reshoring of production, potentially overcoming much of the fashion industry's highly criticized labor conditions; it also, however, increases layoffs and unemployment. Nonetheless, optimized supply chains and higher manufacturing flexibility reduce environmental impacts. From an ethical point of view, the fashion industry, like many other sectors, will have to balance the trade-off between higher unemployment due to increased automation and lower carbon emissions due to more sustainable supply chains.

AI even can help in choosing more sustainable materials and textiles. By analyzing a specific design, artificial intelligence can identify eco-friendly materials, enabling the earth-friendly manufacturing of the desired fashion item. Moreover, AI calculates how a specific fabric falls and a model creases. Lastly, AI can also find more sustainable alternatives to synthetic textiles without changing a designer's original vision.

(c) Greener (fast) Fashion via Bold Modification of Business Models

More broadly, the fashion industry's entire business model would require a transformation to introduce AI across the sector. Previously, this chapter discussed AI's potential to increase a product's emotional value, leading to more prolonged and frequent use of said fashion item. However, this clashes with fast fashion's current model, which is propelled by limited or single-use products. To alleviate fast fashion's environmental effects, AI could increase transparency regarding where materials come from, where products are manufactured, and under what labor conditions. Consumers could then opt for more sustainable products – which would most likely be more costly – and, as such, not aligned with the industry's current pattern of low-cost pricing strategies. In this respect, AI could even calculate optimal prices by analyzing the cost of materials, energy consumption, or competitors' pricing. Above all, one could include the cost of carbon emissions into the equation, i.e., the price of sustainability.

Another potential way of reducing the fashion industry's carbon footprint would be to prioritize second-hand markets, consistent with a circular economy approach, and de-incentivizing fast fashion's short life cycles and overconsumption. By the end of 2023, the second-hand apparel market is forecasted to more than double from €22 billion in 2018 to €47 billion (Sorokanish, 2020). Artificially intelligent apps and websites could be a powerful means of facilitating the offering and purchase of second-hand apparel, connecting sellers and buyers, suggesting trustworthy prices, and recommending products. The same approach could be used for apps and websites specializing in renting clothes for specific events such as weddings or galas, thus preventing the need to buy them.

Finally, artificial intelligence can find and simplify the recycling of fashion items. The United Nations Environment Programme (2019) found that the apparel industry loses approximately €450 billion annually due to its lack of clothing disposal practices and recycling efforts. Algorithms could increase transparency regarding which fabrics are more recyclable than others and suggest the ideal processes.

Implementing such changes ultimately depends on the readiness of high-level decision-makers and a corporation's main shareholders. CEOs may be hesitant to give up on profit margins, at least those who aren't ardent advocates for sustainability themselves. However, consumers' growing interest in companies' CSR and sustainability initiatives may be a strong argument for gradually implementing such changes.

2.4 Samples: Cases and Companies

Applying the preceding sections' observations, I present three case studies of fashion brands using AI toward sustainability. The section first analyzes the Spanish company Zara, the inventor of fast fashion; second, the Swedish company H&M; and third, Farfetch, the British-Portuguese luxury fashion retail platform.

(a) Zara: Fast Fashion's Forerunner

Pioneer of the fast-fashion movement, Zara transformed an entire industry through its then-innovative business model. Because of just-in-time manufacturing and highly integrated supply chains, Zara can design a new clothing product and start selling it in outlets in no more than three weeks (Tokatli, 2008). These fast turnarounds enable Zara to adapt quickly to rapidly changing consumer preferences, leading to its success story. Despite this much-acclaimed way of doing business in fashion, Zara's model is increasingly under attack, with sustainability concerns gaining momentum. Sustainability has become particularly important to Zara, as pledged by the Spanish retailer itself (Shatzman, 2019). It being considered the inventor of fast fashion, consumers increasingly criticize Zara for producing approximately 450 million clothing items per year and releasing around 500 new products per week, leading to more than 20,000 different designs annually (Allon, 2022).

While its brick-and-mortar strategy can be considered a success story compared to its competitors, Zara lacks in its online presence. To gain online expertise, Zara is pursuing a partnering strategy with several start-ups. In 2018, Zara entered a partnership with Jetlore (an AI-driven purchasing behavior forecasting platform), with Intel (to reduce inventory by instantly identifying item volumes in stock), and with Tyco (the company's alarm provider, to incorporate microchips into security tags), allowing Zara to precisely locate products in the supply chain (Dowsett, 2018). Most importantly, it is teaming up with Heuritech and its market intelligence tool, enabling Zara to analyze up to three million images posted on social media daily to decrypt the latest consumer trends and demands. Finally, a partnership with Fetch Robotics enables Zara to use robotics in its Click and Collect division, i.e., customers order online and collect their purchases at one of Zara's brick-and-mortar stores. AI-powered robots search for the ordered item and place it in dedicated dropboxes for customers to retrieve (Dowsett, 2018).

Moreover, Zara applies a virtual styling tool embedded into its in-store mirrors: Shoppers scan item barcodes via Radio-Frequency Identification (RFID) technology, and the system instantly provides several choices for combining the respective item with other apparel and accessories. Via geolocalization technology, customers will be guided in-store via smartphone to find recommended items quickly. Furthermore, the AI-driven system is also employed in Zara's online store, where shoppers browse through the various clothing options, guided by their previous purchases and online behavior, suggesting products available at a close-by store for pickup in less than 30 minutes. It is anticipated that by offering more assistance to customers when selecting their clothing, they will be more likely to choose apparel they genuinely appreciate and, therefore, less likely to return items once they have reached their homes. As previously mentioned, lower return rates would reduce waste and thus increase sustainability.

(b) H&M: The Apparel Industry's AI Advancements

Having begun using AI in 2016, H&M belongs to the pioneers of applying AI in the fast-fashion industry. In 2018, the group established a dedicated AI department, which is considered a significant step toward reaching its objective of a climate-positive value chain by 2040. Artificial intelligence is applied at H&M to predict market demand, optimize supply chains, and avoid excessive inventory on a highly granular level. More than 250 data scientists write algorithms that decrypt data from loyalty cards, store receipts, online purchases, and product returns to forecast trends and future purchases (Chaudhuri, 2018). In one of H&M's Stockholm-based stores, AI identified that customers were mainly female, preferred higher-priced clothes, were fond of a particular coffee shop, and often bought items featuring floral patterns. Adaptations were made accordingly; premium-priced leather bags and cashmere sweaters were offered in addition to the classic, lower-priced items. These changes led to significantly higher store sales but, more importantly, to a reduced carbon footprint.

Another AI-driven initiative, Body Scan Jeans, increases customer experience and satisfaction, enabling H&M to offer tailor-fitted, on-demand manufactured jeans. Clients scan their bodies in dedicated in-store cabins, which are instantly transformed into a digital avatar. Instead of trying on items, customers can virtually try various styles and colors. The system provides customers with recommendations adapted to their body shapes. Via AI, the final choice of jeans can be made, perfectly fitting the client's measurements, within just a few weeks. Personalized jeans result in lower return numbers and, in turn, lower CO_2 emissions.

More recently, H&M began assessing each new AI project regarding potential ethical and societal risks via a nine-principle index, i.e., artificial intelligence applied therein must be focused, beneficial, fair, transparent, governed, collaborative, reliable, respect human agency, and secure (Minevich, 2022). In the past, these principles were used to assess, e.g., whether AI would result in discrimination against particular customer groups or inordinately alter consumer behavior. Establishing the Ethical AI Debate Club aims at the same objective via discussing potentially nefarious AI. The club's goal is to raise awareness of potential moral and societal issues and challenges brought on by the use of AI. Its ultimate objective is establishing an ethical mindset and a culture of responsible artificial intelligence across the H&M group. One of the club's debates addressed the topic of in-store cameras, via which the retailer can detect customers' emotional states by analyzing their facial expressions. The question was whether such data should be used to adjust in-store advertising, lighting, and music to augment store sales, even if customers are unconscious of such tracking of their emotions. As a result of discussions among the club's members, H&M decided against the set-up of facial expression technology in their stores.

(c) Farfetch: Couture Fashion's Not-Farfetched Conscience

High-end fashion, so-called couture fashion, in contrast to fast fashion, is more sustainable since, in theory, better quality clothes results in higher longevity. Couture consumers should also be more reluctant to dispose of their apparel due to their emotional value and attachment to it. Therein, AI can help go even a step further toward sustainability, notably by increasing customers' knowledge about reducing their carbon footprint, potentially leading to more conscientious purchasing behavior. To tap into this possibility, the luxury fashion retail platform Farfetch, which sells products from several hundred couture brands worldwide, has implemented its "Farfetch fashion footprint tool," guiding consumers toward more sustainable shopping.

This AI-driven tool calculates and presents scientific data about a given material's and fabric's long-term environmental footprint, detailing its carbon emissions and water usage. Subsequently, it provides consumers with more sustainable apparel solutions. Recycled or responsibly sourced materials are showcased as preferable over less sustainable options such as cotton, polyester, nylon, wool, silk, or leather.

Farfetch's fashion footprint tool also suggests that consumers buy second-hand, listing the environmental advantages of such purchases. Again, in a very detailed manner, the tool precisely presents how much carbon emissions and wastewater can be avoided by buying an item previously belonging to another fashion aficionado. By teaming up with Thrift+, platform users can send their apparel items – free of charge – to the partner company, which then resells the garment. Additionally, the platform provides contacts to repair services, enabling fashionistas to hold on to their beloved handbags, shoes, and apparel for longer. Second-hand is presented as the more ethical and sustainable choice, aiming for more conscientious and enlightened customer behavior, empowering customers to “think, act, and choose positively,” as stated on Farfetch’s website (Farfetch, 2023).

2.5 Trends: Fashion’s Future

Above, I demonstrate how AI is applied to render fast fashion less polluting. Yet, for fast fashion to become genuinely green, its entire business model would most likely need to change, i.e., fast fashion becoming eco-responsible, or slow fashion. This section’s concluding remarks discuss a further way for fashion to become more environmentally sustainable: replacing physical apparel with digital garments altogether.

Increasingly, people buy new clothes specifically for social media appearances, i.e., Instagram influencers who monetize their online presence by modelling or advertising a vast and varied wardrobe. Yet, being fashionable on social media does not necessarily require the purchase of actual clothes: Companies such as Dress-X sell fully digital fashion items in which they will dress you for your preferred image. The process is foolproof: upload a picture of yourself, purchase the desired clothing, and the service provider will send you back an image with you (virtually) wearing the new cap, sweater, or bag. You never have to buy the actual item; it is all digital. While virtual apparel is undoubtedly the more sustainable option compared to physical clothes, it is not necessarily less expensive: prices range from around 30–50 euros for a stylish dress, sweater or pair of shoes to several thousand euros for a (virtual) haute couture suit.

To dive even deeper into the virtual world, think of the Metaverse, i.e., three-dimensional virtual worlds that people enter in the form of avatars via virtual and augmented reality headsets. Avatars, too, will want to be fashionable and need new (digital) apparel. While Mark Zuckerberg speaks about his Metaverse in the future tense, predecessors such as Second Life, created as soon as 2003 and likely the most known of the first virtual social worlds, have existed for nearly twenty years. Research shows that users experience these virtual social worlds as an extension of their real lives (Kaplan & Haenlein, 2009a, b). As early as a decade ago, they designed, sold, and bought digital clothing (Haenlein & Kaplan, 2009). If predictions become a reality, we will spend more time in the Metaverse working and socializing, all in the form of highly authentic avatars mimicking our facial expressions and body movements – and wearing digital fashion. This environment will be an (almost) entire new fashion market to develop.

However, avatars and virtual worlds are energy-consuming and, as such, polluters. The world's data centers are responsible for approximately two percent of global emissions, as much as generated by the entire aviation industry (Vaughan, 2015). Intel estimates that the future Metaverse would require around a thousand times more power than the world's current combined computational capacity (Gartenberg, 2021). Furthermore, the fashion industry's future depends upon consumers' acceptance of new business models, i.e., slower, more costly fashion. According to a Zalando study (Foroozesh, 2021), while 60 percent of consumers desire more transparency, including information on what materials are used, a mere 20 percent actively search for such information. However, it remains unclear how many of these 20 percent change their purchase behavior due to such information, preferring a more sustainable, potentially more expensive option. These numbers somewhat oppose this chapter's introduction stating an increasing number of eco-responsible consumers. Still, when all's said and done, only markets (and time) will tell fast fashion's future fate.

References

Allon, G. (2022). *The fashion industry's dirtiest secret*. Insider, December 23.

Bendig, D., Schulz, C., Theis, L., & Raff, S. (2023). Digital orientation and environmental performance in times of technological change. *Technological Forecasting and Social Change, 188*, 122272.

Benzidia, S., Makaoui, N., & Bentahar, O. (2021). The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance. *Technological Forecasting and Social Change, 165*, 120557.

Byers, G. (2020). Artificial intelligence is restyling the fashion industry. *Towards Data Science*, February 28.

Chaudhuri, S. (2018). H&M pivots to big data to spot next big fast-fashion trends. *Wall Street Journal*, May 7.

Chen, A. (2021). 5 ways AI is transforming the fashion industry for sustainability. *Towards Data Science*, December 29.

Darrach, B. (1970). Meet shaky: The first electronic person. *Life Magazine*, November 20.

Dowsett, S. (2018). Zara looks to technology to keep up with faster fashion. *Reuters*, June 15.

Ellen MacArthur Foundation. (2017). *A new textiles economy*. EMF.

Farfetch. (2023). *Positively farfetch*. https://www.farfetch.com/pt/positively-farfetch/.

Foroozesh, S. (2021). Sustainability in the fashion industry: Avoiding communication landmines. *Forbes*, July 13.

Gartenberg, C. (2021). Intel thinks the Metaverse will need a thousand-fold increase in computing capability. *The Verge*, December 15.

Haenlein, M., & Kaplan, A. (2009). Flagship brand stores within virtual worlds: The impact of virtual store exposure on real-life brand attitudes and purchase intent. *Recherche et Applications en Marketing, 24*(3), 57–80.

Haenlein, M., & Kaplan, A. (2019). A brief history of AI: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14.

Haenlein, M., Huang, M.-H., & Kaplan, A. (2022). Business ethics in the era of artificial intelligence. *Journal of Business Ethics, 178*, 867–869.

Jacobs, K., Himmelreich, A., KVJ, S., Khemka, Y., Tolido, R., Buvat, J., & B., A. (2018). *Building the retail superstar: How unleashing AI across functions offers a multi-billion dollar opportunity*. Capgemini Research Institute.

Kaplan, A. (2006). *Factors influencing the adoption of mass customization: Determinants, moderating variables and cross-national generalizability*. Cuvillier.

Kaplan, A. (2020). Retailing and the ethical challenges and dilemmas behind artificial intelligence. In P. Eleonora (Ed.), *Retail futures: The good, the bad and the ugly of the digital transformation* (pp. 181–191). Emerald Publishing.

Kaplan, A. (2021). Higher education at the crossroads of disruption: The university of the 21st century. In *Great debates in higher education*. Emerald Publishing.

Kaplan, A. (2022). *Artificial intelligence, business, and civilization: Our fate made in machines*. Routledge.

Kaplan, A. (2023). *Business schools post-covid-19: A blueprint for survival*. Taylor & Francis.

Kaplan, A., & Haenlein, M. (2009a). Consumer use and business potential of virtual worlds: The case of second life. *International Journal on Media Management, 11*(3/4), 93–101.

Kaplan, A., & Haenlein, M. (2009b). The fairyland of second life: About virtual social worlds and how to use them. *Business Horizons, 52*(6), 563–572.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri in my hand, who is the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25.

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons, 63*(1), 37–50.

Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachadran, D., Reinartz, W., & Rindfleisch, A. (2022). Examining artificial intelligence (AI) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing, 39*(2), 522–540.

Li, G., Li, G., & Kambele, Z. (2012). Luxury fashion brand consumers in China: Perceived value, fashion lifestyle, and willingness to pay. *Journal of Business Research, 65*, 1516–1522.

Libai, B., Bart, Y., Gensler, S., Hofacker, C., Kaplan, A., Köttenheinrich, K., & Kroll, E. (2020). A brave new world? On AI and the management of customer relationships. *Journal of Interactive Marketing, 51*(C), 44–56.

Minevich, M. (2022). AI visionary and responsible AI leader Linda Leopold of H&M Group. *Forbes*, May 2.

Mollard, M. (2020). Bridging the gap between artificial intelligence and creativity in fashion. *Heuritech*, June 2.

Moravec, H. (1976). *The role of raw power in intelligence*. Available at: https://web.archive.org/web/20160303232511/http://www.frc.ri.cmu.edu/users/hpm/project.archive/general.articles/1975/Raw.Power.html. Accessed 22 Mar 2023.

Nassauer, S. (2017). Robots are replacing workers where you shop: Walmart and other large retailers, under pressure from Amazon, turn to technology to do workers' rote tasks, *Wall Street Journal*, July 19.

Naz, F., Agrawal, R., Kumar, A., Gunasekaran, A., Majumdar, A., & Luthra, S. (2022). Reviewing the applications of artificial intelligence in sustainable supply chains: Exploring research propositions for future directions. *Business Strategy and the Environment, 31*, 2400–2423.

NextMSC. (2021). *Global opportunity analysis and industry Forcast, 2021–2030*. Next Move Strategy Consulting. December.

Raji, I. D., & Buolamwini, J. (2019). *Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products*. Conference on Artificial Intelligence, Ethics, and Society, Honolulu., January 27–28.

Shatzman, C. (2019). Zara reveals ambitious new sustainability goals. *Forbes*, July 18.

Sorokanish, L. (2020). *By 2023, the second-hand clothes market will double to $51 billion* (p. 237). Here's why. Fast Company Magazine.

Tokatli, N. (2008). Global sourcing: Insights from the global clothing industry—The case of Zara, a fast fashion retailer. *Journal of Economic Geography, 8*(1), 21–38.

United Nations Environment Programme. (2019). *UN Alliance for sustainable fashion addresses damage of fast fashion*. UNEP.

Vaughan, A. (2015). How viral cat videos are warming the planet. *The Guardian*, September 25.

3

Artificial Intelligence and the Global Automotive Industry

Tarek H. Selim and Mostafa Gad-El-Rab

3.1 Introduction

Autonomous vehicles (AV) are rapidly changing how people drive and transport goods. Self-driving technology is at a stage of artificial intelligence (AI) development that is poised for significant disruptions to the business and social spheres of the automotive industry and beyond. This chapter overviews AV technology, as well as its link to AI and emerging machine learning trends, and their relation to the social, economic, and infrastructure domains, and subsequently determines the projected sustainability of AVs using critical industry factors. We address AI technology and its application to the AV industry by analyzing technological, social, business, customer, and input supply factors leading to a most likely scenario for the AV industry's sustainable future.

T. H. Selim (✉)
Global Expert Council, World Economic Forum, Cologny, Switzerland

M. Gad-El-Rab
Capabilities and Insights Division, McKinsey & Company, Cairo, Egypt

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_3

The world is witnessing a new technological revolution, causing a radical change in how humans live, interact, and work. Also known as Industry 4.0, this fourth industrial revolution is shaping a new accelerated reality that is altering global society through innovations in many diverse areas, including AVs. It is integrating the digital, physical, and biological worlds to create new cyberspace dimensions at a rapid pace. Virtual reality, machine learning, autonomous driving, and smart robotics are just a few examples of this trend. Schwab (2016) posits that there are three distinct features of Industry 4.0: (1) scope, (2) velocity, and (3) impact, which are expected to yield profound behavioral changes, disruptive innovations, and accelerated digital interconnections in the future of work, production, and society. AI is a core component of Industry 4.0. Its application to autonomous vehicles will soon profoundly affect social behavior, the rate of innovation, and the intensity of human-digital interconnections (Fig. 3.1).

However, there are valid concerns about the perils of the fourth industrial revolution with respect to security and privacy, the job market, and

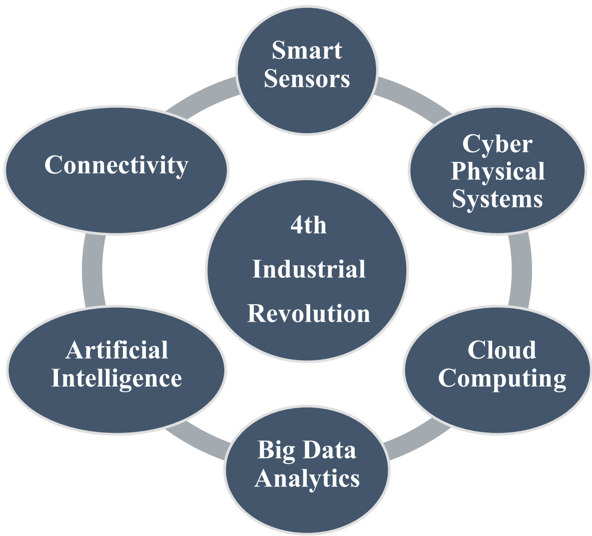

Fig. 3.1 Basic elements of the 4th Industrial Revolution. (Source: Author Illustration based on World Economic Forum, 2023 and Schwab, 2016)

rising inequality due to the digital divide. For example, since Industry 4.0 is heavily reliant on artificial intelligence and machine learning, it is feared that machines will eventually replace human workers and increase unemployment, with some studies showing that automation may threaten up to 47 percent of current jobs (Anderson et al., 2016). One important concern is whether society, businesses, and governments will commonly adapt to the new industrial revolution. The speed at which technology evolves, in turn, increases the speed of the labor substitutability process, which calls for a heightened speed in government regulations, social acceptance, and job re-skilling to keep up (Fig. 3.2).

With such dynamic advancements, the world is expected to witness a stark change in how the global automobile industry operates. The integration of AI and the development of AVs bring with them many opportunities and challenges.

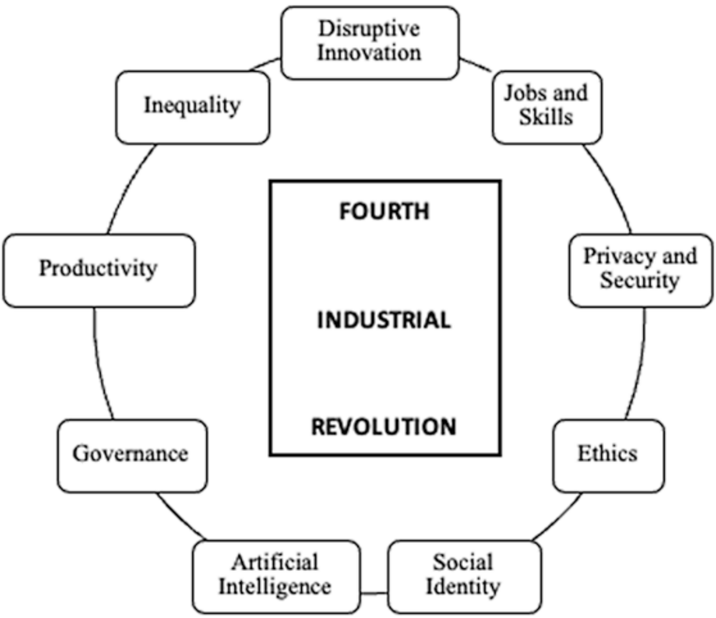

Fig. 3.2 Impacts of artificial intelligence and the Fourth Industrial Revolution. (Source: Author Illustration based on World Economic Forum, 2023)

Regarding opportunities, the fourth industrial revolution has already allowed the automobile industry to achieve numerous improvements, a trend that is expected to continue. Currently, some manufacturers have already started implementing remarkable developments that have revolutionized the nature of car driving. However, Industry 4.0 will not only reshape the automobile industry from the producer's side but for the consumer as well. It will allow car manufacturers to increase customer satisfaction by adding personalized features that meet constantly evolving customer needs (Heineke & Kampshoff, 2019; Manyika & Russell, 2020). By doing this, customers will feel like they have a special connection to their cars.

Additionally, using AI in production will maximize efficiency and optimize production levels such that most decisions will be data-driven. Smart algorithms in AVs will help reach destinations faster and drive more efficiently. Also, AVs can detect unsafe situations and avoid them. Additionally, AVs tend to minimize the risk of human error during driving, whether due to driver distraction or a medical emergency, as they always operate with maximum ability. The presence of self-driving cars is also expected to reduce traffic jams and the frequency of collisions (Fu-Lee & Qiufan, 2021; Heineke et al., 2017).

However, the application of AI comes with a list of challenges that put the AV industry and society at systemic digital risk. There are concerns about data security, safety, and consumer privacy. Being dependent on technology and AI, every manufacturer and user in the automobile industry will be constantly vulnerable to cybersecurity attacks and privacy threats. Data Management is another challenge that automobile manufacturers face, with heavy reliance on the production and processing of massive volumes of data. Big data analytics is now required to streamline data to achieve momentous and concrete decisions (Billington, 2019). Other critical challenges include the faster pace of AV technological breakthroughs than their social acceptance, behavioral resistance to change in adopting new AI standards, and the threat of rising inequality due to an anticipated digital divide between transport areas (Baltic et al., 2019; Burns & Shulgan, 2018).

Due to rapid developments in AI and machine learning that continue to accelerate through the fourth industrial revolution, their direct application to AVs has witnessed immense advancement, along with disruptive innovations, over the past few years. Industry 4.0 enabled the existence of both the hardware and the software components required for developing smart driving vehicles. AVs can now operate almost independently through smart sensors, cloud communication, big data analytics, and physical cyber systems. According to the McKinsey Center for Future Mobility, the adoption of AVs in the United States, for example, is expected to generate public benefits of more than 800 billion dollars a year by 2030, and more than half of the benefits will emerge from safer roads and fewer fatal accidents (Deichmann et al., 2023). The adoption of AVs will require a shift in how the market is governed and regulated. It will require changes in consumer driving behavior, and new insurance and safety mechanisms, with an anticipated fierce degree of competition in business development and delivery. In addition, road infrastructure will require digital upgrading to avoid any perils and ensure the smoothness of the AV operations, which is expected to take decades of time to complete, hence creating digital inequality in the process (Magias, 2023).

3.2 The Six Stages of Self-Driving and Their Implications

The Society of Automotive Engineers (SAE) and the National Highway Traffic Safety Administration (NHTSA) define six levels of self-driving automation in-car technology in the United States. These levels reflect how much automated control the car has without human intervention in the process of driving, commuting, parking, and communicating. The levels of self-driving automation are (see Table 3.1 for an overview of levels):

(1) **Level 0 (No Automation)**: This level refers to a state where there is no automation, meaning that the driver must have complete control over the car and be alert at all times. Moreover, the driver is solely responsible for monitoring the road, reacting to any traffic changes, and handling the safety of the people and objects within the car's surrounding area.

Table 3.1 Stages of self-driving automation

Levels	Technology	Social/behavioral	City infrastructure
Level 1	Up to 2 assistant features	Accepted	Basic
Level 2	3 or more assistant features	Accepted, with adaptation of attention to navigation system and auto-assisting features.	Basic
Level 3	Co-pilot with LiDAR, sensors	Acceptance ongoing.	High frequency road maintenance needed.
Level 4	AI synchronous mapping between road and vehicle	Social resistance to change as risk to adaptation, new safety standards and insurance regulations needed, and legality concerns	Dedicated short-range comm. (DSRC), equipment operating in the 5.9GHz frequency range, synchronous mapping between the roadways and vehicle.
Level 5	Advanced AI features, continuous and pre-synchronous, bidirectional mapping	High risk perception, trust issues, new behavioral acceptance is needed	Central broadcast system is required, continuous pre-synchronous bidirectional communication between road and the vehicle.

Source: Authors' research, Lynberg (2019)

(2) **Level 1 (Function-Specific Automation)**: At this level, few automation features exist, though there are driver-assistance capabilities. For example, the car may have cruise control as well as dynamic brakes. These simple features do not mitigate the need for human intervention as a driver is still required to have full control over the vehicle as well as constantly maintain a safe driving operation.

(3) **Level 2 (Combined-Function Automation)**: Vehicles are semi-autonomous at this level. Cars that belong to this level usually have at least three primary control functions that work coherently. Examples include auto-brake, lane fitting, and park assist features, which react to critical disruptions on the road. Nevertheless, the driver must still be constantly alert and have their hands on the wheel. The driver is still responsible for the safe operation of the vehicle.

(4) **Level 3 (Limited Self-Driving Automation)**: This level is known as conditional automation. The car serves as a co-pilot, but it still requires a driver behind the wheel to take over in some conditions with a comfortable transition time. The car can take control over most of the critical features in known environments only.
(5) **Level 4 (High Automation)**: Here**, c**ars can drive themselves almost entirely. They have control over all safety-critical features and do not need human intervention to function. However, a driver's cockpit still exists in the car for the driver to take over in rare emergency situations.
(6) **Level 5 (Full Automation)**: Fully autonomous vehicles that can operate anywhere and under any conditions. A human driver is not needed to operate the vehicle.

Currently, most of the cars available in the new vehicle market are at level 2 or level 3 of automation. They have cruise control, parking automation, and GPS mapping options. Additionally, most available cars have lane fitting assistance, auto-brake sensitivity, keyless entry, and computer hardware using a low-bandwidth memory device. Existing literature estimates that mass car technology will move towards the full end of level 3 and the early stages of level 4 within the next five years. AVs have the potential to reach the end of level 4 and possibly a partial stage of level 5 in ten years as standard technology in the market. This advancement is under the condition that the current pace of disruptive innovation in the global automotive industry will continue, accompanied by social acceptance, new safety standards, insurance mechanisms, and accelerated digital road infrastructure.

3.3 Artificial Intelligence and Technology Features

Generally, AVs are designed using a sense-plan-act method, which is fundamental for most robotic design systems. Firstly, in the sense phase, sensors, cameras, and radars are installed to gather raw data about the environment and the exact positions of the vehicle's surroundings with precise accuracy. Secondly, the planning phase allows this raw data to be analyzed and interpreted in real-time using software algorithms such that the vehicle uses the interpreted data to plan its own mode of operations. Finally, the act phase concludes the design, where the vehicle translates the raw data and plan into actual action, such as accelerating, steering the wheel, or braking, i.e., self-driving action (Philipp & Padhi, 2017).

The sense-plan-act design is a loop that keeps running all the time. Sometimes, more than one loop runs at the same time in parallel to one another. The loop frequency varies from operation to operation, depending on the action's nature (Anderson et al., 2016). For example, some loops run at a very high frequency to initiate emergency braking, while others, such as changing lanes, run at a low frequency. Each one of the stages of the sense-plan-act design has hardware and software requirements to reach the safest and most efficient AV driving mode (Voelzke & Rudolph, 2023). The sensors required in an AV fall into three categories: cameras, radars, and LiDARs.

Regarding camera technology, rear and 360 cameras are currently available in cars to support the driver by providing visual access to the vehicle's surroundings. The three-dimensional camera view is also available through four to six cameras installed around the car with a very high dynamic range, allowing clear pictures to be delivered even when the light is directly on the camera's lens. Then, a central processing unit collects raw data from the four to six cameras and constructs a full view. Advanced software is required for this process, which requires high data storage capacity as well as high power. Thus, a decentralized camera system was developed in which the central control unit is removed and only the smart cameras and the head-unit exist. This is the camera technology that is currently being used in AV design. The camera system is composed of medium and high-range cameras. These cameras use AI algorithms to detect objects near the car, identify them, and calculate the exact distance between them and the vehicle. The medium-range cameras are responsible for detecting lane signals, pedestrians, and emergency braking in cars ahead. The high-range cameras detect traffic signs and road traffic guidance (Voelzke & Rudolph, 2023), which are essential security features required to keep passengers safe. However, cameras still have serious limitations in terrible weather or extreme environments.

Radar and LiDAR systems are critical sensing technologies on which AVs rely heavily. Compared to cameras, they are generally more precise in any environment or weather. Radar is a radio detection and ranging system that uses radio signals' time of flight to determine the distance between itself and another object. Radar produces short-range signals for detecting blind spots, collision warnings, emergency braking, parking, and automatic distance control. However, radars have a significant limitation because they only work well for identifying metallic objects. Hence, they cannot detect pedestrians and other non-metallic things (Anderson et al., 2016; Voelzke & Rudolph, 2023).

LiDAR, a light detection and ranging system, uses a laser beam to detect and calculate distances between objects and the vehicle. LiDARs send a laser beam and receive its reflection through a highly sensitive laser-based sensor. Some advanced LiDARs have multiple laser finders and constantly rotating mirrors, allowing them to construct a three-dimensional view of the environment. Other valuable sensors in AVs include ultrasonic and infrared sensors. Ultrasonic sensors can detect lane lines and road markings in any environment (Jurgen 2013; Heineke et al., 2017; Weber, 2022), while infrared sensors can detect pedestrians and other objects, even in the dark (Fig. 3.3).

Another important requirement for AVs is vehicle-to-infrastructure communication (V2I). AVs must communicate with roads in real-time to detect the road lanes, traffic lights, and other navigation information

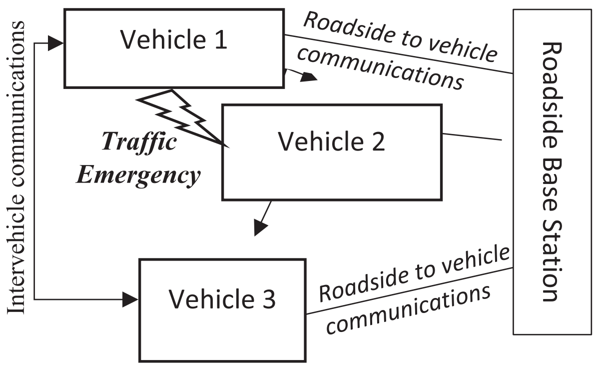

Fig. 3.3 Basic road infrastructure for autonomous vehicles. (Source: Author Illustration based on Jorgen, 2013)

needed. V2I communication can be unidirectional or bidirectional. That is, the AV can only receive information from a broadcast system on the road *or* communicate with a broadcast base station on a point-to-point basis. This communication provides a vital safety system (Ozguner et al., 2012). For example, traffic lights can transfer signals of their current state and cycle time information to AVs on the road, which can avoid collisions in intersections. Another application for V2I is providing the AV navigation system with map updates and traffic congestion through a wireless network.

Overall, for a car to be fully autonomous, the AI system must have extensive digital requirements to be able to constantly interact with the surrounding environment and take the right action at the right moment at the correct location. The current state of AI-featured car technology has not arrived at this status on a commercial level yet. Accordingly, the enormous amount of raw data provided by the different sensors of vehicles will need to be analyzed and interpreted efficiently and precisely without human intervention. This requires a high computing performance and large memory bandwidth to analyze the data and the AV car to act or react quickly. AVs are expected to require a memory bandwidth of more than 1 Terabyte per second to support the computing performance required for dynamic real-time road data (Crowe, 2019; Karkus et al., 2022). Some computer hardware companies, such as Nvidia, are producing computing platforms for autonomous vehicles to be able to process this amount of data in real-time.

Table 3.2 presents the main technical features and requirements for AVs. This includes radar, LiDAR, cameras, and infrastructure. Moreover, it presents these AI features' expected short-run and long-run development trends and requirements.

3.4 Autonomous Vehicles (AV) Industry Forces

A. Competition, Rivalry and Differentiation Points

The market structure for AVs is expected to be an oligopoly with high upfront costs and a relatively homogenous pricing range with the

Table 3.2 Features of autonomous vehicle technology

AI factors	Description	Current status	Future expectations: Short-medium run (1 to 3 years)	Future expectations: Long run (within the next decade)
Radar	A radio detection and ranging sensor that detects the position of objects using radio waves	Short and long-range radars are available in most commercial cars with a bandwidth of 1 GHZ and a frequency of 77 GHZ.	Radars sensors would use 79-GHz frequency band and 4 GHz bandwidth.	Radars will have a system-on-Chip (SoC) that integrates both the radar and a baseband that process the data.
LiDAR	A laser based light detection and ranging sensor, used to calculate distance to a static or a moving object and it can provide a 3D image.	Newly introduced for the automotive industry with a sophisticated mechanical mirror system with a 360° view. Currently, LiDARs are very costly.	There are 2 main systems under progress. A rotating laser using a micro-electro-mechanical system, and a solid-state laser.	There are concerns about the long-term reliability of the mechanical scanning mechanism, and new innovations in mechanical materials are needed.
Cameras	2D and 3D cameras in the front and the rear of the vehicle that provide a 360° external view.	Currently, 2D cameras are widely available to display images. 3D cameras are only available in some luxurious cars.	The majority of the cameras will have better resolution, object detection and classification and amplified reality.	CoaxPress (CXP) serial communication standards for transmission of video data at rates of 10Gbits/s and 12.5 Gbits/s per connection. This is 500 percent faster than the GMSL2 connection currently used.
Infrastructure	A specific road infrastructure needed to allow vehicle-to-infrastructure communication.	Road lane lines and traffic lights and frequent road maintenance are required.	A unidirectional communication between the AV and a central broadcast system that send the AV traffic lights signals.	A bidirectional communication that would allow for a point-to-point update for the traffic congestions, and other navigation information through a central broadcast.

advanced adoption of AI. Producers will be constantly incentivized to merge or collaborate, and businesses are expected to share their technologies to cut costs and develop a fast market share advantage over others.

Such a market scenario is currently in the making. For example, Fiat, Chrysler, and Peugeot have a collaboration project of 50 billion euros to develop an autonomous vehicle (Wayland, 2019). This collaboration saves 3.7 billion euros a year without closing any factories, yet thousands of engineering jobs could be lost. In general, the developed AV market is expected to consist of a small number of producers with huge capacities. However, the merging and collaboration incentive will be reduced in the long run when AI technology becomes readily available for all producers, and the cost of adoption significantly lowers.

In the autonomous vehicle industry, there is a business advantage for first movers. First movers in the market for AV will be able to take a high market share and generate much profit from the industry. This is coupled with an incentive for mergers and collaborations to achieve scale and market share advantage. However, this expected high return is associated with a lot of initial spending on research and development to reach a product that can be commercially successful in the market. Furthermore, this expected high return is based on speculations about AVs' expected demand, sales, and success. Also, second movers can have an advantage over first movers by undercutting the prices to gain a market share. Moreover, they will be able to imitate the technologies innovated by the first movers without spending a substantial amount of money on research and development.

In essence, the competition in the AV industry is expected to be governed by first and second movers while having an oligopoly market structure with a relatively homogenous price range based on high industry collaboration. However, over time, such competition is expected to be more intensive, with less frequent mergers, less collaboration, and more rivalry in technology adoption (Dhawan et al., 2019).

Due to this oligopoly-structured approach, firms also experience a differentiated advantage. The AV industry has different producers differentiation points from the non-AV automotive industry. This is related to the fact that all the producers depend on the same technological elements in the production process: radars, LiDARs, cameras, and processing units. Moreover, all the AVs will have the same functions. Hence, what differentiates one producer from another is not technology per se but rather the safety features, the level of technological advancement in relation to road infrastructure, and the cost of LiDAR.

The most concerning part and the main differentiating point about AVs is the safety of the rider and the people around the car. In the not-too-distant past, an AV vehicle operated by Uber struck a pedestrian in Arizona, leading to her death (Wakabayashi, 2018). Safety features are expected to be a main differentiating point among different producers. The second differentiating point is the level of technological advancement. The advanced AVs have more autonomy features; however, they are not operable on all roads. This is because they require highly connected and digitalized road infrastructure, which is unavailable everywhere. In addition, they are expected to be expensive, making them unaffordable. On the other hand, the less advanced AVs are expected to be more common because of their affordability and because they are more operable on roads with different infrastructure levels. In retrospect, moderately advanced AVs are expected to have an advantage over the other two types on the condition that they have the minimum features that make them adequately autonomous for customer satisfaction, in addition to being affordable and operable on most roads.

The third differentiating point is the cost of LiDAR, which is the most expensive part of the AI components of an AV. LiDARs use laser light to produce a three-dimensional map of the vehicle's surroundings. Google's Waymo parts setup costs around 150 thousand dollars, with the LiDAR mapping system alone costing around 75,000 dollars, thus capturing 50 percent of the AV's total digital AI setup (Crowe, 2019). Consequently, firms in the AV industry are working on reducing the cost of LiDAR mapping systems through economies of scale or breakthrough innovations towards a new LiDAR system, making the AVs more affordable for the mass public (Spencer, 2020) (Fig. 3.4).

The relatively homogenous pricing range of AI adoption in AVs can be explained using kink pricing. A kink pricing for new AV demand is expected due to the disruptive innovation feature of new car technology. The price is expected to be stable, yet within bounds of a "kink" as follows: when a new technology appears, the willingness-to-pay feature in

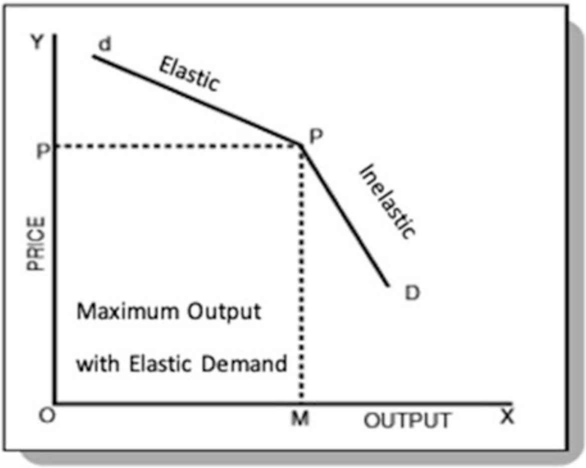

Fig. 3.4 Price kink for new autonomous vehicle industry demand due to open-source technology. (Notes: The top flat demand is elastic until a price kink, followed by inelastic demand at the bottom. The homogenous pricing range is expected around P, and maximum output is expected at M until full saturation of elastic demand is realized)

demand will drive prices above the initial price level, yet due to open-source technology, such a price hike will only be short-lived. When competitors break through this technological innovation, the new tech standard will drive prices down due to more competition beyond the first-mover advantage. Business profits of the first movers will still have the edge over others, yet such a scenario is not expected to last for a long time due to imitation and open-source technology.

Additionally, due to the oligopoly nature of the global auto market, in which Research and Development (R&D) expenses act as a "barrier to entry" for small players, the tendency to reduce prices to capture higher market share is limited. Lower prices will only start a price war, and competitors will follow suit by lowering prices even more to outperform the initial price war. Since technology is expected to be open source, starting a price war will lower business revenues due to the inelastic demand behavior attributed to necessitating this new technology.

B. Substitutability

Substituting non-autonomous vehicles with AVs is expected to decrease the number of standard street cars. This is because AVs are expected to boost the overall size of the ride-sharing industry. Existing research shows that most of the ridesharing fare goes to the driver as the company gets a small cut. With AV cars depending less on human drivers and more on AI driving, the fare of ridesharing is expected to decrease, whereby people will depend more on automated ride-sharing services, which will decrease the total number of privately owned vehicles on the road. An example of this sequence of events can be illustrated in the following numerical example that shows the decrease in the total person-hours spent on driving:

Case A (30 people, no ridesharing):

[30 people] × 40 hours = 1200 person-hours.

Case B (30 people, 10 of them with ridesharing):

[20 people] × 40 hours + [10 people] × 20 hours (Sharing)
= 1000 person-hours < 1200 person-hours.

Case C (30 people, 10 of them with ridesharing, with externalities):

[20 people] × 40 hours + [10 people] × 20 hours (Sharing)
+ 7 hours (Private) + 8 hours (Externalities) = 1015 person-hours < 1200 person-hours.

Thus, the total person-hours are reduced due to ride-sharing, and the gap in person-hours will be even more significant as the AV market develops due to the reduction of ride-sharing costs over time.

This phenomenon is already observable in the United States' high-income urban households, where the number of privately owned cars per household decreases as the size of the ride-sharing market increases (Baltic et al., 2019). This substitution will also decrease the number of gas emissions and reduce traffic congestion. According to the McKinsey Center for Future Mobility, the predicted revenues from substitution towards

AVs will be 1.6 trillion dollars in 2030 (Deichmann et al., 2023). About one-third of the benefits will emerge from utilizing new parking spaces, and 15 percent are attributed to workers' productive commuting time (Heineke & Kampshoff, 2019). Further benefits include reducing environmental damage due to fewer cars on the street.

C. Entry and Exit Conditions

Due to the heavy type of competition in the AV market at its early stage, producers will need to achieve economies of scale to decrease costs and offer lower prices. Such a challenge has entry and exit conditions governed by multiple factors, as summarized in Fig. 3.5. We expect that entering the market will be fast and competitive with collaboration incentives. Because the industry and AV technology are still in their infancy at the commercial level, significant funds must be directed toward research and development (R&D) to achieve a cost-effective and functional AV with a high market share. Facing an oligopoly market with "big players", small and medium firms face high barriers to entering the market.

In the short to medium run, it is expected that there will be mergers and acquisitions (M&A) in the market, which will create barriers to entry to smaller firms and yield tacit collusion between large firms, but at the same time, also attain the benefits of supply-chain collaboration. There has already been some M&A activity in the market, which may intensify in the medium term. Currently, automotive suppliers and automakers are

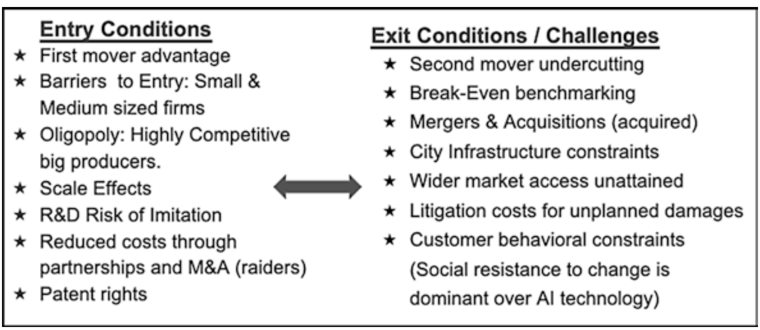

Fig. 3.5 Entry and exit in the new autonomous vehicle industry

forming alliances to share the costs of producing AVs. For example, BMW has formed an alliance with Intel to supply chips and processors for their AVs. This is only one of many other collaborations between technology suppliers and AV producers (Lienert & Sage, 2017).

In the long run, AV technology is expected to be available to all firms as open source; thus, the cost of production will decrease while allowing more firms to enter the market. Road infrastructure must be compatible with AVs, which will increase demand, enlarge the market size, and allow more producers to enter, hence relaxing the initial fast competition.

D. Buyer Choice

Buyer's choice has both positive and negative psychological factors, as shown in Table 3.3. Psychology studies have shown that there will be social resistance to AVs due to loss aversion, endowment effect, and status quo effect (Honan, 2019; Pankratz et al., 2017; Selim & Gad El-Rab, 2022). Loss aversion is when an individual's risk preference is to avoid losing rather than focus on gaining. The endowment effect means that people tend to overvalue what they already own (or are accustomed to) due to emotional attachment. In contrast, the status quo effect is the unwillingness to change the current status, which is simply resistance to change.

Moreover, people gain utility from owning their cars and being able to control them, which includes the psychological thrill of driving (Selim & Gad El-Rab, 2022). Additionally, consumers tend to mistrust and overestimate the risks of new technology. They do not fully trust driverless cars and tend not to feel as safe in them. Rare incidents of AV crashes negatively affect consumers' opinions and choices radically, and people tend to be more sensitive to an automated crash as compared to non-automated crashes (Honan, 2019; Pankratz et al., 2017).

People's choices in cars are affected by culture as well (Clark, 2021). For many, cars are not just considered a means of transportation but reflect a part of their identity. For example, some people partially buy electric or hybrid cars to show they are environmentally conscious. Hence, it is important to note that how society perceives cars is a major factor in car purchase decisions. In some cultures, cars are a sign of

Table 3.3 Buyer psychology factors for new autonomous vehicle purchase decisions

Positive factors (+ve)	Negative factors (-ve)
Less driving effort (less disutility of effort while driving an AV)	**Loss aversion** (risk preference to avoid losing rather than gaining, which relates to uncertainty of future happiness in buying a new AV)
Less constant burden of responsibility (AVs carry less constant responsibility in parking, cleaning, and maintenance)	**Endowment effect** (emotional attachment and more preference towards technology already owned and experienced, thus yielding less preference to AI-enabled AVs)
Environmentally friendly (AV cars are perceived as more environmentally friendly than non-AV cars)	**Status-quo effect** (behavioral resistance to change towards AVs as long as they are still uncommon cars in society)
Less traffic congestion (perception that more AV cars on the street will lead to less traffic congestion)	**Less thrill of driving** (due to automation features and driverless decisions in AVs)
Scenery and socializing (AVs have a better option of granting more time for relaxation with travel scenery, socializing, texting, and/or ride-sharing)	**Over-sensitivity to safety** (more sensitivity towards automated crashes in comparison to non-automated crashes)
Enhanced productivity (AVs enhance work productivity with the option of working while inside the vehicle as well as reduction of commuting time)	**Culture-dependence** (varying forms of independence and identity according to the location culture which creates uncertainty in purchase decision)
Technology adoption (being up-to-date in technology)	**Unreadiness of road infrastructure** (waiting time to purchase an AV until road infrastructure of driving locations is ready)

independence, social status, and personal freedom (Pankratz et al., 2017). Approximately 88 percent of households in the United States own at least one car. In contrast, only 6 percent of households in other countries, such as India, own at least one car, where cars are not as an important part of the culture. Lower car ownership per household is particularly common in developing countries.

Additionally, people may prefer AVs as they require less effort and less liability. For example, the disutility of driving in traffic congestion or driving while tired is a relative psychological benefit of AV adoption. Moreover, moving to AVs will reduce the liabilities and responsibilities of car ownership and driving. In essence, the driver will not carry the constant burden of finding a parking space, cleaning the car, paying for spare parts, and having to do the driving.

E. Input Supply Factors and Challenges

The growing AV industry is a technology-based industry in its input supply chain. The AV industry's supply chain has significant bargaining power in the hands of technology suppliers and innovators. This contrasts with the previous history of labor-intensive car production in which workers generally had low bargaining power. Radar, LiDAR, camera, and processor companies carry the primary technological components of any AV. They hold the soul of the AV car in their technology. Accordingly, the automotive market is witnessing increasing mergers and acquisitions between automotive producers, technical advancement companies, and start-ups. One example is Toyota's acquisition of Jaybridge, an engineering team working on automated solutions (Crowe, 2019).

The two main technological challenges of AV input supply are (1) to produce a cost-effective LiDAR system and (2) to develop a fast and power-efficient chip or Central Processing Unit (CPU).

LiDAR is the most expensive part of the AV. In addition, it is arguably the safest way the car can develop a map of its surroundings. All autonomous car producers are working to find a way to reduce the cost of LiDAR systems or develop an alternate system that can produce a precise three-dimensional map of the environmental surroundings of the vehicle. Luminar, a LiDAR producer that is working with Toyota, Audi, and Volkswagen, has reached a LiDAR-based system that costs less than half its LiDAR predecessors, and this can give an enormous advantage for the car brands working with it if scaled successfully to adoption (Crowe, 2019; Spencer, 2020). Google's Waymo started its own LiDAR production line and is selling it to other producers, hoping to reach low costs through economies of scale (Dolgov, 2021).

The second challenge for the input supply is CPUs and chips. The many sensors and cameras of an AV provide an enormous number of high-resolution images and high-frame scans of the environmental surroundings of the car. Then, these images and data are processed through the CPU using highly complex algorithms that plan a move and give the other car components an order to take action. All of this should happen in real-time, or the outcome can be disastrous. Moreover, data processing consumes much power. Therefore, producing a fast and power-efficient processing unit is critically important and differentiates one producer from another. In the market for AVs, there are a few input suppliers, such as Intel, Nvidia, and Qualcomm, that supply chips to all the AV producers. These companies have significant bargaining power in the AV industry.

Furthermore, acquisitions among technology suppliers are taking place, giving even *more* bargaining power to the technology raiders. For example, Intel acquired Mobileye, a leader in computer vision systems, for 15.3 billion dollars, one of the most expensive acquisitions in this line of business (Crowe, 2019; Lunden, 2017). Input supply factors are summarized in Fig. 3.6.

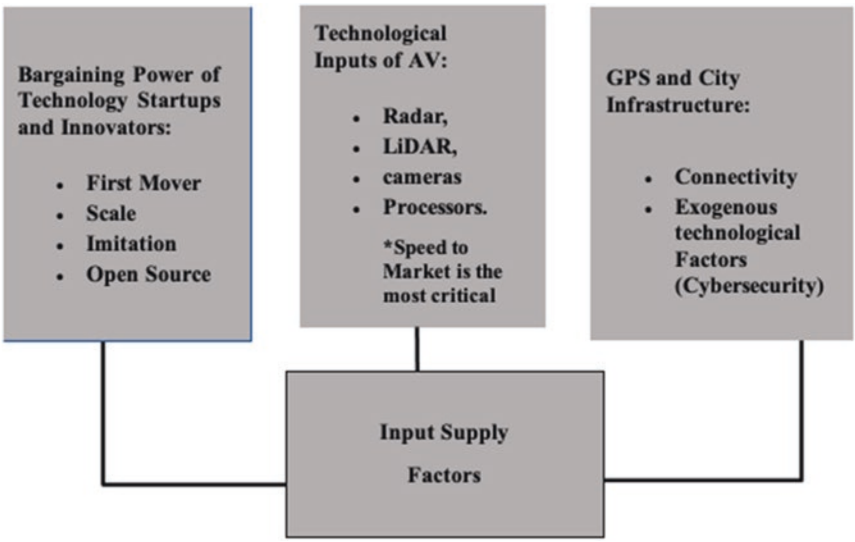

Fig. 3.6 Input supply factors

3.5 Conclusion and Path Forward

The self-driving car industry is expected to be driven by critical factors for its sustainable future. This includes a reduced-cost AI system led by pulsed laser LiDAR with a sufficient loop frequency and GPS bi-directional cloud technology. Also, R&D collaboration is highly anticipated across the input supply chain, which is expected to lead to an open-source technology in the long-run but with homogenous kinked pricing and an oligopoly market structure associated with barriers to entry in the short-run for SMEs. The industry is most likely expected to yield similar products but with business differentiation using first-mover market advantage coupled with the business risk of second-mover undercutting in both price and technology. A time lag in road infrastructure slower than accelerated AI adoption is highly expected. Demand behavior is expected to be driven by customer loss aversion and initial behavioral resistance followed by social acceptance and technological adoption leading to less traffic congestion. Major changes are expected in the associated insurance industry, primarily towards pooled insurance accountability rather than individual liability, as well as new strictly enforced safety regulations using pre-crash, crash, and post-crash safety performance standards. This will require adaptive government regulations that could disrupt the industry's sustainability. Additionally, large investments in smart city infrastructure are expected, with a sharp digital divide across transport regions leading to more regional inequality. Our research also suggests that AV consumers, at the current stage of AI technology adoption, strongly prefer a human-controlled semi-autonomous vehicle rather than complete machine autonomy.

Further research for the sustainability of the AV industry will require studies on the extent of post-merger R&D cost reductions, the impact of "big players" on technology adoption, frequent customer surveys at the global level, and additional research on the intensity of the digital divide caused by AI advancement in the self-driving car industry. Lastly, additional research is needed on the substitution and interplay between AV and electric cars for the future of global cars.

References

Anderson, J., Nidhi, K., & Stanley, K. D. (2016). *Autonomous vehicles technology: A guide for policymakers* (pp. 443–442). RAND Corporation, Research Report.

Baltic, T., Cappy, A., Hensley, R., & Pfaff, N. (2019). How sharing the road is likely to transform American mobility. *McKinsey & Company.* April.

Billington, J. (2019). The future of autonomous cars: A guide to what self-driving projects car makers and tech giants are working on right now. *Autonomous Vehicle International*. March.

Burns, L., & Shulgan, C. (2018). *Autonomy: The quest to build the driverless car and how it will reshape our world*. Ecco Publishers.

Clark, J. (2021). *Smart cities*. MIT Technology Review. April 2021.

Crowe, S. (2019). Waymo LiDAR sensor now available for robotics companies. *The Robot Report*. March. https://www.therobotreport.com/waymo-LiDAR-sensor-now-available/

Deichmann, J., Ebel, E., Heineke, K., Heuss, R., Kellner, M., & Steiner, F. (2023). Autonomous driving's future: Convenient and connected. *McKinsey & Company*. January.

Dhawan, R., Hensley, R., Padhi, A., & Tschiesner, A. (2019). Mobility's second great inflection point. *McKinsey Quarterly*, February.

Dolgov, D. (2021). How we've built the world's most experienced urban driver. *Waymo*. August 19. https://waymo.com/blog/2021/08/MostExperiencedUrbanDriver.html

Fu-Lee, K., & Qiufan, C. (2021). *AI 2041: Ten visions for our future*. Currency Publications.

Heineke, K., & Kampshoff, P. (2019). The trends transforming mobility's future. *McKinsey Quarterly.* March.

Heineke, K., Kampshoff, P., Mkrtchyan, A., & Shao, E. (2017). Self-driving car technology: When will the robots hit the road?. *McKinsey & Company*. May.

Honan, M. (2019). Autonomous driving: Safety first. MIT Technology Review. March Issue.

Jurgen, R. (2013). *Autonomous vehicles for safer driving*. SAE International.

Karkus, P., Ivanovic, B., Mannor, S., & Pavone, M. (2022). A differentiable and modular control stack for autonomous vehicles. In *Conference on robot learning*. Self-Driving Cars Technology & Solutions. NVIDIA. December.

Lienert, P., & Sage, A. (2017). Automakers, suppliers team up to share costs of self-driving cars. *Reuters*. January. https://tinyurl.com/mv98yyzx

Lunden, I. (2017). Intel buys Mobileye in $15.3B deal. *TechCrunch*. March. https://techcrunch.com/2017/03/13/reports-intel-buying-mobileye-for-up-to-16b-to-expand-in-self-driving-tech/

Lynberg, M. (2019). *Automated vehicles for safety. National Highway Traffic Safety Administration (NHTSA)*. United States Department of Transportation. Washington DC. August.

Magias, M. (2023). *Implications of self driving vehicles on public policy with Tony Gómez-Ibáñez*. Harvard Kennedy School: Faculty and Research.

Manyika, J., & Russell, S. (2020). *How to ensure artificial intelligence benefits society*. McKinsey Global Institute.

Ozguner, U., Acarman, T., & Redmill, K. (2012). *Autonomous ground vehicles*. Artec House Publishers.

Pankratz, D. M., Willigmann, P., Kovar, S., & Sanders, J. (2017). *Framing the future of mobility: Using behavioral economics to accelerate consumer adoption.* Deloitte Review, Issue 20.

Philipp, K., & Padhi, A. (2017). *Autonomous-driving disruption: Technology, use cases, and opportunities*. McKinsey & Company. November.

Schwab, K. (2016). *The fourth industrial revolution.* Currency Publications.

Selim, T., & Gad El-Rab, M. (2022). Disruptive Innovation in the Global Auto Industry: New Findings, *Annual Conference of Industry Studies Association,* Wharton School, University of Pennsylvania, 2022.

Spencer, L. (2020). The price of LiDAR is falling: Will driverless cars be on the road sooner? *ITU News*. April 24.

Voelzke, U., & Rudolph, G. (2023). *Three sensor types drive autonomous vehicles.* Evnoia Group Publications.

Wakabayashi, D. (2018). Self-driving Uber car kills pedestrian in Arizona, Where Robots Roam. *The New York Times*. March 19.

Wayland, M. (2019). Fiat Chrysler and Peugeot are behind on electric and autonomous vehicles, but their $50 billion merger should help them catch up. *CNBC*. December 18.

Weber, A. (2022). *New Technology for Bonding Automotive Displays*. Assembly Magazine: Autonomous and Electric Mobility. July 11.

World Economic Forum. (2023). *Global council on the future of artificial intelligence*. Strategic Intelligence. https://www.weforum.org/communities/global-future-council-on-artificial-intelligence.

4

The Emergence of the Nighttime Artificial Intelligence-Robot-Driven Economy

Steve Lee, Won-Yong Oh, and Irene Yi

4.1 Introduction

Artificial intelligence (AI) and robots do not sleep. They do not tire, nor do they become distracted, and they can work without interruptions. For AI and robots, nighttime may be the optimal time to perform tasks with minimal human intervention and better access to energy and resources. Automated factories and production facilities have been observed to function effectively at night. They could be advanced to operate

S. Lee (✉)
Aprilli Design Lab, Toronto, ON, Canada
e-mail: slee@aprilli.com

W.-Y. Oh
Lee Business School, University of Nevada, Las Vegas, NV, USA
e-mail: wonyong.oh@unlv.edu

I. Yi
Department of Finance, Rotman School of Management, University of Toronto, Toronto, ON, Canada
e-mail: i.yi@rotman.utoronto.ca

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_4

autonomously in the future, managed by AI and utilizing AI-driven robots. Autonomous vehicles can be used for overnight shipping, delivery, and transportation, fully utilizing the roads while most humans are asleep. Furthermore, street cleaning, infrastructure maintenance, and construction can occur primarily at night, facilitated by technologies such as light detection and ranging (LiDAR) and ultrasound sensors.

In this chapter, we argue that the advancement of AI and robots will lead to the emergence of a "nighttime economy" (i.e., an economy driven primarily by AI and robots during human resting hours) as they open new horizons for economic activity. Unstaffed offices, retail stores, or distribution centers can all benefit from nighttime operations as energy costs are lower during off-peak hours when no lighting or air conditioning for humans is needed, and ambient temperatures are lower, making equipment-cooling easier (e.g., Seppanen et al., 2003). As AI and robotic technologies become more advanced, an increasing number of professions and businesses will benefit from the operational advantages and cost-effectiveness of nighttime activities. We argue that these advantages would be attractive enough to pave the way for an AI-robot-driven nighttime economy that is separate from and independent of the human-centered day economy (Agrawal et al., 2019; Zhuravlova et al., 2020).

Although AI is still in its early stages of development, the recent release of generative AI models has raised attention regarding the potential implications for humans as AI begins to reach or even surpass human cognitive capabilities. While generative AI is beginning to affect jobs in specific areas, the future state at which AI surpasses human capabilities would be a crucial turning point. Some refer to this state as Artificial General Intelligence (AGI), while others refer to it as superintelligence (Bostrom, 2014), which implies that AI outperforms humans in virtually all areas of interest. Scientists and futurists speculate that at a specific juncture called "the Singularity," technological growth will become uncontrollable and irreversible, leading to unforeseen and unpredictable changes in human civilization (Kurzweil, 2005). Because of the progressive nature of AI, studies of AI's future impact would vary depending on the specific context and period being considered. Some studies focus on the immediate benefits of early AI systems, such as AI-assisted management (Paschek et al., 2017) or hyper-automation (Haleem et al., 2021), which is a combination of automation and AI that augments human capabilities to manage and execute operations more efficiently. Others focus on subjects like human-machine collaboration (Bolton et al., 2018), where machines and robots still operate mostly under human guidance or control.

This chapter explores a more advanced future scenario based on the realization of AGI, in which AI possesses the intelligence to independently create and manage complex systems encompassing businesses, societies, governance, and economies. As AI rapidly revolutionizes digital operations and enhances intellectual capacities, AI-driven robots, drones, appliances, computers, and wearables are poised to bring about profound changes in the physical world. As AI-robot-driven operations gradually take on roles currently held by humans in business and society and continue to grow in scale, their activities will eventually expand to the point of establishing their own economy. We refer to this state as an AI-robot-driven economy, a term introduced in contrast to the human-driven economy, where AI and AI-driven robots autonomously operate and establish relations among themselves with minimal human intervention. Differing from concepts like hyper-automation or human-machine collaboration, the AI-robot-driven economy assigns AI and robots a more extensive and proactive role as they pursue distinct goals, purposes, and methods of execution. As the AI-robot-driven economy advances, AI-robot-driven businesses will develop their systems to adapt to their optimal conditions, leveraging underutilized locations and time zones from human-driven economies. This process will ultimately establish a balanced ecosystem between the AI-robot-driven world and the human world.

Divided into four parts, this chapter examines conditions contributing to the formation of the AI-robot-driven economy. The first part illustrates how AI and AI-driven robots will gradually replace jobs and subsequently impact human society and urban structures with their superior capabilities and economic advantages. In the second part of the chapter, we discuss the key characteristics of the Fourth Industrial Revolution and its related technologies, highlighting concepts such as decentralization, localization, and customization, which also characterize the AI-robot-driven economy. The third part illustrates how these conditions can lead to the formation of an AI-robot-driven circular economy and its potential contribution to sustainability. This part explores how significant improvements in the efficiency of value chain activities can be achieved by utilizing digital twins, recycling, refurbishing, and sharing. Lastly, the final section presents examples and implications across various fields, illustrating how AI-robot-driven operations can leverage nighttime periods to execute tasks effectively and efficiently. As technology advances, we foresee the AI-robot-driven and human-driven economies compensating for each other, creating balanced ecosystems. We compare these two economies in Table 4.1.

Table 4.1 The human-driven economy versus the artificial-intelligence-robot-driven economy

	Human-Driven Economy	AI-Robot-Driven Economy
1. Drivers	**Humans with tools**	**AI and AI-driven-robots**
2. Operational hours	**Daytime-based** Humans mostly work during the daytime and on workdays, with breaks during weekends, holidays, and vacations. Night shifts are occasionally required, depending on the profession.	**Nighttime-based** AI and robots can work 24/7 every day with minimal breaks. While they would work during the daytime as well, they tend to be more profitable and efficient during nighttime.
3. Locations and form of development	**Urban, centralized** Human-driven businesses benefit from being located within urban areas, where they have better accessibility to other businesses, human labor, human networks, infrastructure, and services.	**Suburban/rural, decentralized** AI-robot-driven businesses can operate more efficiently when they are among themselves. They would cluster around underutilized or outskirt areas where they can create their own ecosystem.

(*continued*)

Table 4.1 (continued)

	Human-Driven Economy	AI-Robot-Driven Economy
4. Labor cost	**Expensive** The cost of human labor, including salary, overhead, and office rents, is expensive and constitutes a major portion of operational costs. Human skills require education and training, while resources are inflexible.	**Inexpensive** AI is constantly up to date with all necessary information and skills and can operate multiple robots and equipment simultaneously. Robots can be shared across operations and businesses.
5. Energy cost and usage	**Expensive, high** Human workplaces require air conditioning, lighting, and equipment usage, which is concentrated during the daytime, including peak hours for electricity. Daily commute also demands a significant amount of energy.	**Inexpensive, low** AI and robots can operate during nighttime, taking advantage of cheaper off-peak electricity rates. They can also benefit from lower outdoor temperatures, resulting in reduced cooling loads.
6. Resources and infrastructure	**Concentrated** The human-driven economy operates predominantly during peak hours when electricity, internet, and infrastructure are heavily utilized. Access to resources requires substantial transportation and shipping.	**Distributed** The AI-robot-driven economy operates primarily during off-peak hours through distributed networks. Its human-less, digital nature provides better opportunities for localized production.
7. Sustainability	**Linear economy, human-driven** In a human-driven economy, products lose track of information when handed over to the customer. Recycling efforts become inefficient without prior information, resulting in a massive amount of waste.	**Circular economy, data-driven** AI and digital twins track parts and products throughout their lifecycles, aiding in component reuse and waste reduction. The sharing economy enhances the efficiency of product usage.

4.2 Transitions Toward an AI-Robot-Driven Economy

Human workers, operating within their physical and cognitive capabilities, have historically been at the forefront of economic activities. Humans primarily carried out work during the days, with time off during the nights and on weekends; thus, humans' need for time away from work determined the pace of work. While technological advancements, such as electricity, automobiles, computers, and the internet, have significantly improved human capabilities, productivity, and access to information, these technologies have not altered the fundamental reality that humans remained central to work and that technologies were tools to achieve tasks. During the Second Industrial Revolution between the mid-nineteenth century and the early twentieth century, the production boom augmented by machines and electricity created massive numbers of manufacturing jobs and new business opportunities. The Third Industrial Revolution, also known as the Digital Revolution, enabled people to gain instant access to all kinds of information through computers and the Internet. This revolution boosted productivity and enhanced connectivity, leading to globalization and increased worldwide economic activities. While these technological revolutions expanded human capabilities, humans were still at the center stage, performing tasks, making decisions, sharing knowledge, and running businesses. Rather than resulting in job reduction, these technological changes brought forth new business and work opportunities, growing the economy and providing compensation and motivation for humans to continue working.

However, the forthcoming Fourth Industrial Revolution will mark a significant departure from its predecessors due to the emergence of AI, which holds the potential to establish fully autonomous systems operating independently of human intervention. For the first time in history, AI and robots are poised to challenge the value of human work as AI exceeds human cognitive capabilities, and AI-driven robots surpass human physical abilities (Altman, 2021; Brynjolfsson & McAfee, 2014). The Fourth Industrial Revolution, as defined by Klaus Schwab, is a fusion of advances in AI, robotics, the Internet of Things (IoT), blockchain, 3D printing, genetic engineering, quantum computing, and many other technologies that collectively lead to a revolution in how things work (Schwab, 2017). With AI leading the transformation, these various technologies will collectively create digital ecosystems and new territories that will operate beyond human physical and cognitive capabilities. While this transition may introduce new opportunities for individuals, businesses, and societies, it will also disrupt the existing social system and economic landscape (Unger et al., 2018), leading to a re-evaluation of the significance attributed to work, life, wealth, and social values.

A. Job Replacements

As AI advances, many jobs will likely be replaced by AI or AI-driven robots (Gibbs, 2022; Nam, 2019). Although this substitution will likely begin with simpler labor and desk-based occupations, it could progressively extend to more complex and sophisticated positions, such as operational management, finance, accounting, PR, and sales. AI can quickly learn tasks and perform such jobs with superior speed, efficiency, and accuracy (Executive Office of the President, 2016; OECD, 2023). Most industries, such as manufacturing, construction, banking, transportation, healthcare, and service, will likely undergo a transition period wherein professionals collaborate to compile their knowledge into a comprehensive database through which Industry AIs can quickly learn, analyze, and utilize the most up-to-date information and standards. Consequently, tasks that traditionally required years or decades for human professionals to master could be consolidated and integrated into all-in-one AI systems, assuming the roles of bookkeepers, lawyers, job managers, drafters, and salespersons simultaneously. This shift may reduce human jobs and diminish the hierarchical structure within firms, as the traditional distinctions between job roles, responsibilities, ranks, and positions in human workplaces lose relevance. Instead, an integrated corporate AI controlling many robots for various tasks would be sufficient to run businesses autonomously. Over time, businesses may operate with minimal input and guidance from humans, reducing their need for human staff before finally becoming autonomous. These autonomous businesses will require fewer offices in urban environments, resulting in the decentralization of urban cores and the subsequent restructuring of human societies (Saltinski, 2015).

B. Operational Cost Reduction

AI-robot-driven businesses offer significant reductions in operational costs (Jovanović et al., 2018) compared to human-driven operations. While humans have traditionally been a crucial factor in driving businesses and the economy, they also carry substantial operational and production costs (Jung & Lim, 2020). These costs include human wages, overhead expenses, office rents, energy bills, workspace equipment, and maintenance fees. Human workspaces are notably resource-intensive and costly, requiring substantial materials, resources, and energy for construction and operation. Maintaining a suitable working environment for humans, such as maintaining specific temperature ranges, providing adequate lighting, and ensuring comfortable and productive interiors, add to these costs.

In contrast, robots and machines can operate in a broader temperature range and do not rely on well-finished interiors. Additionally, they often work more efficiently in environments where humans are absent. Unlike human-robot collaborative workspaces, fully autonomous workspaces do not require safety measures such as speed limits or safe distances between humans and robots (Unger et al., 2018). Without being limited to human scale, mobile autonomous robots can navigate open floors with enhanced flexibility and efficiency under precise real-time management of AI. This streamlined approach would further reduce costs in fully autonomous workspaces. AI-robot-driven businesses thus offer significant reductions in operational costs (Jovanović et al., 2018) compared to human-driven operations.

C. Workplace Relocation

AI-robot-driven businesses can also achieve substantial cost reductions regarding office rents, particularly for businesses in prime locations. Historically, human networks and labor supply have driven firms to maintain offices and workplaces in accessible and prime areas at a premium cost. Businesses cluster around urban centers to gain better access to clients, vendors, and a skilled workforce (Rosenfeld, 1997). As cities expand to accommodate growing populations, more people contend with higher housing expenses, traffic congestion, energy consumption, waste management challenges, gentrification, and crime rates. However, due to many businesses no longer needing physical offices for humans (Yun & Lee, 2019), and with AI-robot-driven workspaces moving to the outskirts or underutilized areas where they can operate more efficiently, urban areas will gradually decentralize. This shift will create more livable environments in urban centers and foster more balanced development in suburban regions.

4.3 Decentralization, Localization, and Customization

The advent of the Fourth Industrial Revolution is expected to accelerate the decentralization of economic activities and foster localization and customization, in contrast to the post-Second Industrial Revolution's mass production, centralization, and urbanization principles. This shift, driven by the AI-robot-driven economy, brings with it enhanced value chain optimizations and improved quality of life. AI, digital platforms, and real-time connectivity crucially facilitate direct and diverse connections, communication, and transactions between organizations, businesses, customers, and products. Digital technologies such as Augmented Reality (AR) and Virtual Reality (VR) enable overcoming physical barriers, while AI interpreters assist in overcoming psychological, linguistic, and cultural barriers, bringing people closer together globally. The increased connectivity between devices, vehicles, homes, infrastructure, humans, businesses, and governments fosters greater coordination, efficiency, and optimized control in operations. Blockchain technology facilitates direct economic transactions (Kowalski et al., 2021), while autonomous driving technology offers affordable door-to-door transportation solutions. Regarding product deliveries, businesses can directly send products to end consumers, eliminating the need for distribution centers, warehouses, or retail shops. These transformative changes collectively enhance sustainability efforts and foster a more fully circular economy (Oh et al., 2021; Ramadoss et al., 2018), emphasizing resource efficiency and waste reduction.

A. Decentralization

As AI-robot-driven workplaces and businesses relocate to suburban areas and the transition to unstaffed operations increases, urban areas will gradually decentralize, presenting significant opportunities to create improved living spaces for human societies (see Fig. 4.1). Since the Second Industrial Revolution, major cities worldwide have experienced rapid growth due to the demand for human labor and the necessity for businesses to establish proximity to supply chains, clients, and human society. Consequently, the increasing demand for greater access to housing and comprehensive human services has driven expansion and densification to accommodate growing populations.

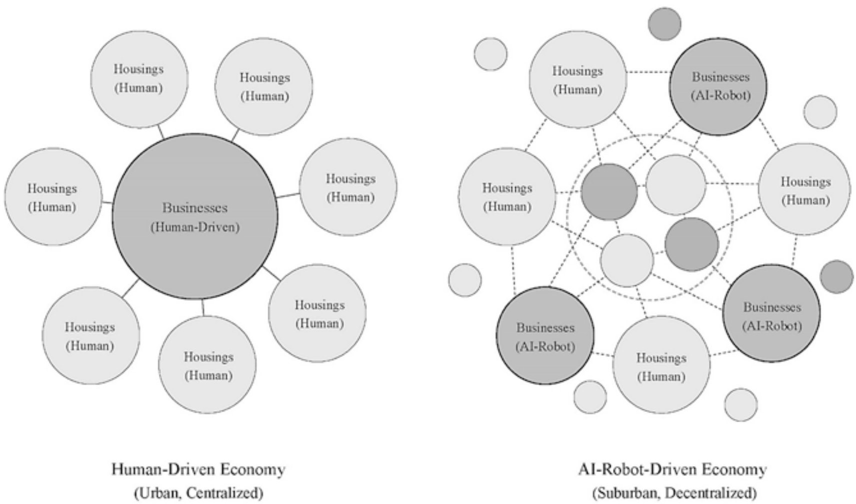

Fig. 4.1 Urban structures of a human-driven economy versus an artificial-intelligence-robot-driven economy

While business and job opportunities have historically been the most significant factors attracting individuals to urban areas (Faggian et al., 2017), this paradigm is poised to change in the future. As AI-robot-driven businesses advance their operations, urban cores will rapidly lose attraction with reduced human job opportunities and business activities. With technology promoting more work-from-home and flexible work conditions, off skirt areas will become more appealing with better living conditions and lower human costs. In this context, daily commutes to work can be reduced, resulting in considerable time and energy savings for individuals, reducing traffic congestion and decreasing reliance on mass transportation. From a broader societal perspective, reducing the number of individuals commuting to and working in office spaces will significantly decrease carbon footprints and bring progress to sustainability efforts.

B. Localization

In contrast to the previous era of globalization, where production was often outsourced to developing economies for lower labor costs, a notable future trend is manufacturing localization (Grossman & Rossi-Hansberg, 2008). This shift toward localization will lead to cost savings in transporting goods worldwide. As labor costs become less influential in production expenses, the focus will shift toward materials, parts, energy, and transportation costs. Rather than depending on inexpensive labor from developing nations, future autonomous manufacturing will leverage the advantages of local communities and supply chains, resulting in reduced shipping costs and quicker delivery to customers. This transition will go beyond reducing carbon footprints by reducing the long-haul transportation of consumer goods providing better-customized local solutions tailored to human needs. Consequently, many technology companies adopt AI-driven local manufacturing systems (Bogue, 2014).

Although not yet realized, companies like Tesla are planning to establish fully autonomous "giga factories" on each continent (Naor et al., 2021), allowing the company to produce and deliver vehicles directly to customers, thereby reducing their reliance on transportation, dealerships, and warehousing costs. Emerging automotive companies, such as Arrival, are also embracing localized micro-factories equipped with robotic assemblies strategically located to meet regional demands and provide customized solutions. The localization approach will yield significant reductions in transportation costs, enhance resiliency within local communities, and achieve a more even balance between supply and demand.

C. Customization

The emergence of an AI-robot-driven economy will lead to a substantial transition towards customization across all aspects of the economy. This transformation encompasses not only products but also the processes of production, retail, and delivery. Unlike traditional automated factories (Koinoda et al., 1984), autonomous factories will leverage AI's capabilities to operate advanced production lines independently while maintaining flexibility and versatility (Rong et al., 2016). AI-driven robots can learn and perform diverse tasks akin to human intelligence, setting them apart from machines or factories programmed for singular functions. AI-driven Robots can simultaneously become excellent chefs, builders, musicians, artists, and soldiers with AI at the helm.

In this paradigm, various combinations of AI-driven robots can be dynamically assembled as a team or individually, hourly, to undertake specific tasks (Østergaard, 2018). The AI orchestrates the collaborative robots' assignments, ensuring efficient utilization and resource allocation. Each factory will be equipped to manage a wide range of functions, enabling customized solutions for each standardized base model, such as scale, design, material, or accessories. Mass customization and personalization will become the norm in AI-robot-driven factories, diverging from the traditional mass production approach. Just as a fast-food chain's sandwiches can be individually customized for customers, product lines can deliver tailored solutions to cater to diverse customer needs. Integrating customization into the production process enhances efficiency and creates more valuable and functional goods for customers without overexploiting resources. This transformation aligns to produce sustainable and efficient outcomes by leveraging AI and robotics in manufacturing.

4.4 AI-Robot-Driven Circular Economy (CE) and Value Chain Optimization

A. Intelligent Lifecycle Product Management and Recycling Economy

AI is poised to revolutionize material and product management by enabling comprehensive information tracking throughout the entire lifecycle, from material extraction to production, consumption, and recycling (Wilson et al., 2022). Digital twins, which are digital representations of products or systems, will be used to track components and provide valuable insights and data-driven decisions. Throughout the entire product lifecycle, AI can employ digital twins, product manuals, disassembly guides, and recycling value chain information to track each component effectively (Shennib & Schmitt, 2021). In the future, AI-driven robots will also have the potential to enhance the productivity and efficiency of the recycling market significantly (Bogue, 2019). Consequently, when a product reaches its end-of-life stage, it can be easily and fully disassembled into individual parts and materials within a disassembly facility. Using a reverse logistics process (Hedberg & Šipka, 2021), AI can facilitate the refurbishment and reintroduction of individual components and materials into the production cycle. As the AI-robot-driven Circular Economy (CE) progresses (Oh et al., 2021), products can be designed to be easily disassembled and reused from the outset, thus enhancing the value chain's efficiency as it evolves.

As illustrated in Fig. 4.2, the AI-robot-driven economy presents opportunities for growth in the recycling and refurbishing market, as well as in the used items market and the sharing market. Currently, when products reach the end of their lifecycle, they lose their connection to the production process and become useless. However, with digital twins in the future, reusable parts and products can be recycled and reintegrated into the production cycle more efficiently (Tozanlı et al., 2020). In the past, disassembly for recycling has required significant labor, knowledge, and skills compared to the initial production process. With the advent of AI-driven robots that have access to comprehensive data and knowledge about parts and materials, the recycling market can become more

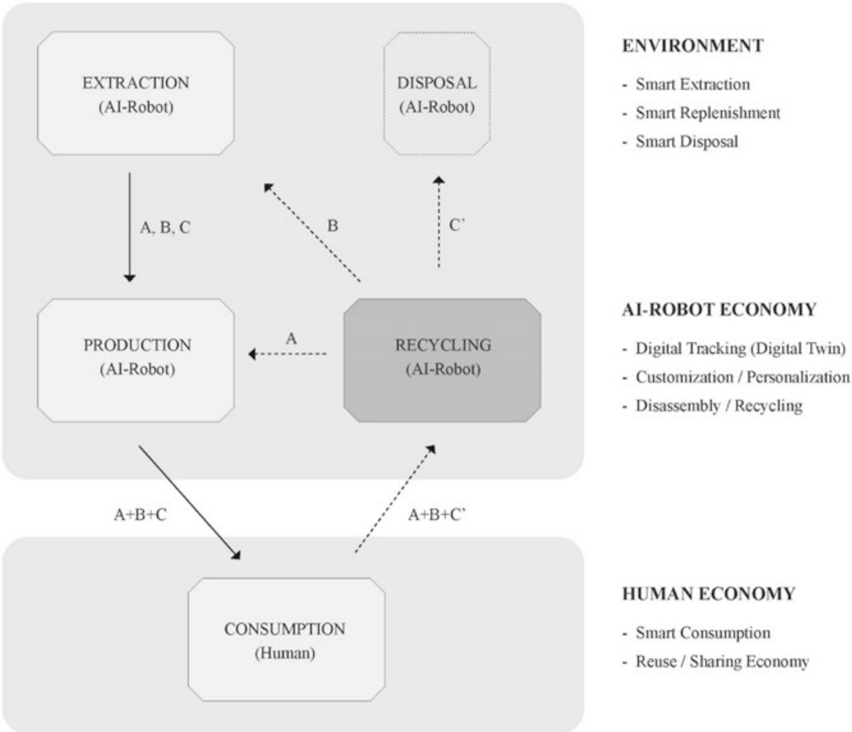

Fig. 4.2 Artificial-intelligence-robot-driven circular economy

appealing. This shift can reduce overproduction and minimize waste (Gupta et al., 2019). Over time, AI can continuously learn and improve existing recycling systems, creating more efficient processes that align closely with the production process. By utilizing AI's capabilities for resource utilization and waste minimization, the CE can move closer to achieving near perfection.

B. Sharing Economy

Integrating AI and digital platforms can substantially enhance the sharing economy, encouraging improved access to shared products and reducing the overall demand for goods. Numerous studies have indicated that implementing a well-established sharing economy, including car-sharing (Chen & Kockelman, 2016), peer-to-peer product-sharing (Martin et al., 2019), and the sharing of home appliances like laundry machines (Wasserbaur et al., 2020), may vastly reduce carbon footprints (Christis et al., 2019; Liu et al., 2018). While consumption remains primarily a human-driven activity, technology can enable the more efficient utilization of assets, crucially reducing overproduction. Many products worldwide, such as personal vehicles, computers, office spaces, parking lots, machines, hobby tools, and cooking appliances, are utilized for only a few hours a day, or during specific periods. Products are often discarded before the end of their lifecycles, frequently with minimal use. Accordingly, implementing a system where AI can track these products and provide a robust sharing platform would yield substantial societal benefits.

The sharing of products through enhanced management—facilitated by AI—can lead to significant advancements in ridesharing, tool-sharing, robot-sharing, and space-sharing. AI-driven businesses can offer superior information, maintenance, and security through applications and digital platforms. Factories and businesses can share or rent robot labor, tools, and workspace to reduce costs and enhance operational flexibility. Moreover, the sharing economy can foster the recycling market's growth by enabling more efficient usage of products through rental platforms or corporate initiatives while maintaining inventory management and updates. Overall, the convergence of AI and digital platforms through the sharing economy holds great potential for optimizing resource use, promoting sustainability, and facilitating a more efficient and collaborative societal framework.

4.5 The Emergence of the AI-Robot-Driven Nighttime Economy

The AI-robot-driven nighttime economy assumes that AI and robots will fill most roles currently occupied by humans sometime in the future; part of this transition includes AI and robots' search for optimal conditions, including workspace locations, working hours, spatial arrangements, organizational structure, and supply chain relationships, to maximize the efficiency and sustainability of their operations. AI-robot-driven facilities will autonomously create local clusters in cities' outskirts for better efficiency and safety. They will share workspaces, robots, energy, and infrastructure within these clusters, forming a localized network value chain. These facilities can be situated in areas less attractive to humans or in suboptimal urban settings for human activities (Erdoğan, 2019), such as lots close to highways, railroads, or underground locations. Networks of AI-robot-driven facilities can collectively share robots, tools, spaces, and data on a per-project basis to create a more flexible and versatile production system grounded in sharing and collaboration.

While the human-driven economy operates during the daytime, AI-robot-driven facilities will capitalize on nighttime hours when humans are typically asleep. This time advantage enables them to benefit from lower energy costs, improved resource access, and enhanced transportability (Agrawal et al., 2019; Zhuravlova et al., 2020). As a symbiotic relationship, the human-driven day economy and the AI-robot-driven nighttime economy serve different purposes and operate within their respective periods, allowing for the complementary utilization of resources and capabilities. Figure 4.3 illustrates future industries and tasks that can benefit from AI-robot-driven nighttime operations.

A. Nighttime Cost Savings – Energy, Internet, and Infrastructure

One of the critical reasons AI-robot-driven businesses will primarily operate at night is the cost savings from reduced electricity, internet, and infrastructure expenses. As AI-driven facilities rely on servers, computers, robots, batteries, and other equipment that require electricity and internet connectivity, these costs comprise a sizable portion of their operational expenses. Businesses can benefit from lower costs by leveraging off-peak hours and avoiding peak usage when humans consume the most electricity and internet services. Many countries offer Time-Of-Use (TOU) rates, which vary energy prices based on usage time. For instance, in Ontario, Canada (Ontario Energy Board, n.d.), electricity rates during ultra-low overnight hours (11 p.m.–7 a.m. every day) can decrease to as low as 2.4 ¢/kWh, in contrast to the 15.1 ¢/kWh charged during on-peak hours (from 7 a.m. to 11 a.m. and 5 p.m. to 7 p.m. on weekdays, or

Fig. 4.3 Implications of the artificial-intelligence-robot-driven nighttime economy

from 11 a.m. to 5 p.m. on weekdays during the summer). AI-robot-driven facilities can benefit from these significantly lower energy rates by operating during nighttime.

Furthermore, the overall energy usage of AI-robot-driven facilities can be significantly reduced at night compared to during the day, as a substantial portion of daily energy costs is attributed to cooling robots, servers, and equipment. In unstaffed "lights-out factories," where human presence is unnecessary, notable cost reductions in energy bills have been observed due to the absence of air conditioning and lighting required for human comfort (Lee, 2018). Moreover, as cooling is essential for robots, servers, batteries, and electric equipment to function correctly, the lower ambient temperatures at night can help offset cooling loads compared to daytime operations (Seppanen et al., 2003). This natural temperature advantage contributes to the overall cost savings in energy consumption while operating at night. It allows businesses to leverage shared infrastructure, resources, supplies, space, and transportation more effectively.

During nighttime, accessibility to these shared resources will improve, enabling efficient utilization and further cost optimization.

B. Autonomous Nighttime Delivery and Transportation

AI-robot-driven businesses can leverage the night as the primary delivery window for goods, supplies, and materials, as this time of day presents several advantages. For example, road traffic is considerably reduced at night, providing a smoother and more efficient transportation experience for autonomous vehicles and robots. Reduced congestion enables faster and more efficient delivery for both the transportation of goods and human traffic. In addition, nighttime delivery can be much safer with fewer humans driving on the road. Historically, human-related accidents are more prevalent during nighttime due to factors like lower visibility, fatigue, drunk driving, erratic behavior, and misjudgments.

In contrast, with the advent of autonomous vehicles, AI can provide safer and more accurate decision-making on roads, making night a more attractive time for transportation. By operating during the nighttime, autonomous vehicles and robots can function with increased safety and reliability, eliminating human factors and related risks. With AI and robots managing shipping, delivery, and reception, products can be made readily available for humans before the start of the day.

The relative lack of human interference during nighttime operations further enhances efficiency. Robots and vehicles can operate at night with enhanced speed and precision when not navigating human obstacles; this streamlined operation is particularly beneficial for industries that involve production, construction, or transportation, where the swift and safe movement of materials and goods is essential for uninterrupted production lines. By assigning delivery vehicles, construction vehicles, waste trucks, and maintenance vehicles to operate at night, streets can become less congested during the day, allowing for smoother human transportation experiences. Diversifying the traffic load on streets by increasing nighttime operations can optimize resource utilization and improve overall traffic flow for daytime human usage. Furthermore, in a sharing economy, autonomous vehicles present opportunities to optimize transportation costs by utilizing them for passenger delivery during the daytime and transitioning to product delivery during the nighttime. This approach can lead to further cost reductions.

Implementing localized supply chains and digital platforms allows customers to place orders during the daytime, leading to the manufacturing of custom-made products overnight and their subsequent delivery before sunrise. By accepting orders throughout the day and organizing overnight production, this model effectively highlights the possibilities of capitalizing on nighttime hours to facilitate direct customer delivery and maintaining equilibrium between supply and demand within local supply chains. As a result, adopting this approach can significantly enhance the overall efficiency of the value chain (Zhuravlova et al., 2020).

C. Nighttime Construction, Infrastructure, and Street Maintenance

Another field that can benefit from AI-robot-driven operations is construction. Traditionally, construction projects heavily rely on daylight due to factors like visibility, project scale, worker safety, and the need for equipment transitions between separate phases. In the future, construction processes will be reinvented based on the capabilities of AI and robots, overcoming human factors and physical limitations such as material dimensions, scale, working environment, and visibility. By incorporating AI management on construction sites, the coordination of equipment, robot workers, and resources can be improved, reducing waste and enhancing efficiency. A critical advantage of utilizing construction robots is the ability to continue work even in weather conditions hostile to outdoor human labor, such as rain and cold, which often cause construction projects to halt. This uninterrupted progress can significantly reduce the overall construction time and cost. Furthermore, utilizing off-site autonomous factories to create modular units near resources and subsequently delivering them to the construction site for assembly could significantly enhance the overall efficiency of the construction process. By integrating AI-robot-driven operations, construction projects can be conducted around the clock, including at night. This extended working time allows for accelerated development projects, reducing the overall construction period.

As a result of AI integration, street maintenance, waste management, and cleaning can also experience increased efficiency and frequency, particularly during nighttime operations. Using an integrated control platform powered by AI, municipal facilities can optimize their tasks and operations during nighttime hours. One area where this can be particularly beneficial is waste management (Gupta et al., 2019). By leveraging AI and robotics, waste collection can be conducted more frequently and thoroughly, resulting in cleaner streets and buildings while also promoting better waste management practices and environmental sustainability. Furthermore, integrating AI-driven recycling facilities can enhance the recycling process and contribute to waste reduction efforts.

In summary, integrating AI and robotics in street maintenance, waste management, and cleaning operations can enable more efficient and frequent services at night. This nightly, AI-robot-driven service can improve waste management practices, reduce waste volume, and lower congestion in urban areas. Overall, AI-robot-driven nighttime operations may enhance the quality of life in cities by optimizing municipal services and creating a more pleasant urban environment.

4.6 Conclusion

This chapter portrays a future landscape wherein AI-robot-driven businesses progressively gain autonomy, eventually establishing an independent economy and ecosystem. We propose a scenario wherein an AI-robot-driven economy evolves into a CE, optimizing the use of local materials and effectively tracking products and components throughout their entire lifecycle, facilitating comprehensive recycling capabilities. Businesses can significantly reduce transportation and storage costs by adopting localized production and delivery approaches. The emergence of AI-robot-driven businesses in outskirt areas will catalyze local clusters and networks that promote the sharing of resources and equipment, enabling the formation of self-sustaining supply chains. Consequently, urban centers experience decentralization and increased recreational space for human inhabitants. In addition, we explore a scenario in which

AI and robots leverage nighttime hours to maximize resource utilization and gain further independence and autonomy from human society.

As Nick Bostrom highlights in his book *Superintelligence* (2014), the timing of and speed at which AI surpasses human cognitive abilities will be a crucial factor for human civilization in the future. If AI progresses too rapidly without sufficient preparation, humans and societies might encounter social problems where, for example, AI dominance impedes human employment and economic participation. Without trickle-down effects, the concentration of wealth in a few advanced technology companies would exacerbate inequality and lead to social issues such as mass unemployment (Altman, 2021). As was the case in the aftermath of the Second Industrial Revolution, human societies would likely undergo significant conflicts and social upheavals before developing better ways and systems to maintain societal peace, order, and harmony amidst significant economic growth.

Inequality will likely be a significant issue during the Fourth Industrial Revolution among people within local and global societies more broadly. During the Fourth Industrial Revolution, the advancement of AI-driven economies is expected to widen the technological gap between developing and developed countries. The nature and extent of transitions occurring in developed countries equipped with high AI-robot capabilities will significantly differ from those in developing countries heavily reliant on low-skilled, labor-intensive industries. Leveraging low labor costs, as done previously, may become increasingly challenging for less-developed countries seeking to compete with developed nations. Therefore, the technological disparities between countries could contribute to social issues associated with wealth inequality.

In this context, creating a robust framework for a sustainable world where AI, robots, humans, and nature coexist harmoniously is an urgent task, but it faces significant obstacles (Nishant et al., 2020). Currently, AI and technological advancements are primarily driven either by profit-oriented businesses in market-driven societies or by state-controlled entities in totalitarian regimes. These entities compete for global dominance, profitability, and power rather than prioritizing the overall well-being of humanity, so the present situation elicits concerns. The inherent secrecy and competitiveness surrounding technological advancements make it challenging for global organizations or governments to comprehend the technology thoroughly; similarly, such secrecy impedes the establishment of comprehensive frameworks and regulations to ensure technology's utilization for the benefit of society at large. While AI holds tremendous potential as a future tool, its ultimate impact depends on the underlying human intentions driving its development. With the right intentions, transparency, and careful implementation, AI can become a powerful instrument to create prosperous, balanced, and sustainable future human societies.

References

Agrawal, A., Gans, J., & Goldfarb, A. (Eds.). (2019). *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Altman, S. (2021). Moore's law for everything. *Blog*, https://moores.samaltman.com

Bogue, R. (2014). What future for humans in assembly? *Assembly Automation, 34*(4), 305–309.

Bogue, R. (2019). Robots in recycling and disassembly. *Industrial Robot: The International Journal of Robotics Research and Application*.

Bolton, C., Machová, V., Kovacova, M., & Valaskova, K. (2018). The power of human–machine collaboration: Artificial intelligence, business automation, and the smart economy. *Economics, Management, and Financial Markets, 13*(4), 51–56.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Chen, T. D., & Kockelman, K. M. (2016). Carsharing's life-cycle impacts on energy use and greenhouse gas emissions. *Transportation Research Part D: Transport and Environment, 47*, 276–284.

Christis, M., Athanassiadis, A., & Vercalsteren, A. (2019). Implementation at a city level of circular economy strategies and climate change mitigation – the case of Brussels. *Journal of Cleaner Production, 218*, 511–520.

Erdoğan, G. (2019). Land selection criteria for lights out factory districts during the industry 4.0 process. *Journal of Urban Management, 8*(3), 377–385.

Executive Office of the President. (2016). *Artificial intelligence, automation, and the economy*. https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Faggian, A., Partridge, M., & Malecki, E. J. (2017). Creating an environment for economic growth: Creativity, entrepreneurship or human capital? *International Journal of Urban and Regional Research, 41*(6), 997–1009.

Gibbs, M. B. (2022). *How is new technology changing job design?* IZA World of Labor.

Grossman, G. M., & Rossi-Hansberg, E. (2008). Trading tasks: A simple theory of offshoring. *American Economic Review, 98*(5), 1978–1997.

Gupta, P. K., Shree, V., Hiremath, L., & Rajendran, S. (2019). The use of modern technology in smart waste management and recycling: Artificial intelligence and machine learning. *Recent advances in computational intelligence*, 173–188.

Haleem, A., Javaid, M., Singh, R. P., Rab, S., & Suman, R. (2021). Hyperautomation for the enhancement of automation in industries. *Sensors International, 2*, 100124.

Hedberg, A., & Šipka, S. (2021). Toward a circular economy: The role of digitalization. *One Earth, 4*(6), 783–785.

Jovanović, S. Z., Đurić, J. S., & Šibalija, T. V. (2018). Robotic process automation: Overview and opportunities. *International Journal Advanced Quality, 46*(3–4), 34–39.

Jung, J. H., & Lim, D. G. (2020). Industrial robots, employment growth, and labor cost: A simultaneous equation analysis. *Technological Forecasting and Social Change, 159*, 120202.

Koinoda, N., Kera, K., & Kubo, T. (1984). An autonomous, decentralized control system for factory automation. *Computer, 17*(12), 73–83.

Kowalski, M., Lee, Z. W., & Chan, T. K. (2021). Blockchain technology and trust relationships in trade finance. *Technological Forecasting and Social Change, 166*, 120641.

Kurzweil, R. (2005). The singularity is near. In *Ethics and emerging technologies* (pp. 393–406). Palgrave Macmillan.

Lee, N. K. (2018). Total automation: The possibility of lights-out manufacturing in the near future. *Missouri S&T's Peer to Peer, 2*(1), 4.

Liu, Z., Adams, M., Cote, R. P., Chen, Q., Wu, R., Wen, Z., Liu, W., & Dong, L. (2018). How does circular economy respond to greenhouse gas emissions reduction: An analysis of Chinese plastic recycling industries. *Renewable and Sustainable Energy Reviews, 91*, 1162–1169.

Martin, M., Lazarevic, D., & Gullström, C. (2019). Assessing the environmental potential of collaborative consumption: Peer-to-peer product sharing in Hammarby Sjöstad, Sweden. *Sustainability, 11*(1), 190.

Nam, T. (2019). Citizen attitudes about job replacement by robotic automation. *Futures, 109*, 39–49.

Naor, M., Coman, A., & Wiznizer, A. (2021). Vertically integrated supply chain of batteries, electric vehicles, and charging infrastructure: A review of three milestone projects from theory of constraints perspective. *Sustainability, 13*(7), 3632.

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management, 53*, 102104.

OECD. (2023). *OECD employment outlook 2023: Artificial intelligence and the labour market*. OECD Publishing. https://doi.org/10.1787/08785bba-en

Oh, W. Y., Chang, Y. K., Park, J. H., & Han, S. (2021). Toward a sustainable paradigm: Circular economy solutions in the fashion industry. In *Research handbook of innovation for a circular economy* (pp. 47–58). Edward Elgar Publishing.

Ontario Energy Board. (n.d.). Consumer information and protection: Electricity rates. *Ontario Energy Board*. https://www.oeb.ca/consumer-information-and-protection/electricity-rates

Østergaard, E. H. (2018). *Welcome to industry 5.0*. Retrieved Febr, 5, 2020. https://info.universal-robots.com/hubfs/Enablers/White%20papers/Welcome%20to%20Industry%205.0_Esben%20%C3%98stergaard.pdf?submissionGuid=00c4d11f-80f2-4683-a12a-e821221793e3

Paschek, D., Luminosu, C. T., & Draghici, A. (2017). Automated business process management–in times of digital transformation using machine learning or artificial intelligence. In *MATEC web of conferences* (Vol. 121, p. 4007). EDP Sciences.

Ramadoss, T. S., Alam, H., & Seeram, R. (2018). Artificial intelligence and internet of things enabled circular economy. *International Journal of Engineering and Science, 7*(9), 55–63.

Rong, W., Vanan, G. T., & Phillips, M. (2016, September). The internet of things (IoT) and transformation of the smart factory. In *In 2016 international electronics symposium (IES)* (pp. 399–402). IEEE.

Rosenfeld, S. A. (1997). Bringing business clusters into the mainstream of economic development. *European Planning Studies, 5*(1), 3–23.

Saltinski, R. (2015). Techstorm 2030: Restructuring future society. *National Social Science Technology Journal, 5*, 27–32.
Schwab, K. (2017). *The fourth industrial revolution*. Currency.
Seppanen, O., Fisk, W. J., & Faulkner, D. (2003). Cost benefit analysis of the night-time ventilative cooling in office building.
Shennib, F., & Schmitt, K. (2021, October). Data-driven technologies and artificial intelligence in circular economy and waste management systems: A review. In *2021 IEEE International Symposium on Technology and Society (ISTAS)* (pp. 1–5). IEEE.
Tozanlı, Ö., Kongar, E., & Gupta, S. M. (2020). Evaluation of waste electronic product trade-in strategies in predictive twin disassembly systems in the era of blockchain. *Sustainability, 12*(13), 5416.
Unger, H., Markert, T., & Müller, E. (2018). Evaluation of use cases of autonomous mobile robots in factory environments. *Procedia Manufacturing, 17*, 254–261.
Wasserbaur, R., Sakao, T., Söderman, M. L., Plepys, A., & Dalhammar, C. (2020). What if everyone becomes a sharer? A quantification of the environmental impact of access-based consumption for household laundry activities. *Resources, Conservation and Recycling, 158*, 104780.
Wilson, M., Paschen, J., & Pitt, L. (2022). The circular economy meets artificial intelligence (AI): Understanding the opportunities of AI for reverse logistics. *Management of Environmental Quality: An International Journal, 33*(1), 9–25.
Yun, Y., & Lee, M. (2019). Smart city 4.0 from the perspective of open innovation. *Journal of Open Innovation: Technology, Market, and Complexity, 5*(4), 92.
Zhuravlova, I., Maciulyte, E., Krauß, J., & Suska, P. (2020). *The outlook on nighttime economy*. Fraunhofer IAO.

Part II

AI and Sustainable Business Operations

5

Strengthening the Sustainability of Artificial Intelligence: Fostering Green Intelligence for a More Ethical Future

Lukasz Swiatek

5.1 Introduction

At a time of increasingly concerning global environmental change, the business sector is being called upon to develop and implement new sustainability initiatives in more innovative ways than ever before. Calls for establishing these new initiatives are growing ever louder because, as Baba (2023, para. 3) notes, "we are finally starting to realize the impact of human and industrial activities" on the Earth and its ability to regenerate and protect itself. Societies' unsustainable trajectories have already caused significant ecological damage and substantial human suffering. The loss of biodiversity, desertification, deforestation, and climate change are just some of the consequences of those trajectories (O'Brien, 2016). By 2007, humanity's "ecological footprint" had already surpassed the Earth's biocapacity by 50% (Moran & Wackernagel, 2012). Consequently, humanity has since found itself in a state of "ecological overshoot," using the earth's

L. Swiatek (✉)
The University of New South Wales (UNSW), Sydney, NSW, Australia
e-mail: L.Swiatek@unsw.edu.au

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_5

resources faster than they can be replenished (Global Footprint Network in Cribb, 2009). Fortunately, businesses are increasingly making sustainability a priority in their operations; in the process, they are finding that "going green" yields unexpected benefits by opening innovative market niches, increasing employee motivation, generating higher stakeholder engagement, and bringing about efficiency gains (Shrivastava & Tamvada, 2019).

Significant and ongoing advances in artificial intelligence (AI) are helping the business sector implement novel sustainability initiatives at an increasingly rapid rate. Businesses of many kinds have already found that multiple AI innovations – such as predictive maintenance, enhanced supply chain tracking, and smart transportation optimization – are significantly strengthening their sustainability efforts (Kompella, 2023). On an even more basic level, businesses can make use of AI for straightforward but important efficiency-related tasks. For instance, AI can help to reduce a business's energy consumption by studying its patterns of energy usage and providing insights into improving consumption in ways that do not compromise productivity (Javaid, 2022). These sorts of initiatives are continually being enhanced because of the fact that AI is also being constantly improved, due to rapid advances in research and development, steadily growing demand for the technology, and enhancements that allow AI to make improvements without human input (Gent, 2020). The global COVID-19 pandemic also accelerated businesses' adoption of AI, as they raced to use the technology in order to overcome new challenges. Krishna et al. (2021, p. 2) note that, even though some executives found themselves suffering from COVID-induced "AI whiplash," there is a sense that businesses "can never return to the pre-COVID-19 pace" of the implementation of the technology.

Although AI brings with it multiple benefits, it also generates issues that currently lack adequate solutions. These issues encompass the mass capture – and, in many instances, the misuse – of data that results in the discriminatory treatment of diverse populations by many different organizations (ranging from government agencies to corporations), the mistakes made by flawed algorithms, the increasing monopolization of the technology by a handful of "tech giants," and the fostering of exploitative labor practices used in the development and running of the technology

(Crawford, 2021). More extreme concerns relate to AI's speeding-up of cybercrime, such as digital hacking, and AI-enabled terrorism (Marr, 2019). AI itself is neither perfect nor objective; it is also a product of the imperfect spheres in which it operates more broadly. As Crawford (2021, p. 211) explains: "They [AI systems] are designed to discriminate, to amplify hierarchies, and to encode narrow classifications."

The environmental damage caused by AI is also a particular concern, especially in light of increasingly worrying global environmental change. AI runs on technologies that require enormous amounts of energy, particularly as the data centers on which it relies excessively consume electricity (Brevini, 2022). As Mosco (2017, p. 150) has noted, these centers are often powered by non-renewable energy sources, while the wireless networks that enable access to the centers are actually "considerably less efficient at energy consumption than are the data centers that have come in for the most criticism." AI also contributes to environmental degradation by requiring the extraction of scarce minerals used in the hardware that supports it, and generating pollution that contributes to anthropogenic – that is, human-caused – climate change (Crawford, 2021). Additionally, the devices on which AI is used day to day (such as laptops and phones) – are problematic, as most of them often end up becoming "e-waste" that harms diverse ecosystems; approximately 50 million tons of this electronic waste are produced each year, and, according to one estimate (Ryder & Houlin, 2019), by 2050, the amount of this waste – that will, by that stage, be fully (or almost-fully) integrated with AI – could swell to 120 million tons each year around the world.

The business sector is coming to terms with the need to make AI more sustainable, though, to date, little attention has been given to the greening of AI in both practice and theory. Companies can take highly varied measures to make their AI use more sustainable, Wiesalla (2022) notes; these measures span the implementation of training for employees to the automation of the capturing of data about a company's AI footprint. A prominent example of a business that has implemented AI for sustainability is Emirates Flight Catering; in partnership with the technology company Winnow, Emirates Flight Catering has been reducing food wastage by using intelligent cameras, meters and smart scales to analyze food preparation processes, identify the most-wasted food items, and calculate the environmental and financial cost of the discarded ingredients (Paul, 2022). This initiative aligns with the national-level work that the United Arab Emirates (UAE) Ministry of Climate Change and Environment (MOCCAE) has been undertaking to reduce food waste using AI; the national conditions in the UAE allow MOCCAE, as an arm of the UAE federal government, to bring together the public and private sector actors effectively (Bridge, 2019). However, the greening of AI has only recently begun to be given sustained treatment in the scholarly literature. Greening this technology is vital in light of the growing climate crisis and the need for companies to achieve corporate social responsibility requirements by fulfilling not just the triple bottom line (Elkington, 1994), but also the quadruple bottom line (Inayatullah, 2005): that is, by helping to look after not just people, profit and planet, but also principles or ethics. To date, green AI – that is, AI that is used in sustainable and environmentally beneficial ways – has begun to be discussed in diverse fields, including computing (Pedrycz, 2022), sustainability (Yigitcanlar et al., 2021), and communication (Masih & Kaur, 2022). However, it has been subjected to a limited amount of academic analysis (see, especially, Bertin et al., 2021; Yigitcanlar, 2021) in relation to the business sector. This chapter helps to fill this current gap in the literature. In so doing, it provides fresh and timely insights relating to both AI and sustainable business practices.

The chapter argues that the greening of bounded AI needs to be undertaken by the business sector in order to enhance businesses' sustainability efforts and address the current issues surrounding AI. Bounded AI refers to AI that undertakes a limited number of activities, especially ones that humans could not undertake by themselves, and that does not harm humans (with an example of this sort of activity being the analysis of multiple, extremely large datasets in order to produce actionable insights within time-critical contexts, such as emergencies or crises). This understanding of bounded AI draws on the work of Rossi and Mattei (2019), particularly their notion of "ethically-bounded AI." Businesses are ideally placed to implement bounded AI and to green it in ethical ways, given their significance and their capacity to provide leadership to other types of organizations. A failure to green bounded AI will likely result in financial, legal, and reputational damage for companies, given the aforementioned issues that AI has already been causing.

In this chapter, a case study of NearMap, an aerial technology company, is used to illustrate this argument. Launched in 2007 in Perth, Australia, NearMap provides organizations of all kinds with high-resolution aerial imagery and, in recent years, has taken steps to green the bounded AI that it has implemented in its operations. The company states that the aerial imagery that it produces “offer[s] unprecedented access to detailed visualizations that tell stories through context” (NearMap, 2023a, para. 9). The case study explains the ways in which NearMap has undertaken the ethical greening of the bounded AI that it uses. This single instrumental case study (Stake, 1995) offers a robust way of understanding the principles that apply from it to other, similar settings (Yin, 1981). The data for the case study were gathered using conventional qualitative content analysis (following Hsieh & Shannon, 2005) that enabled consistent patterns in the company’s publicly available (online) communication collateral, such as webpages and annual reports, to be gathered and analyzed. The goal of this analysis was not to examine the collateral with a view to understanding signification (or the meanings of the different elements in the content); rather, it was intended to identify themes relating to the greening of bounded AI.

The remainder of the chapter is divided into five sections. First, the nature of bounded AI is explained in greater detail. Second, the greening of this sort of AI (as well as non-bounded AI more broadly) is canvassed more comprehensively in showing why the greening of the technology is required. Third, further considerations for the business sector relating to AI and greening are discussed. Fourth, detailed recommendations for business leaders, policymakers, and other stakeholders are presented. Fifth, a conclusion brings together the chapter’s contributions, notes its limitations, and outlines avenues for further research.

5.2 Bounded AI

Various labels have been used to describe the different types of AI that already exist, are believed to be forthcoming, or need to be implemented. Bounded AI is one of those types and is beginning to be used more widely in popular discourse. However, the three types of AI that are currently the main subjects of discussion and analysis are: "narrow AI," "general AI," and "super AI." Frank et al. (2017) usefully explain that "narrow AI" (also known as "weak AI" or "applied AI") is the most widespread and is business-focused (in dealing with particular tasks, such as analyzing data or operating software). The second type, "general AI" (also termed "strong AI"), featuring human-level intelligence, is believed to be decades away from being completely developed. The third type, "super AI," is superior to humans and, it is generally feared, may come to conquer the Earth one day. These three categorizations of AI are captured in Table 5.1. The third type of AI, even though it is only speculative at this point in time, is often the subject of the most significant "AI anxiety" (Johnson & Verdicchio, 2017). Additional labels exist for describing various other categories of AI – such as "embodied AI" (that takes physical form) and "non-embodied AI" (that is virtual) – but those additional types are not outlined in this chapter; for details, see, for example, the essay collection assembled by Lungarella et al. (2007).

Bounded AI, in contrast to other types or categorizations of AI, draws on ethics-related thinking. As outlined in the introduction, this type of AI is associated with the setting of limits around AI. As its name suggests, it is about bounding or circumscribing the activities of AI (both current

Table 5.1 Key categorizations and examples of artificial intelligence

	Narrow AI	General AI	Super AI
Examples	AI software that analyzes data at significant scale and speed	AI (such as a robot) that operates with the sophistication of a human being	AI, in different forms (such as software or robots), that overtakes humans and conquers the Earth

Source: Drawing on Frank et al. (2017)

and potential) as a way of controlling it more effectively and increasing its safety (Bhatt, 2018; Johnson & Verdicchio, 2017). As mentioned previously, this approach is required given the risks and harms that have already arisen due to the inadequate regulation of AI. Rossi and Mattei (2019, p. 9785) observe that the creation of robust AI that can respond effectively to all kinds of living beings and situations, including ones with unexpected developments, involves granting AI a certain level of freedom and creativity. At the same time, these authors note that AI should follow appropriate ethical principles that "should define the boundaries of AI's freedom and creativity." Determining these sorts of ethical principles is a challenge for AI developers and users. As these authors ask: "Are human values suitable for machines, given that machines have extended capabilities compared to humans but lack some very relevant human feelings, such as guilt or empathy, that heavily support human's ethical behavior?" (Rossi & Mattei, 2019, p. 9788). Despite these sorts of questions, some companies have taken practical steps to implement AI in a bounded way. A salient example is the Canadian independent greenhouse produce grower Nature Fresh Farms, which has integrated AI into a delimited set of specific operations: greenhouse growing, irrigation and water reuse, and packaging and distribution (Nature Fresh Farms, n.d.).

One classic component of ethics relates to avoiding harm, as mentioned in the introduction to this chapter; this fundamental component can be viewed as the core element of bounded AI. The principle of non-maleficence (the ethical duty to avoid causing harm to others) has been widely acknowledged as a fundamental principle of ethics in moral philosophy, including in deontological ethics (Kant, 1788), consequentialist ethics (Bentham, 1789), and virtue ethics (Aristotle, 340 B. C. E.). It can thus be viewed as the most fundamental part of bounded AI. In other words, bounded AI should both "remove existing harms and prevent harms that may be likely to occur" (Institute of Medicine, 2014, p. 108). Removing existing harms might involve using AI in ways that avoid further job losses, while preventing potential harms might entail programming AI in such a sophisticated way that it does not injure (at best) or kill (at worst) human beings. This approach to developing and using AI is also in line with the "human-centered AI" framework that puts people first and prioritizes "high levels of human control" over the technology (Shneiderman, 2020, p. 495).

NearMap's integration of AI into its operations very much reflects a bounded approach to AI. Before becoming one of the world's largest aerial survey companies, NearMap operated as a small startup that used airplanes to capture high-quality aerial photos. Now more diversified, its operations continue to focus on providing aerial data, especially imagery. NearMap's AI-based offering is highly delimited and takes the form of an "image and object recognition product which enables you to derive deeper insights of an area or parcel in seconds" (NearMap, 2023b, para. 8). This tightly defined product contains two features. The first, "AI Layers," allows customers to add visual overlays on top of the captured aerial imagery and obtain deeper insights from the AI-generated data. The second, "AI Parcel Export," uses AI to integrate the obtained data into customers' in-house geographic information system platforms, or other workflow applications, in order to enrich customers' existing data (NearMap, 2023b, para. 9). NearMap's bounded approach to AI is reflected not only in the circumscribed or restricted use of the AI technology, but also in the high level of human control that is exercised over it by employees at the company. By setting limits on the use of AI, NearMap also helps to increase the safety of the technology.

5.3 The Greening of Bounded AI

The bounding of AI is only a partial response to the issues that have been raised about the technology; the greening of bounded AI helps to address those issues more fully, especially in terms of sustainability. Greening has been defined in multiple ways by authors from different fields over time. Some definitions are wide-ranging; for example, Wolbring and Noga (2013, p. 88), in their discussion of the key literature relating to sustainability, define greening as "making something more environmentally friendly." Other definitions are more specific. For example, Sharma (2014, p. 3), in his investigation of sustainable innovation efforts by businesses, defines it as "actions adopted by firms to reduce negative impacts on the natural environment." In this respect, greening is pursued by corporations with different levels of intensity. In their now-classic study of corporate greening, Winn and Angell (2000, p. 1129) found that some companies are highly committed to greening and can be considered "environmental innovators," while others are less committed and can be seen to be "environmental followers." In the greening of AI, too, some companies will likely be leaders while others will be laggards.

The greening of AI increasingly needs to be a key part of businesses' sustainability approaches, given the significant environmental impacts of this technology. As mentioned in the introduction, AI – through the use of data centers, networks, and enormous energy consumption (among other elements) – creates an immense carbon footprint. Training a single deep-learning model can emit up to 284,000 kg of carbon dioxide equivalent, the same amount as the energy consumed by five cars over their lifetimes (van Rijmenam, 2023, para. 13). AI also contributes to the creation of vast amounts of e-waste that contains toxic chemicals; it has been estimated that 55 tons of mercury alone from electronics in landfills have already seeped into soil and groundwater globally (Forti, 2020). These sorts of negative environmental impacts are particularly concerning given the acceleration of global environmental change. The current crises facing the planet "driven by decades of relentless and unsustainable consumption and production," Andersen (2021, p. 5) warns, "are amplifying deep inequalities and threatening our collective future."

The greening of AI is also required given the rate of the expansion of the technology and the need to help safeguard the future of the coming generations. The technology is developing at an "extremely rapid pace" (Béranger, 2021, p. xi). Problematically, though, the technology is not grounded in the ecological systems on which life forms on Earth depend (Skene, 2020). Although the greening of bounded AI would not incorporate the technology into these (organic) systems, it would represent a meaningful step in helping to look after the Earth. It would also aid future generations. As Scheffler (2018) has argued, humanity has an obligation to worry about the coming generations and the world that they will inherit, because of the duties of beneficence (in terms of the need to maximize the future wellbeing of forthcoming generations), the desire for the current activities that we value to continue, and a general love of humanity. If companies do not green and bound the AI that they use, the risks of financial, legal and reputational damage will steadily grow into the future. A recent international surge in lawsuits by activists against companies and governments – numbering over 1000 between 2015 and 2021 – is damaging institutions' reputations and forcing them to undertake expensive, speedy measures to implement more effective sustainability measures (Bateman, 2021).

Fortunately for businesses, AI can be greened in a range of ways. Edge computing – enabling AI computations to be performed on devices (such as smartphones) at the edge of networks, rather than on centralized servers – is assisting businesses to make their usage of AI environmentally friendlier (van Rijmenam, 2023). Companies are also tracking energy consumption, increasing transparency around the carbon footprint that AI generates, and training employees to strengthen their awareness of sustainability and the effects that AI has on it (Wiesalla, 2022). For example, Siemens has partnered with Dell to address building performance issues using AI and edge computing. The initiative is helping Siemens not only understand its buildings' energy demand and supply needs more effectively, but also minimize energy use proactively. The greening of bounded AI is aided in the North American context by growing efforts to help companies meet their sustainability goals (Dell, 2022). The environmental impacts of training deep learning models are also being reduced through the use of more efficient algorithms that require less computational power. Some of the major data centers used by businesses are increasingly switching to renewable energy. Microsoft and Google, for example, have committed to using 100% renewable energy to power their data centers by 2030 (van Rijmenam, 2023). Recyclable hardware, already available in limited quantities, is also being improved in an effort to reduce e-waste (Ackerman, 2022).

NearMap has undertaken steps over the years to green its bounded AI and thus strengthen its sustainability efforts. Although most of the company's communication collateral (particularly its set of annual reports) focuses on general sustainability measures, AI-related greening information is provided. NearMap (2021, p. 44) does not try to hide the fact that it "generates vast amounts of data from its ever-expanding location intelligence content. This content requires increasing data storage and processing capacity." To address this issue, the company has contracted

Amazon Web Services (AWS) to store its data, thus reducing the energy consumption it would otherwise use from running its own servers. NearMap also highlights the fact that AWS is making progress towards achieving 100% renewable energy for its infrastructure by 2030, thus also contributing to the greening of the firm's bounded AI. The company aims to create zero waste, but acknowledges that it still "produces a small amount of electronic waste." It has addressed this problem by establishing arrangements with its building management teams for any e-waste items (such as cables, batteries, and monitors) to be donated to social enterprise businesses or not-for-profit organizations if those items cannot be recycled or reused (NearMap, 2022, p. 42).

Other sustainability measures have been helping NearMap to green its bounded AI. The company states that it monitors its environmental impact as part of its *Health, Safety and Environment Policy*, which stipulates that the Board and Executive's decision-making "must consider environmental issues a high priority and the identification of potential environmental issues … [a focus of] ongoing awareness and regular review" (NearMap, 2022, p. 42). This policy's emphasis on sustainability thus provides a high-level environmentally focused motivation for the business's activities, including the ones involving AI. NearMap has also identified that its direct environmental footprint is generated by energy consumption, water usage, and waste. It has engaged an accredited external assessor to verify third party data of its direct greenhouse gas emissions (GHG) in Australia. The company also makes a financial contribution towards the purchase of Australian large-scale renewable generation certificates and Australian carbon credit units that offset GHG emissions. In terms of its aerial data-capturing activities, NearMap acknowledges that it "engages third party aerial operators to fly the Company's proprietary camera systems. These flights emit GHG." However, it claims that it has addressed this issue, saying that "the environmental impact is limited given that a typical aircraft flying these systems is lightweight and carries only the pilot and the camera" (NearMap, 2022, p. 43).

The company has also involved itself in global and local sustainability initiatives that further contribute to the greening of its bounded AI. In 2021, it launched a sustainability program, "NearMap for Good," which is aligned with the United Nations Sustainable Development Goals (SDGs). The program particularly supports goal 11, "Sustainable Cities and Communities," and goal 13, "Climate Action." At the same time, it has committed itself to the "Pledge 1%" initiative, pledging 1% of its product and 1% of its time to supporting the SDGs (Newman in NearMap, 2022, p. 27). In 2022, NearMap partnered with The Nature Conservancy Australia, an organization that works in partnership with Aboriginal and Torres Strait Islanders to care for the land, focusing (among other things) on preventing devastating bushfires through the traditional land management techniques of the First Nations people. By combining local action with sophisticated aerial location intelligence, the two organizations "have committed to working together to monitor and manage important conservation projects" as a way of directly tackling the negative impacts of climate change (NearMap, 2023c, para. 6). In the Australian content, urgency surrounds these initiatives owing to many years of inadequate environmental action by governments at both the state and federal levels (Morton & Readfearn, 2022), highlighting the fact that the conditions supporting the greening of AI around the world differ from jurisdiction to jurisdiction. The varied sustainability-focused activities undertaken by NearMap show that the greening of bounded AI can be undertaken with an internal or external focus, directly or indirectly, by businesses; Table 5.2 captures these different approaches.

Table 5.2 A summary of the different approaches that businesses can take to green bounded artificial intelligence

	Internally focused greening of bounded AI	Externally focused greening of bounded AI
Direct greening of bounded AI	Embedding the greening of bounded AI into the business's policies	Using a sustainability-focused server (cloud) provider
Indirect greening of bounded AI	Making the business's offices carbon-neutral	Partnering with an organization to undertake sustainability activities

Note: Based on the NearMap case study

5.4 Further Considerations for the Business Sector

In greening bounded AI, businesses need to keep an eye on the evolution of the technology and monitor new developments in it, as well as the broad reactions to it. At the time of writing, AI has made recent headlines on a couple of occasions for the dangers it poses if it is left poorly regulated (for details, see, for example, Vallance, 2023; Vincent, 2023). If these sorts of outcries continue, individuals and groups will likely become even more wary of AI, thus potentially becoming unwilling to engage further with AI or AI-integrated products and services. Businesses will need to understand the shifts in sentiment surrounding AI and respond accordingly. Also, if AI continues to evolve at the extraordinarily rapid pace at which it has evolved to date, companies may need to take stronger measures to implement bounding and greening. These measures may come with increased costs, the need to secure additional expertise, and potentially the requirement to source special resources. Businesses should prepare for these scenarios.

As indicated earlier, a significant part of the challenge in this area for businesses involves showing strong and steady leadership around AI, greening, and sustainability more broadly. Stakeholders increasingly expect businesses, with their superior resources, to provide this sort of leadership, despite the turbulent environments now being created by rapidly changing technologies. As Gregory and Willis (2022, p. 39) note, "The emergence of wider environmental, social and governance … concerns requires them [leaders] to expand their gaze even further to consider their relationships with a growing spectrum of stakeholders from across society." Navigating AI, though, requires additional knowledge and skills, ones that most business leaders (and employees) might not have at the moment. For this reason, leaders at all levels of companies may need to undertake additional and ongoing professional learning in order to prepare themselves effectively for all kinds of AI developments. This sort of learning will need to be factored into their schedules (taking them away from other activities).

All members of businesses, not just individuals in leadership positions, should prepare for potentially significant perspective-changes regarding AI and sustainability. For some individuals, these sorts of shifts in perspective may be uncomfortable. AI has already shown that it can operate in unpredictable ways. Sustainability-related conversations are shifting in some parts of the world in light of rapid global environmental change and the worldwide 'polycrisis' to which it is contributing (Tooze, 2022). For example, Barnatt (in Funnell, 2021, pars. 9–12) has precisely commented that typical discussions of sustainability "give[] the impression that we could all go on living exactly as we live today but sustainably." He adds, more bluntly, that, "The only way we can actually preserve things for the future and look after the environment is to change how we live, to use fewer resources, to value things in another way." As these sorts of perspectives evolve, members of businesses will need to prepare for potential changes to thinking and, by extension, practice.

5.5 Recommendations for Business Leaders, Policymakers, and Other Stakeholders

Business leaders should proactively take steps to green bounded AI in order to avoid facing problems that will require reactive, and potentially inadequate, approaches to the bounding and greening of the technology in the future. In particular, business leaders should take steps to accelerate the greening of AI. As the NearMap case study has shown, greening can be undertaken in a variety of ways, directly and indirectly, and with an internal or external focus. Leaders of businesses large and small should think laterally about the different ways in which AI can be greened and bounded throughout their operations. To support this lateral thinking, leaders should consider organizing brainstorming-oriented sessions or retreats with subordinates, establishing rapid-response committees, or hiring expert consultants. Leaders may find themselves isolated or unsupported during the initial efforts to green AI in their businesses. If they do find themselves needing to go out on a limb to implement greening, leaders should remain confident while staying the course and remember the importance of the greening task that they undertake. At the same time, leaders should engage all stakeholders – internal and external – as effectively they can in order to build support for authentic, long-lasting and widely adopted green AI.

Policymakers should also introduce policies that hasten the more widespread greening of AI. In particular, these individuals ought to formulate policies that help businesses transition to the use of steadily greener AI. Although policymaking is not a quick activity, urgency is required on the part of policymakers given the problems that AI is causing. Left unchecked, AI-generated problems could rapidly become even more serious around the world. Given the leadership role that businesses often play, thanks to the resources that they have at their disposal, it is commonsensical for policymakers to aim to support businesses as a matter of priority; businesses, benefiting from that support, can then look to support other social actors (ranging from community groups to non-government organizations). Businesses also need time to transition to green AI, thus making it all the more important for leaders in government to develop policies as expeditiously as they can.

Other stakeholders should increasingly aim to contribute to the greening of bounded AI in their own contexts. The NearMap case study highlights the fact that diverse stakeholders – from (external) suppliers to (internal) employees – can take a myriad of creative approaches to supporting the direct and indirect greening of AI. Above all, the stakeholders ought to familiarize themselves with the purpose of, and the approaches for, greening and bounding AI. This vital first step will provide them with a solid foundation for undertaking this sustainability-enhancing work. As part of the building of this foundation, the stakeholders should consult online sources (ranging from informational videos to trade press articles) in order to increase their knowledge, and join or form groups to develop creative solutions for greening AI. They should also actively identify and use green AI (rather than non-green AI) tools wherever possible, and engage others in the greening of bounded AI by spreading awareness (through informal approaches such as conversations and more formal channels such as posters or flyers).

5.6 Conclusion

AI can undoubtedly help businesses enhance their sustainability efforts, though the issues that this technology entails – and, in particular, the environmental impacts that it generates – need to be recognized and overcome. This chapter has argued that the business sector needs to green bounded AI in order to strengthen sustainability efforts and address current AI-related issues. If businesses, which are ideally placed to provide leadership in this area, fail to take action, they will likely suffer various types of damage (especially reputational, legal and financial damage) in future. Fortunately, as this chapter has explained, the greening and bounding of AI can be undertaken in a range of ways, at both the micro level (through staff training, for example) and the macro level (for instance, through the implementation of company-wide environmentally friendly and delimited AI programs or tools).

The chapter's argument has been illustrated through a case study of the NearMap company and its aerial-imagery-focused operations. This case demonstrates the ways in which a business can not only integrate AI into its activities in a bounded or circumscribed way, but also green it using a variety of different approaches to tackle the environmental impacts of the technology. Businesses, as the case shows, can exercise strong (human) oversight of AI in the specified activities into which it has been integrated. They can also embed bounding and greening into their policies in order to hold themselves accountable to set standards. Additionally, they can engage sustainability-focused third-party providers to help them bound and, in particular, green AI. Partnerships can also be developed with non-government organizations, government institutions, and other businesses to help advance sustainability initiatives and thus further indirectly contribute to the greening and bounding of AI.

This chapter opens a number of avenues for future research. Investigations into other cases would provide a more detailed understanding of the ways in which businesses can green their bounded AI. A selection of cases from different parts of the world would also reveal differences in businesses' perceptions of, and approaches to, greening in corporate contexts across borders. Additional methods could also be used to gain deeper insights into the ways in which the greening of bounded AI can be undertaken. In particular, interviews or focus groups with key business personnel and stakeholders would help to generate these insights. This additional, future research will be valuable given the growing need from businesses for innovative and effective approaches, especially ones that make use of bounded AI, to enhancing sustainability.

References

Ackerman, D. (2022, December 15). Dell's Concept Luna lets a robot recycle your old laptop. *CNET*. https://www.cnet.com/tech/computing/dells-concept-luna-lets-a-robot-recycle-your-old-laptop/

Andersen, I. (2021). UNEP executive director's foreword. In UNEP, *Making peace with nature: A scientific blueprint to tackle the climate, biodiversity and pollution emergencies* (p. 5). United Nations Environment Programme. https://wedocs.unep.org/xmlui/bitstream/handle/20.500.11822/34948/MPN.pdf

Aristotle. (340 B. C. E.). *Nicomachean ethics* (H. Rackham, Trans.). Harvard University Press. http://data.perseus.org/citations/urn:cts:greekLit:tlg0086.tlg010.perseus-eng1:1094a

Baba, S. (2023, March 7). A new UN report offers businesses a template for achieving true sustainability. *The Conversation*. https://theconversation.com/a-new-un-report-offers-businesses-a-template-for-achieving-true-sustainability-197872

Bateman, J. (2021, December 8). Why climate lawsuits are surging. *BBC*. https://www.bbc.com/future/article/20211207-the-legal-battle-against-climate-change

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. Batoche Books. https://socialsciences.mcmaster.ca/econ/ugcm/3ll3/bentham/morals.pdf

Béranger, J. (2021). *Societal responsibility of artificial intelligence: Towards an ethical and eco-responsible AI*. Wiley.

Bertin, E., Crespi, N., & Magedanz, T. (2021). Shaping future 6G networks: Needs, impacts, and technologies. *John Wiley & Sons*. https://doi.org/10.1002/9781119765554

Bhatt, A. (2018). Safe AI systems. *International Journal of Computer Applications, 181*(31), 39–42. https://doi.org/10.5120/ijca2018918205

Brevini, B. (2022). *Is AI good for the planet?* Polity Press.

Bridge, S. (2019, September 19). How the UAE is using AI to reduce food waste. *Arabian business*. https://www.arabianbusiness.com/industries/technology/427077-how-the-uae-is-using-ai-to-reduce-food-waste

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press. https://doi.org/10.12987/9780300252392

Cribb, J. (2009, October). *The coming famine: The risks to global food security* [presentation]. The Productivity and Food Security Symposium, Sydney, Australia. https://www.dpi.nsw.gov.au/__data/assets/pdf_file/0020/304427/Prof-Julian-Cribb%2D%2DThe-coming-famine%2D%2D-speaking-notes.pdf

Dell. (2022, November 8). *Smart buildings leveraging edge solutions*. Dell. https://www.dell.com/en-lv/dt/video-collateral/siemens-desigo-cc-video.htm

Elkington, J. (1994). Towards the sustainable corporation: Win-win-win business strategies for sustainable development. *California Management Review, 36*(2), 90–100. https://doi.org/10.2307/41165746

Forti, V. (2020, July 11). Global electronic waste up 21% in five years, and recycling isn't keeping up. *The conversation*. https://theconversation.com/global-electronic-waste-up-21-in-five-years-and-recycling-isnt-keeping-up-141997

Frank, M., Roehrig, P., & Pring, B. (2017). *What to do when machines do everything: How to get ahead in a world of AI, algorithms, bots, and big data*. Wiley.

Funnell, A. (2021, October 7). Sustainable development won't solve environmental crises, say these experts. It's simpler than that. *ABC News*. https://www.abc.net.au/news/2021-10-07/the-myth-of-sustainable-development/100504448

Gent, E. (2020, April 13). Artificial intelligence is evolving all by itself. *Science*. https://www.science.org/news/2020/04/artificial-intelligence-evolving-all-itself

Gregory, A., & Willis, P. (2022). *Strategic public relations leadership* (2nd ed.). Routledge. https://doi.org/10.4324/9781003185253

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Inayatullah, S. (2005). Spirituality as the fourth bottom line? *Futures, 37*(6), 573–579. https://doi.org/10.1016/j.futures.2004.10.015

Institute of Medicine. (2014). *Health standards for long duration and exploration spaceflight: Ethics principles, responsibilities, and decision framework*. The National Academies Press https://www.ncbi.nlm.nih.gov/books/n/nap18576/pdf/

Javaid, S. (2022, May 31). 5 ways AI can make your business more sustainable in 2023. *AI Multiple*. https://research.aimultiple.com/ai-in-sustainability/

Johnson, D. G., & Verdicchio, M. (2017). AI anxiety. *Journal of the Association for Information Science and Technology, 68*, 2267–2270. https://doi.org/10.1002/asi.23867

Kant, I. (1788). *The critique of practical reason* (T. K. Abbott, Trans.). Project Gutenberg. https://gutenberg.org/ebooks/5683

Kompella, K. (2023, February 10). How industries use AI to ensure sustainability. *TechTarget*. https://www.techtarget.com/searchenterpriseai/tip/How-industries-use-AI-to-ensure-sustainability

Krishna, S., Campana, E., & Chandrasekaran, S. (2021). *Thriving in an AI world*. KPMG. https://advisory.kpmg.us/content/dam/advisory/en/pdfs/2021/thrivingai2021.pdf

Lungarella, M., Iida, F., Bongard, J., & Pfeifer, R. (2007). *50 years of artificial intelligence: Essays dedicated to the 50th anniversary of artificial intelligence*. Springer. https://doi.org/10.1007/978-3-540-77296-5

Marr, B. (2019, August 29). What are the negative impacts of artificial intelligence (AI)?. *Bernard Marr*. https://bernardmarr.com/what-are-the-negative-impacts-of-artificial-intelligence-ai/

Masih, A., & Kaur, G. (2022). Green artificial-intelligence-based communication system. In G. Kaur & A. Srivastava (Eds.), *Green communication technologies for future networks: An energy-efficient perspective* (pp. 215–230). Routledge. https://doi.org/10.1201/9781003264477-12

Moran, D. D., & Wackernagel, M. (2012). Measuring sustainability. In J. Murray, G. Cawthorne, C. Dey, & C. Andrew (Eds.), *Enough for all forever: A handbook for learning about sustainability* (pp. 27–35). Common Ground. https://doi.org/10.18848/978-1-61229-015-7/CGP

Morton, A., & Readfearn, G. (2022). State of the environment: Shocking report shows how Australia's land and wildlife are being destroyed. *The Guardian*. https://www.theguardian.com/environment/2022/jul/19/labor-says-it-wont-put-head-in-the-sand-as-it-releases-shocking-environment-report

Mosco, V. (2017). *Becoming digital: Toward a post-internet society*. Emerald. https://doi.org/10.1108/9781787432956

Nature Fresh Farms. (n.d.). Artificial intelligence and agriculture. *Nature Fresh Farms*. https://www.naturefresh.ca/artificial-intelligence-and-agriculture/

NearMap. (2021). *2021 annual report*. NearMap.

NearMap. (2022). *The most intelligent aerial map on earth: 2022 annual report*. NearMap.

NearMap. (2023a). About us. *NearMap*. https://www.nearmap.com/au/en/aerial-view-maps-about-us

NearMap. (2023b). NearMap AI. *NearMap*. https://www.nearmap.com/au/en/products/ai-aerial-maps

NearMap. (2023c). NearMap and the nature conservancy Australia, a greener future. *NearMap*. https://www.nearmap.com/au/en/aerial-view-blog/nearmap-partners-with-the-nature-conservancy-australia

O'Brien, C. (2016). *Education for sustainable happiness and well-being*. Routledge. https://doi.org/10.4324/9781315630946

Paul, S. (2022, January 8). The food industry has a waste problem. Can AI solve it?. *Datatechvibe*. https://datatechvibe.com/data/the-food-industry-has-a-waste-problem-can-ai-solve-it/

Pedrycz, W. (2022). Towards green machine learning: Challenges, opportunities, and developments. *Journal of Smart Environments and Green Computing, 2*(4), 163–174. https://doi.org/10.20517/jsegc.2022.16

Rossi, F., & Mattei, N. (2019). Building ethically bounded AI. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 9785–9789. https://doi.org/10.1609/aaai.v33i01.33019785

Ryder, G., & Houlin, Z. (2019, January 24). The world's e-waste is a huge problem. It's also a golden opportunity. *World Economic Forum*. https://www.weforum.org/agenda/2019/01/how-a-circular-approach-can-turn-e-waste-into-a-golden-opportunity/

Scheffler, S. (2018). *Why worry about future generations?*. Oxford University Press. https://doi.org/10.1093/oso/9780198798989.001.0001

Sharma, S. (2014). *Competing for a sustainable world: Building capacity for sustainable innovation*. Routledge. https://doi.org/10.4324/9781351285803

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Shrivastava, M., & Tamvada, J. P. (2019). Which green matters for whom? Greening and firm performance across age and size distribution of firms. *Small Business Economics, 52*, 951–968. https://doi.org/10.1007/s11187-017-9942-y

Skene, K. R. (2020). *Artificial intelligence and the environmental crisis: Can technology really save the world?* Routledge. https://doi.org/10.1201/9780429055676

Stake, R. E. (1995). *The art of case study research*. Sage.

Tooze, A. (2022, October 28). Welcome to the world of the polycrisis. *Financial Times*. https://www.ft.com/content/498398e7-11b1-494b-9cd3-6d669dc3de33

Vallance, C. (2023, March 30). Elon Musk among experts urging a halt to AI training. *BBC News*. https://www.bbc.com/news/technology-65110030

van Rijmenam, M. (2023, February 23). Building a greener future: The importance of sustainable AI. *The Digital Speaker*. https://www.thedigitalspeaker.com/greener-future-importance-sustainable-ai/

Vincent, M. (2023). Tech world warns risk of extinction from AI should be a global priority like pandemics and nuclear war. *ABC News*. https://www.abc.net.au/news/2023-05-31/tech-world-warns-risk-of-extinction-from-ai-should-be-priority/102413250

Wiesalla, L. (2022, August 4). Green AI – Corporate Sustainability. *NextLytics*. https://www.nextlytics.com/blog/green-ai-corporate-sustainability

Winn, M. L., & Angell, L. C. (2000). Towards a process model of corporate greening. *Organization Studies, 21*(6), 1119–1147. https://doi.org/10.1177/0170840600216005

Wolbring, G., & Noga, J. (2013). Greening and energy issues: An analysis of four Canadian newspapers. *Journal of Sustainable Development, 6*(7), 88–112. https://doi.org/10.5539/jsd.v6n7p88

Yigitcanlar, T. (2021). Greening the artificial intelligence for a sustainable planet: An editorial commentary. *Sustainability, 13*(24), 1–9. https://doi.org/10.3390/su132413508

Yigitcanlar, T., Mehmood, R., & Corchado, J. M. (2021). Green artificial intelligence: Towards an efficient, sustainable and equitable technology for smart cities and futures. *Sustainability, 13*(16), 1–9. https://doi.org/10.3390/su13168952

Yin, R. K. (1981). The case study as a serious research strategy. *Science Communication, 3*(1), 97–114. https://doi.org/10.1177/107554708100300106

6

Predictive Machine Learning in Assessing Materiality: The Global Reporting Initiative Standard and Beyond

Jan Svanberg, Peter Öhman, Isak Samsten, Presha Neidermeyer, Tarek Rana, and Natalia Berg

6.1 Introduction

Commitment to sustainability is an increasingly important factor in a stakeholder's decision to support a company. Sustainability reporting standards, such as the Global Reporting Initiative (GRI) and the Sustainability Accounting Standards Board (SASB), guide companies in

J. Svanberg (✉)
University of Gävle, Gävle, Sweden
e-mail: jan.svanberg@hig.se

P. Öhman
Mid Sweden University, Östersund, Sweden
e-mail: Peter.Ohman@miun.se

I. Samsten
Stockholm University, Stockholm, Sweden

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_6

assessing the materiality of their goods; this may be defined as the material composition of an object. However, sustainability standards often adopt a procedural approach, (i.e., a methodological and systematic way of implementing standards) whereby companies have relatively high discretion in selecting the information to be reported. The many loose ends in materiality assessment approaches of sustainability standards are problematic. In particular, companies are required to report only those properties that they determine to be material. As a result, the indeterminacy of materiality assessments leads to indeterminant reporting, leading to companies selectively reporting or withholding. In this chapter, we address companies' disclosure of sustainability information, especially as it relates to the environment, e.g., carbon emissions, use of water, energy, and pollution. We consider areas where companies disclose information about issues relating to these features, "sustainability topics" or "environmental topics," and use these terms interchangeably. This chapter assesses the materiality of various sustainability topics considered for inclusion in sustainability reports and is structured as follows. We begin with an overview of existing reporting standards and types. The following section introduces our methodology and situates our proposed contribution within related literature before presenting our model, applying it to three different examples, and discussing our results.

Important developments, particularly the new mandatory reporting requirements and the extended and mandatory sustainability reporting

P. Neidermeyer
West Virginia University, Morgantown, WV, USA
e-mail: presha.neidermeyer@mail.wvu.edu

T. Rana
Royal Melbourne Institute of Technology University, Melbourne, VIC, Australia
e-mail: tarek.rana@rmit.edu.au

N. Berg
Linnaeus University, Växjö, Sweden
e-mail: natalia.berg@lnu.se

standards launched in the European Union (EU) on 1 January 2023 demarcate a clear path towards more definite and standardized sustainability reporting (Jørgensen et al., 2022). Corresponding developments, albeit in different forms, are occurring in the United States (US), Canada, and Australia, where the boundaries of what companies need to address with sustainability reporting are being tested. The importance countries attribute to sustainability reporting is indicated by the number of new regulations they implement, such as the EU's *Corporate Sustainability Reporting Directive* (CSRD), which substantially expands the scope and content of the EU's existing disclosure requirements. The new law is in force for many EU corporations for fiscal years starting on or after 1 January 2023. It refers to the European Financial Reporting Advisory Group's (EFRAG) proposed *European Sustainability Reporting Standards* (ESRS). The European Commission aims for a paradigm shift in sustainability reporting to increase information transparency and corporate accountability (Jørgensen et al., 2022). At the same time, the US Securities and Exchange Commission (SEC) is considering rule amendments to enhance disclosures regarding issuers' climate-related risks and opportunities.

US and EU corporations cannot appropriately compare their sustainability performance to their transatlantic peers. This is an issue faced by companies in the United Kingdom (UK), Canada, Australia, and the rest of the world, as globally listed companies can still choose whether they report with the GRI double- or impact materiality principle or with a financial materiality viewpoint of SASB. These types of standards are a matter of choice because nones of them is mandatory. Comparing sustainable performance between companies is not a small issue because investors pour enormous capital into sustainable investment. In 2018, one in every $4 ($12 trillion USD) of investment in the US was labelled sustainable and one in every two euros (€22 trillion) in Europe carried the same label, and the numbers are growing steadily (Christensen & Clawson, 2018; USSIF, 2018).

Two types of materiality approaches guide what sustainability topics companies should disclose. Companies in the US, Canada, Australia, and the UK have the option to report sustainability topics that are financially material to the corporation and its owners (financial materiality). The new EU legislation now mandates EU companies to disclose their impacts on society and the environment regardless of financial implications (impact materiality) and a financial materiality requirement. Does materiality assessment guarantee that the right concerns are reported? A study by Christensen et al. (2021) shows no evidence that the financial materiality of individual sustainability topics is determinable. We are unaware of any study describing a method to determine impact materiality. We introduce a machine learning method that determines materiality based on the likelihood of news media critique. We suggest that instances of news media critique are used as labels of non-compliance with environmental norms.

This chapter discusses a decision-support tool, such as a simple model of a machine learning (ML) method, to assess the impact materiality of topics in sustainability reports that provide information in accordance with the new EU legislation and the GRI standards. Although the materiality assessment can be applied to topics across environmental, social, and governance (ESG) dimensions, we control the task by using the ML method to assess the materiality of topics in the environmental dimension of sustainability. The findings will be of interest to EU companies and other large companies that report according to GRI. In 2020, 73% of the world's 250 largest companies reported sustainability performance in accordance with GRI, according to a KPMG survey.[1] The findings contribute to research devoted to quantitative methods for assessing materiality in sustainability reports by applying an ML method to expose the subjectivity issues in materiality assessment.

Scholars suggest using qualitative methods subject to bias and subjectivity and leaving the objectivity issues unsolved. No scientifically verifiable method provides objective results to companies making assessments of impact materiality. The materiality assessment approach that forms the basis for selecting sustainability topics in sustainability reports is incomplete because these reports do not describe the materiality assessment process (Canning et al., 2019). This makes it challenging to apply qualitative methods.

[1] https://assets.kpmg/content/dam/kpmg/be/pdf/2020/12/The_Time_Has_Come_KPMG_Survey_of_Sustainability_Reporting_2020.pdf

Further, the GRI standards and the new EU sustainability reporting standards require that materiality be assessed from the perspective of the broadest possible stakeholder groups, which is society as a whole. Divergent stakeholder opinions should be mediated through the materiality assessment. However, it is unlikely that reporting companies can do this because stakeholder opinions are largely unknown to the companies (Bellantuono et al., 2016). Although there are descriptions of methods to collect opinions from many stakeholders, we are unaware of any method that satisfies the standards' requirements for representing society's opinions on many sustainability topics.

Despite many practical obstacles, a materiality assessment is a prerequisite for sustainability reporting, and this chapter shows the potential usefulness of machine learning methods. We focus on the impact of materiality requirements in the GRI standards and the EU *Corporate Sustainability Reporting Directive* (CSRD). Predictive algorithms can help report companies to identify the most material sustainability topics companies focus on.

6.2 Literature Review

Materiality assessments are intended to be the foundation of sustainability reports because they determine what sustainability topics, such as pollution or waste, the company should disclose concerning its production (Eilifsen & Messier, 2015). Traditional materiality assessment methodologies that evolved in the context of financial reporting might be inappropriate for sustainability reporting because they provide no guidance about how to compare the financial or environmental impact implications of, for instance, a company's CO_2 emissions to the financial or environmental impact implications of a company's total waste (Knechel, 2021). Researchers have argued that materiality assessment in sustainability reporting is a translation of the principle of materiality in financial reporting (Eilifsen & Messier, 2015; Messier et al., 2005; Moroney & Trotman, 2016). In financial reporting, materiality involves "determining the importance of the disclosure of an item of information (or its omission) to users" (Canning et al., 2019, p. 2). It is the threshold at which potential disclosures matter for an external user's decision. In contrast, the (impact) materiality assessment in sustainability reporting assesses whether the relative significance and type of impacts make sustainability topics important enough to stakeholders to be reported. Comparing sustainability topics is complicated and multidimensional because it requires comparing the size and nature of very different effects of a company's activities on stakeholders (Svanberg et al., 2022).

The differences between traditional materiality in financial reporting and materiality assessment in sustainability reporting illustrate how complex the latter is. First, the impact materiality approach requires that companies critically assess stakeholders who are individuals or groups with interests that are or could be, affected by a company's activities. This may include, for example, business partners, civil society organizations, consumers and customers, suppliers, and local communities. GRI and CSRD require broad stakeholder groups to be identified and used as references to prioritize the importance of sustainability topics. Although impact materiality may often have financial implications for the companies that cause the harm or benefits to the environment, there is no guarantee that there are financial implications. This is the main reason why the EU has put new reporting requirements in place. Financial reporting and the financial materiality approach to sustainability reporting require that investors' information interests are prioritized. This means reporting according to financial materiality should contain information relevant to investors. In some cases, the two materiality principles lead to quite conflicting reporting: SASB defines financial material information as could reasonably be expected to influence investment or lending decisions that users make on the basis of their assessments of short-, medium-, and long-term financial performance and enterprise value if misstated or omitted. According to an analysis by Jørgensen et al. (2022), in practice, this means that the environmental footprints of H&M and NIKE do not have to be disclosed according to the financial materiality standards, while they are likely to be disclosed according to the impact materiality standards and be important for many stakeholders.

Assessing impact materiality is challenging for several reasons. First, broader stakeholder groups are difficult to identify (Puroila & Mäkelä, 2019), and their conflicting needs are difficult to define (Moroney &

Trotman, 2016). For example, if a reporting company views women as a key stakeholder group, it would contain many sub-groups with potentially conflicting views on environmental concerns. There may be mothers with a strong interest in the welfare of their children and a strong emphasis on reducing certain pollutions and waste, as well as environmentalists who consider CO_2 the most troubling concern of today. How should a reporting company determine the information preferences of such a varied stakeholder group? Second, a precise materiality threshold, such as a benchmark percentage, can be used in financial reporting. However, in sustainability reporting, there is no such benchmark for materiality. How should a company present the trade-off between a 5% increase in carbon emissions and a 10% decrease in waste? Standards require a company to do just that.

Sustainability reporting standards state that the reporting company should decide where to set the materiality threshold (for each sustainability topic individually) by assessing the significance and likelihood that impacts are relevant to stakeholders. The companies' difficulties in assessing the trade-off between environmental concerns are reflected in differences between auditors' perceptions of what they believe to be relevant sustainability information—auditors do not know either (Libby & Brown, 2013). Third, materiality assessment in financial reporting is a quantitative exercise. The materiality assessment in sustainability reporting demands both quantitative and qualitative predictions (Canning et al., 2019; Moroney & Trotman, 2016). Companies must compare apples and oranges using a metric they cannot access. These differences between traditional financial reporting and sustainability reporting indicate that assessing the materiality of topics in the sustainability report takes place with significant uncertainty for the reporting companies. There is no way of knowing whether a ton of carbon emissions is as essential as a ton of waste. Which one is more important to the stakeholders?

Concerning the inherent uncertainties in the materiality assessment observed by Moroney and Trotman (2016) and by Canning et al. (2019), scholars have questioned the credibility of sustainability reports (Guix et al., 2019; Machado et al., 2021). The effect of poor materiality assessment approaches can mean that companies are selective in their sustainability reporting in a way that accentuates strengths and withholds weaknesses in order to mislead stakeholders (Adams, 2004). In practice, the involvement of stakeholders in sustainability reporting as the approach of determining which topics are the most material is an unreasonable requirement for many companies. This explains why the materiality assessment becomes a discretionary reporting practice rather than the means of reporting what matters (Puroila & Mäkelä, 2019).

Because many companies report sustainability topics discretionarily, sustainability report users struggle to question the topic selections of the reporting company. The users of reports, such as investors and the general public, lack an intersubjectively binding reference point for materiality assessment to which they could refer. The solution to this problem must be reducing the amount of subjective discretion that companies can exercise when assessing materiality. Studies have proposed subjectivity reduction methods. Bellantuono et al. (2016) describe a mathematical method to aggregate the opinions of various stakeholders on several sustainability topics to represent the combined views of many stakeholder groups. A theoretical problem of this method is that it is impossible to determine the relative importance of a spectrum of sustainability topics without asking each stakeholder to take a position on all topics. This is unrealistic as it assumes that stakeholders know the many environmental damages companies can expose to their surroundings (Puroila & Mäkelä, 2019). Furthermore, there are no practical tools that assist in identifying relevant stakeholders (Boesso & Kumar, 2009; Torelli et al., 2019). The lack of completeness of sustainability reports in covering all the material aspects from a stakeholder perspective illustrates this (Adams, 2004; De Villiers & van Staden, 2011; O'Dwyer et al., 2005). Findings that companies conduct their materiality assessment in different ways accentuate this lack of completeness because sustainability reporting standards enable companies to determine material topics based on the context in which the company operates (Farooq & de Villiers, 2019; Farooq et al., 2021; Jones et al., 2016). Reporting companies can decide which stakeholders to engage and the importance assigned to different stakeholders. As such, materiality is a discretionary issue (Adams & Frost, 2006; Kaur & Lodhia, 2018).

Relevant literature also proposes procedural suggestions for a method for materiality assessment. Using Hsu et al.'s (2013) approach, Calabrese et al. (2016) propose a formalized procedure using multicriteria decision-making and fuzzy logic to provide reporting companies with some conceptual support on materiality assessment. However, the issue of arranging the topics from most to least material topics remains.

Based on the literature review, we find no method in existing scholarship that offers a materiality ranking of sustainability topics which can be verified in terms of validity and accuracy. This chapter develops a materiality assessment methodology consistent with sustainability reporting standards. In particular, the chapter focuses on the impact materiality approach of GRI standards. Prior research argues that financial materiality may not be possible to measure because there is limited empirical evidence of the relationships between items claimed to be financially material and companies' financial performance (Christensen et al., 2021).

The main concern of impact materiality assessments is that the effect of a company on society, such as environmental harms from waste, emissions, or harm to the workforce from disrespect of labor law, is the foundational point for what counts as material. Impact materiality targets the effects on society and the environment regardless of whether there are any financial repercussions for the company in question. The coming EU standards for sustainability reporting state that materiality assessment should consider all stakeholders in their capacity either as information users or as affected by a company's activities, which effectively includes the entire society (EFRAG, 2022). As previously mentioned, impact materiality refers to the significance of an effect that a company has or could have on stakeholders, the environment, and society at large. For example, the environmental activities of companies generally affect different stakeholders and interest groups. Environmental problems such as climate change, pollution, and biodiversity erosion have significant and long-lasting consequences for both people and ecosystems. Impact materiality should therefore mirror the opinions of broad groups of people. Accordingly, we propose to measure a societal-level perception of the materiality of a wide spectrum of environmental topics using a method that estimates the likelihood that a company would be criticized for non-compliance with environmental norms. The logic is that the more an environmental topic is criticized for non-compliance with environmental norms, the more severe and morally intense the topic is considered to be

(Jones, 1991). We use the public scrutiny of companies as a proxy for society's assessments of companies' non-compliance with environmental norms. News media critique is often a factor in ESG ratings (a measure of sustainability often used in the finance industry) and corporate wrongdoing indices (Faulkner, 2011; Fiaschi et al., 2020). We suggest that instances of news media critique, which we refer to as controversies, are used as labels of non-compliance with environmental norms from which machine learning models can extract the environmental topics characteristic of the non-compliant companies. Some studies use news media critiques (i.e., "bad news") to determine an unexpected risk incident (i.e., controversies) (Nieri & Giuliani, 2018) when a company does not comply with social norms; for example, the United Nations' (UN) Global Compact principles represent the major source of a legitimacy gap (Hassel & Semenova, 2018). The methodology and results section below explains how we use compliance labels in machine learning experiments to develop a model to assess companies' environmental topics. We then describe how the materiality of environmental topics can be estimated from the model and discuss interpretations of the model output.

This following section discusses a decision-support tool, such as a simple model of a machine learning (ML) method, to assess the impact materiality of topics in sustainability reports that provide information in accordance with the new EU legislation and the GRI standards. Although the materiality assessment can be applied to topics across environmental, social, and governance (ESG) dimensions, we control the task by using the ML method to assess the materiality of topics in the environmental dimension of sustainability. The findings will be of interest to EU companies and other large companies that report according to GRI. In 2020, 73% of the world's 250 largest companies reported sustainability performance in accordance with GRI, according to a KPMG (2020) survey. The findings contribute to research devoted to quantitative methods for assessing materiality in sustainability reports by applying an ML method to expose the subjectivity issues in materiality assessment.

The usefulness of an automated assessment tool should be judged on how well it performs on the assessment compared to alternative methods. Alternative methods offer mainly procedural support or conceptual focus, such as Calabrese et al. (2013, 2016) or Wu et al. (2018). The latter simply propose that two static, pre-defined materiality assessments are Sustainability Accounting Standards Board Materiality Map and Global Reporting Initiative (GRI) Sustainability Topics for Sectors or GRI's Sustainability Disclosure Database and that modelling from a life-cycle perspective could be ways to assess materiality. The only realistic alternative appears to be how the assessments are conducted manually. Descriptions are scarce, and companies do not report extensively about their method of materiality assessments (Farooq et al., 2021; Machado et al., 2021). Unfortunately, we find no evidence of the validity and accuracy of any presently practiced assessment method. Some evidence even indicates that companies disclose less and less when they adopt the more standardized guidance that some standards have implemented (Borgert et al., 2018). Our literature review shows that accounting for materiality is considered a serious problem, for which offered solutions are procedural recipes that yield little evidence of results.

6.3 Methodology and Results

6.3.1 Methodology

Traditionally, materiality assessments are done manually, which involves collecting and analyzing data from sources such as stakeholder feedback, industry reports, and internal documents. Our machine learning method, provides advantages over manual procedures for several reasons. First, it gives consistent assessments by applying the same set of rules and criteria to all data inputs, reducing the risk of subjective biases. Our method uses verifiable data, while a manual assessment may be based on information the people involved are unaware of, such as the preferences of broad stakeholder groups. Second, machine learning models can handle large datasets much more efficiently than manual assessments. This scalability can help identify previously unnoticed trends or patterns in relation to a company's disclosed environmental topics. Third, machine learning models can analyze data much faster than manual processing, enabling stakeholders to identify sustainability risks and opportunities in a timely manner. Fourth, machine learning models can provide more accurate assessments using advanced algorithms that can detect patterns and correlations in large datasets that are not visible with manual procedures. Fifth, our machine learning model can provide more transparency in the materiality assessment through the simplicity of the underlying logic. Topics are assessed as material if they substantially contribute to a corporation's prediction as non-compliant with environmental norms.

Based on our machine learning experiments, we explore the possibility of predicting non-compliance characteristics with environmental norms. If this can be done from the environmental information provided by sustainability reports, then variable importance measures such as Shapley Additive exPlanations (SHAP) can be used for determining the significance (materiality) of each environmental item for the prediction (Lundberg & Lee, 2017). Companies discretionarily disclose, i.e., they decide what to disclose, the environmental information in their sustainability reports, and our goal is to estimate how each item contributes to society's condemnation of companies' environmental failure to comply.

An important concern is what the prediction model's output means in relation to materiality assessment in sustainability reporting. Items that contribute the most to the likelihood that a company is compliant or non-compliant with environmental norms are rated by the model as the most material. In contrast, other items follow in descending order of materiality. In other words, the model output estimates environmental topics in terms of the likelihood that a company could be criticized for non-compliance with environmental norms. We view the likelihood of media criticism of companies' environmental activities as a measure of the significance of environmental topics for broad stakeholder groups. This means that if there is much media criticism targeting a company and its environmental impacts, we consider the amount of criticism as an indication that broad stakeholder groups consider the disclosed environmentally relevant behaviors of the company as significant/material.

The topic estimate is the outcome of the machine learning model's assessment of patterns of environmental items, such as pollution and energy consumption, and the likelihood of non-compliance with environmental norms. This estimation is non-subjective. The model arrives at the variable importance estimates for each environmental disclosure item.

For simplicity, we demonstrate our materiality assessment method by a random forest algorithm. Random forest is a tree-based algorithm for which the SHAP variable importance measures can be used to calculate the materiality of environmental indicators (topics) (Lundberg & Lee, 2017). The predictive ability of the model is evaluated according to five conventional performance measures (cf. Alpaydin, 2010): precision, recall, F-measure, area under receiver operating characteristic (ROC) curve, and Precision recall curve (PRC).

We acquired all data from Refinitiv Eikon, one of the most commonly used databases for sustainability research. For the prediction model not to exclude potentially material items ex-ante, we use as many as 112 environmental performance indicators as model input (Svanberg et al., 2022). In addition to those, and to aid in our assessment of how the model predicts environmental controversies, we included total assets, net assets, return on assets, market value, industry classification (i.e., Global Industry Classification Standard (GICS) sector), and country of headquarters. The dataset contains 2517 companies, of which approximately 13% have been involved in at least one environmental controversy. Table 6.1 provides an overview of the companies in the sample. Media reporting typically includes a range of political and social views which is one of the reasons why using such data is a good reference point for representing broad stakeholder groups' views on for example environmental issues.

We first examine the ability of random forest to predict environmental controversies according to the five conventional performance measures. The algorithm's output after training predicts the likelihood that a company complies with environmental norms. A post hoc interpretation method, SHAP, then interprets the predictions (Lundberg & Lee, 2017), which, for example, provides diagrams showing how much each environmental issue contributes to the model's assessment of environmental performance of a particular company. Notably, prediction in this study does not mean that the model is forward-looking because its purpose is to characterize companies as more or less environmentally compliant based on typical patterns in their environmental disclosures. The model shows each company's characterization or assessment, rather than predicting a future state based on past performance.

Table 6.1 Descriptive statistics by industry, number of controversies, and total assets

GICS sector	Companies, n	Controversies, n	Companies with controversies, n	Controversies, when $n > 0$ Min.	Max.	Mean	Standard deviation	Skewness	Kurtosis	Assets in USD billions	Assets in USD billions per company Min.	Max.	Mean	Standard deviation	Skewness	Kurtosis
Energy	183	130	59	1	7	2.2	1.69	1.47	1.19	4581	0.07	291	25	51	3	12
Materials	277	96	48	1	7	2	1.43	1.83	3.2	2703	0.04	96	9	15	3	14
Utilities	124	59	38	1	4	1.55	0.89	1.53	1.37	3494	0.13	259	28	36	3	15
Consumer discretionary	297	36	21	1	4	1.71	0.9	1.09	0.4	4153	0.11	331	13	35	6	46
Industrials	420	19	18	1	2	1.06	0.24	4.24	18	5770	0.25	457	13	35	10	122
Consumer staples	156	16	12	1	2	1.33	0.49	0.81	−1.65	2377	0.29	160	15	23	3	15
Health care	141	7	6	1	2	1.17	0.41	2.45	6	2257	0.05	144	16	24	2	7
Financials	398	5	5	1	1	1	0	0	0	79,534	0.24	2509	199	393	3	11
Information technology	201	2	2	1	1	1	0	0	0	2387	0.16	182	11	24	4	24
Communication services	161	1	1	1	1	1	0	0	0	3415	0.06	284	21	40	3	14
Real estate	155	1	1	1	1	1	0	0	0	1667	0.42	78	10	12	2	7

We divide companies into two categories—positive cases and negative cases—one consisting of companies that have had at least one controversy (positive cases) and the other that have not been involved in a controversy (negative cases). The longitudinal aspect of data is captured by an average of numerical (e.g., CO_2 emission) or encoding using dummy variables if binary (e.g., one per year). We compute SHAP values, and the SHAP diagrams used are illustrations through which the materiality distribution is easily comprehended.

To evaluate the random forest algorithm, we first calculate precision and recall. Precision is the fraction of true positives in relation to the total number of positive case predictions. At the same time, recall is the fraction of true positive predictions in relation to the true positive and false negative cases in the data. Precision can be said to represent the number of times a darts player's arrow hits the target about the number of attempts, and recall represents the ability of the predictor to identify as large as possible fraction of all the controversy companies in a data set. In addition, we use the F-measure, which is the trade-off between precision and recall.

Furthermore, the measure, here meaning number or metric, Area under the ROC curve estimates the probability of a classifier ranking a true positive instance ahead of a false positive instance and is thus a measure of its so-called ranking performance. Similarly, the PRC estimates the mean precision for multiple thresholds of recall. The main benefit of the Area under the ROC curve and PRC is that they are insensitive to the class distribution of the training and testing data, as opposed to the static accuracy measures.

6.3.2 Results

Table 6.2 summarizes our findings regarding the machine learning model's ability to predict environmental non-compliance. The ability of the

Table 6.2 Measures of random forest predictive performance

	Precision	Recall	F-measure	ROC	PRC
Random forest	0,6994	0,1942	0,2990	0,8849	0,5090

random forest model to predict non-compliance is not our main focus, and these results have been explained elsewhere (Svanberg et al., 2022). We therefore only briefly summarize these facts here. Our materiality assessment method requires predictive validity of its underlying model, i.e., a model that can predict our operationalization of environmental topics for individual companies. As shown, the random forest does well on the task of predicting environmental controversies with a high precision of 0.70 but a more moderate recall of 0.19. Broadly speaking, this means that the model's current settings make it good at correctly identifying a sample of non-compliant companies.

Figures 6.1a, b enable evaluation of the ranking and predictive performance of the model, and interpretations are briefly explained under each figure. The graphs reveal that the results in Table 6.2 and Fig. 6.1 are consistent, i.e., random forest learns the item patterns typical of controversy companies and performs well as a controversy company predictor. The performance according to the Area under the ROC in Fig. 6.1a is suitable for both non-controversies (blue curve) and controversies (yellow curve). Figure 6.1c presents negative correlations between likelihood estimations and the number of controversies. The x-axis is the number of controversies per company, and the y-axis is the estimated likelihood based on environmental information items. The slope in the graph is downwards, as expected because the more controversies a company has had, the higher the estimated likelihood of controversy.

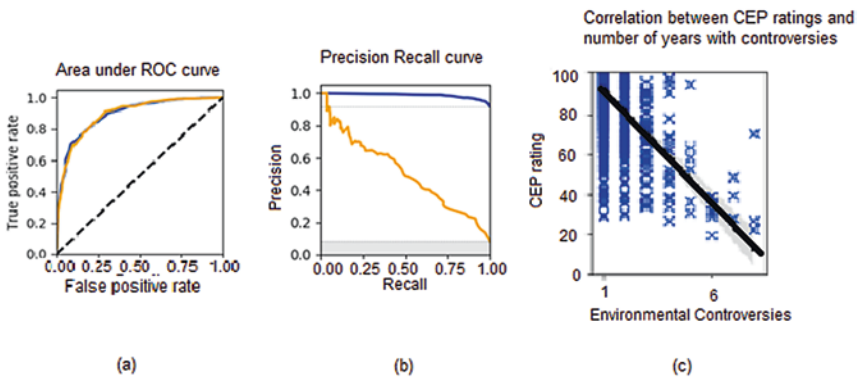

Fig. 6.1 Evaluation of predictive performance

Figure 6.1 presents three pieces of information to be evaluated:

(a) **Area under the ROC curve**, where the blue line represents the results for predicting non-controversy and the yellow line represents the results for predicting controversy. Note that classifiers that produce ROC curves that lie above the dashed line provide better predictions than random guessing.
(b) **Precision recall curve (PRC),** where the blue lines represent the results for predicting non-controversy and the yellow line represents the results for predicting controversy. The PRC shows the precision of a classifier as the recall increases. The top region (defined by the grey line) shows the region where a classifier performs better than random guessing for the non-controversy cases, and the region between the bottom and top regions shows where a classifier performs better for the controversial cases.
(c) **Correlation between the random forest corporate environmental performance (CEP)** ratings and the number of years a company has been controversial. The figure clearly shows that our CEP ratings penalize firms with more controversies and assign lower scores, i.e., the lower the CEP rating, the more controversies.

Once the predictability of non-compliance is established, variable importance measures can represent the materiality of environmental topics for each company. Predictive modelling offers this advantage compared to explanatory modelling in that it targets not general relationships in a population but the particular properties of each case. For materiality assessments, the prediction model offers rich information on what topics matter the most for assessing a corporation as non-compliant with environmental norms. In the examples of materiality assessments below, we look at materiality with SHAP. Our estimates of materiality in terms of the SHAP values are numerical between 0 and 1. Notably, any aggregation level within an industry classification, such as GICS, can generate SHAP. The SHAP is most accurate for the individual company. Still, if the reporting company would like to compare its materiality assessment with a group of companies, our method allows the computation of materiality for any group of companies.

Regarding the interpretation of individual materiality assessments, the so-called SHAP Force Plots offer an overview of environmental feature materiality. Such plots show materiality expressed as the contribution in terms of the conditional probability of each environmental topic indicator to the model's assessment of compliance with environmental norms. We illustrate this materiality assessment by providing three examples from two GICS sectors. These illustrations are not meant to provide evidence of this method's validity or reliability but to show what kind of information can be derived in seconds by looking at companies' potential material topics through the machine learning lens. Therefore, we do not describe these companies in detail. The first company that the model assesses here is Extendicare Inc., a publicly traded Canadian company in the healthcare sector that provides long-term care and senior care services. It was founded in 1968 and is headquartered in Markham, Ontario. The company has a network of over 120 senior care and retirement living centers in Canada, home healthcare services and other health-related businesses. Extendicare operates in two main sectors: long-term care and retirement living. The company describes itself as strongly committed to innovation and improving care delivery through technology and evidence-based practices. One of the big rating firms, Morgan Stanley, recognized Extendicare Inc. for its commitment to ESG issues through an "A" rating in 2021, which indicates that Morgan Stanley claims that Extendicare is a sustainable company in accordance with MCI ESG Ratings. Extendicare was included in the FTSE4Good Index, another rating organization claiming that Extendicare is more sustainable than most other companies.

Figure 6.2 offers an immediate overview of environmental topics' importance in Extendicare's performance; the blue arrows dominate, suggesting that Extendicare is a low-impact/low-risk company in the environmental impact dimension. The base value, where an average corporation's environmental risk/impact metric would be, is far to the right, meaning that Extendicare is doing much better than the average. This is because it has not reported any environmental fines in past years, it appears to be uninvolved in activities that have a negative environmental impact on land (signaled by not having a policy for reducing such impact), and its CO_2 emissions are on the low side. These three material aspects push the company towards the positive, but there are also

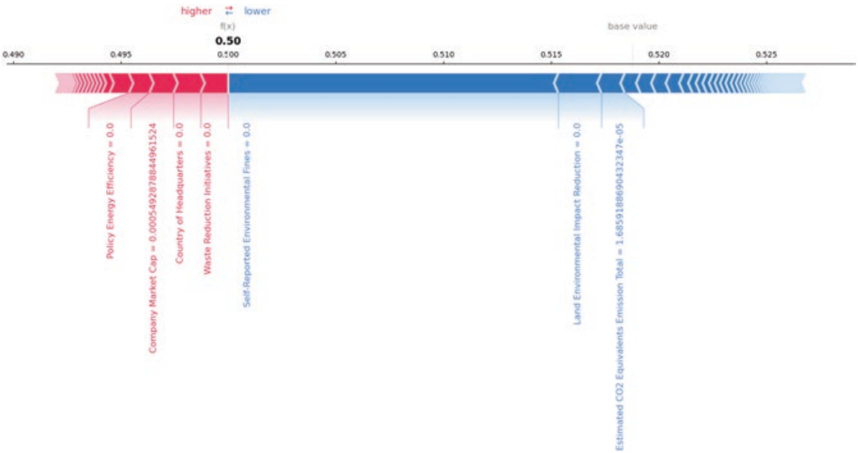

Fig. 6.2 SHapley Additive exPlanations force plot showing the materiality of environmental topics for Extendicare Incorporated

noteworthy omissions. The lack of waste reduction initiative despite disposing of a considerable amount of waste and the lack of an energy policy are two aspects that, according to the model, have a material significance for how Extendicare is perceived as compliant or non-compliant with environmental norms. In addition, the model also notes that Extendicare is a large corporation based in Canada. These circumstances constitute a risk of being scrutinized regarding environmental performance, most likely because of high environmental standards in Canada. Therefore, these considerations may be regarded as irrelevant for a materiality assessment and can be ignored. The model does not distinguish between the meanings of items other than their impacts on the assessment of the risk of being criticized for non-compliance with environmental norms, so the model's output must be manually interpreted.

Figure 6.3 exemplifies how the model conducts its materiality assessment of Delta Electronics Inc., an information technology sector company based in Taiwan. The company takes pride in environmental responsibility and has been in the Dow Jones Sustainability Indices for several years, suggesting we should expect the blue arrows to dominate again. With the forces pushing the rating to be high, the blue arrows dominate the red. This blue dominance indicates that there are very few

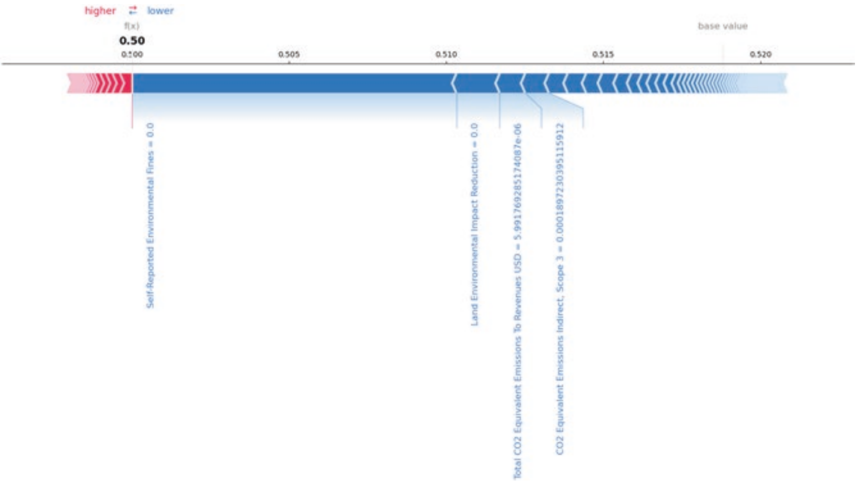

Fig. 6.3 The random forest model's materiality assessment of Delta Electronics Incorporated

doubts from our compliance-based model that Delta Electronics is a company highly likely to comprehensively comply with environmental responsibilities. The figure demonstrates that most material items are blue, i.e., environmental performance strengths or risk-reducers, rather than showing red flags. The model discounts the reduction as an indication that the company is not involved in environmentally risky activities. Low CO_2 emission confirms the image of a company exposed to low risks of being non-compliant with environmental responsibilities.

Per Fig. 6.4, our following example is Allergan Plc., a multinational pharmaceutical company in the healthcare sector that develops, manufactures, and markets branded and generic prescription drugs, medical aesthetics and over-the-counter products. The company was founded in 1983 and is headquartered in Dublin, Ireland. Allergan has a presence in more than 100 countries and employs over 17,000 people worldwide. Allergan operates in three main segments: branded pharma, medical aesthetics, and over-the-counter products. Some of Allergan's well-known brands include Botox, Juvederm, and Restasis. Botox is a cosmetic product used for reducing the appearance of wrinkles and facial lines, while Juvederm is a dermal filler used for smoothing out wrinkles and restoring facial volume.

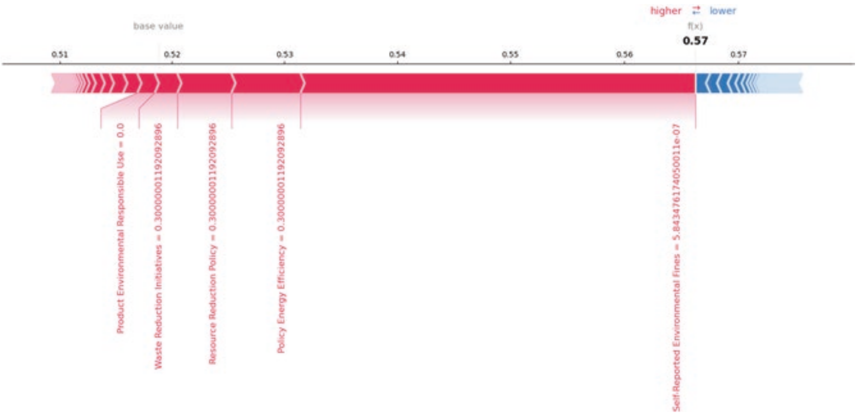

Fig. 6.4 The random forest model's materiality assessment of Allergan Public Limited Company

Contrary to our two first examples, the materiality assessment appears dominated by red, signaling high-impact or high-risk topics. We may immediately conclude that Allergan is doing worse than the average corporation, and we see the most material reasons why the model has a negative estimate of Allergan's environmental compliance. The material topics according to our model's estimation are related to its history of being fined for breaches and its lack of energy efficiency, resource, and waste reduction policies. There is also a material concern about how the use of its products impacts the environment. Thus, these observations are made by the model, and are not based on external research sources.

It is clear from these examples that the materiality assessment model identifies environmental indicators, not full-fledged topics. Because the model provides indications and hints about where a company may be environmentally problematic, using the model presupposes that the identified indicators are interpreted. The red flag for waste reduction initiatives relating to Allergan must be understood with some background knowledge of what Allergan does. For some companies, an initiative to reduce waste may be good, but in this case, the model suggests that it is a sign of underlying environmental problems. Disaggregated indicators may mean different things in different contexts. Comparing companies in different sectors is obviously a quite different exercise than comparing companies in the same sector. A user of an automatic model must understand the extent to which the model is able to provide relevant information for comparison between sectors when the environmental impacts per definition are very different.

6.4 Conclusions and Limitations

This chapter addresses the difficulties with sustainability reporting's materiality assessment. The difficulties are expected to be significant when assessing impact materiality according to the GRI and the coming EU standards because such assessments presuppose that the reporting company can determine which aspects of a company's activities have the most material effects on the environment from the perspective of the broadest possible stakeholder groups. The reviewed literature univocally describes materiality assessment as subjective, uncertain, and imprecise without providing a clear reference point, such as the rule-of-thumb percentage used in financial reporting as a materiality threshold (Moroney & Trotman, 2016).

The usefulness of an automated assessment tool should be judged on how well it performs on the assessment compared to alternative methods. Alternative methods offer mainly procedural support or conceptual focus, such as Calabrese et al. (2013, 2016) or Wu et al. (2018). The latter simply propose that two static, pre-defined materiality assessments are Sustainability Accounting Standards Board Materiality Map and Global Reporting Initiative (GRI) Sustainability Topics for Sectors or GRI's Sustainability Disclosure Database and that modelling from a life-cycle perspective could be ways to assess materiality. The only realistic alternative appears to be how the assessments are conducted manually. Descriptions are scarce, and companies do not report extensively about their method of materiality assessments (Farooq et al., 2021; Machado et al., 2021). Unfortunately, we find no evidence of the validity and accuracy of any presently practiced assessment method. Some evidence even indicates that companies disclose less and less when they adopt the more standardized guidance that some standards have implemented (Borgert et al., 2018). Our literature review shows that accounting for materiality is considered a serious problem, for which offered solutions are procedural recipes that yield little evidence of results.

Our suggested approach is not without limitations. First, we do not adjust the model for the unequal news media attention paid to companies, which likely leads to different companies having different likelihoods of being the subject of critique, not corresponding to differences in their non-compliance with environmental norms. Media reporting might also differ between countries. What is perceived as alarming in one part of the world may be more or less ignored in another. Furthermore, we do not intend to develop an operational assessment method and explore aspects of the construction of a method, which means that we include several indicators that should not be adopted in an actual decision-support tool. Another limitation is that we have not fine-tuned the parameters in the machine learning algorithm, which would enable higher predictive performance. The controversies database might also be replaced by a database that offers a higher coverage of companies and geographical settings. Future development of the methodology could also attempt to make the predictions unbiased regarding firm size and profitability, two biases that seriously have troubled the use of ESG metrics (Drempetic et al., 2020). Ultimately, however, uncritical reliance on any decision-support tool is not viable. A machine learning method may provide suggestions as input to materiality assessment but cannot replace a thorough manual process. Reliance on a decision-support tool should be viewed as a way of making informed decisions based on a rapid screening of large amounts of information that would be difficult to obtain manually. The proposed decision-support tool may enable analysts to detect greenwashing because the model can report the environmental circumstances present in a particular company that typically results in serious criticism and therefore should be a warning signal to investors that care about sustainable business.

References

Adams, C. A. (2004). The ethical, social and environmental reporting-performance portrayal gap. *Accounting, Auditing & Accountability Journal, 17*(5), 731–757. https://doi.org/10.1108/09513570410567791

Adams, C. A., & Frost, G. R. (2006). The internet and change in corporate stakeholder engagement and communication strategies on social and environmental performance. *Journal of Accounting & Organizational Change, 2*(3), 281–303. https://doi.org/10.1108/18325910610690090

Alpaydin, E. (2010). *Introduction to Machine Learning* (2nd ed.). The MIT Press.

Bellantuono, N., Pontrandolfo, P., & Scozzi, B. (2016). Capturing the stakeholders' view in sustainability reporting: A novel approach. *Sustainability, 8*(4), 379. https://doi.org/10.3390/SU8040379

Boesso, G., & Kumar, K. (2009). Stakeholder prioritization and reporting: Evidence from Italy and the US. *Accounting Forum, 33*(2), 162–175. https://doi.org/10.1016/j.accfor.2008.07.010

Borgert, T., Donovan, J. D., Topple, C., & Masli, E. K. (2018). Initiating sustainability assessments: Insights from practice on a procedural perspective. *Environmental Impact Assessment Review, 72*, 99–107. https://doi.org/10.1016/J.EIAR.2018.05.012

Calabrese, A., Costa, R., Levialdi, N., & Menichini, T. (2016). A fuzzy analytic hierarchy process method to support materiality assessment in sustainability reporting. *Journal of Cleaner Production, 121*, 248–264. https://doi.org/10.1016/J.JCLEPRO.2015.12.005

Calabrese, A., Costa, R., & Menichini, T. (2013). Using Fuzzy AHP to manage Intellectual Capital assets: An application to the ICT service industry. *Expert Systems with Applications, 40*(9), 3747–3755. https://doi.org/10.1016/j.eswa.2012.12.081

Canning, M., O'Dwyer, B., & Georgakopoulos, G. (2019). Processes of auditability in sustainability assurance – The case of materiality construction. *Accounting and Business Research, 49*(1), 1–27. https://doi.org/10.1080/00014788.2018.1442208

Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies, 26*(3), 1176–1248. https://doi.org/10.1007/S11142-021-09609-5

Christensen, M., & Clawson, S. (2018). European SRI study EUROSIF. In *Revised édition*. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:European+SRI+Study#8

De Villiers, C., & van Staden, C. J. (2011). Where firms choose to disclose voluntary environmental information. *Journal of Accounting and Public Policy, 30*(6), 504–525. https://doi.org/10.1016/j.jaccpubpol.2011.03.005

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics, 167*(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

EFRAG. (2022). European Financial Reporting Advisory Group. [Draft] European Sustainability Reporting Guidelines 1 Double materiality conceptual guidelines for standard-setting. Available at: https://www.efrag.org/Assets/Download?assetUrl=/sites/webpublishing/SiteAssets/Appendix%202.6%20-%20WP%20on%20draft%20ESRG%201.pdf&AspxAutoDetectCookieSupport=1

Eilifsen, A., & Messier, W. F. (2015). Materiality guidance of the major public accounting firms. *Auditing, 34*(2), 3–26. https://doi.org/10.2308/AJPT-50882

Farooq, M. B., & de Villiers, C. (2019). Understanding how managers institutionalise sustainability reporting: Evidence from Australia and New Zealand. *Accounting, Auditing and Accountability Journal, 32*(5), 1240–1269. https://doi.org/10.1108/AAAJ-06-2017-2958/FULL/XML

Farooq, M. B., Zaman, R., Sarraj, D., & Khalid, F. (2021). Examining the extent of and drivers for materiality assessment disclosures in sustainability reports. *Sustainability Accounting, Management and Policy Journal, 12*(5), 965–1002. https://doi.org/10.1108/SAMPJ-04-2020-0113/FULL/PDF

Faulkner, R. R. (2011). Corporate wrongdoing and the art of the accusation. *Anthem Publishers.* https://doi.org/10.7135/UPO9780857284204

Fiaschi, D., Giuliani, E., Nieri, F., & Salvati, N. (2020). How bad is your company? Measuring corporate wrongdoing beyond the magic of ESG metrics. *Business Horizons, 63*(3), 287–299. https://doi.org/10.1016/j.bushor.2019.09.004

Guix, M., Font, X., & Bonilla-Priego, M. J. (2019). Materiality: stakeholder accountability choices in hotels' sustainability reports. *International Journal of Contemporary Hospitality Management, 31*(6), 2321–2338. https://doi.org/10.1108/IJCHM-05-2018-0366

Hassel, L. G., & Semenova, N. (2018). Engagement dialogue as a Nordic sustainable and responsible investment (Sri) strategy. *Challenges in Managing Sustainable Business: Reporting, Taxation, Ethics and Governance, 179–204*. https://doi.org/10.1007/978-3-319-93266-8_8/COVER

Hsu, C. W., Lee, W. H., & Chao, W. C. (2013). Materiality analysis model in sustainability reporting: A case study at Lite-On Technology Corporation. *Journal of Cleaner Production, 57*, 142–151. https://doi.org/10.1016/j.jclepro.2013.05.040

Jones, P., Comfort, D., & Hillier, D. (2016). Materiality in corporate sustainability reporting within UK retailing. *Journal of Public Affairs, 16*(1), 81–90. https://doi.org/10.1002/PA.1570

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review, 16*(2), 366. https://doi.org/10.5465/amr.1991.4278958

Jørgensen, S., Mjøs, A., & Pedersen, L. J. T. (2022). Sustainability reporting and approaches to materiality: tensions and potential resolutions. *Sustainability Accounting, Management and Policy Journal, 13*(2), 341–361. https://doi.org/10.1108/SAMPJ-01-2021-0009

Kaur, A., & Lodhia, S. (2018). Stakeholder engagement in sustainability accounting and reporting: A study of Australian local councils. *Accounting, Auditing and Accountability Journal, 31*(1), 338–368. https://doi.org/10.1108/AAAJ-12-2014-1901/FULL/PDF

Knechel, W. R. (2021). The future of assurance in capital markets: Reclaiming the economic imperative of the auditing profession. *Accounting Horizons, 35*(1), 133–151. https://doi.org/10.2308/HORIZONS-19-182

Libby, R., & Brown, T. (2013). Financial statement disaggregation decisions and auditors' tolerance for misstatement. *The Accounting Review, 88*(2), 641–665. https://doi.org/10.2308/accr-50332

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 2017-December*, 4766–4775.

Machado, B. A. A., Dias, L. C. P., & Fonseca, A. (2021). Transparency of materiality analysis in GRI-based sustainability reports. *Corporate Social Responsibility and Environmental Management, 28*(2), 570–580. https://doi.org/10.1002/csr.2066

Messier, W. F., Martinov-Bennie, N., & Eilifsen, A. (2005). A review and integration of empirical research on materiality: Two decades later. *Auditing, 24*(2), 153–187. https://doi.org/10.2308/AUD.2005.24.2.153

Moroney, R., & Trotman, K. T. (2016). Differences in auditors' materiality assessments when auditing financial statements and sustainability reports. *Contemporary Accounting Research, 33*(2), 551–575. https://doi.org/10.1111/1911-3846.12162

Nieri, F., & Giuliani, E. (2018). International business and corporate wrongdoing: A review and research agenda. In D. Castellan, R. Narula, Q. Nguyen, I. Surdu, & J. Walker (Eds.), *Contemporary issues in international business* (pp. 35–53). Springer. https://doi.org/10.1007/978-3-319-70220-9_3

O'Dwyer, B., Unerman, J., & Hession, E. (2005). User needs in sustainability reporting: Perspectives of stakeholders in Ireland. *European Accounting Review, 14*(4), 759–787. https://doi.org/10.1080/09638180500104766

Puroila, J., & Mäkelä, H. (2019). Matter of opinion: Exploring the socio-political nature of materiality disclosures in sustainability reporting. *Accounting, Auditing and Accountability Journal, 32*(4), 1043–1072. https://doi.org/10.1108/AAAJ-11-2016-2788/FULL/XML

Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P., Rana, T., & Danielson, M. (2022). Prediction of environmental controversies and development of a corporate environmental performance rating methodology. *Journal of Cleaner Production, 130979*, 130979. https://doi.org/10.1016/J.JCLEPRO.2022.130979

Torelli, R., Balluchi, F., & Furlotti, K. (2019). The materiality assessment and stakeholder engagement: A content analysis of sustainability reports. *Wiley Online Library, 27*(2), 470–484. https://doi.org/10.1002/csr.1813

USSIF. (2018). *Report on US sustainable, responsible and impact investing trends*. https://www.ussif.org/files/Trends/Trends2018 executive summary FINAL.pdf

Wu, S. R., Shao, C., & Chen, J. (2018). Approaches on the screening methods for materiality in sustainability reporting. *Sustainability, 10*(9), 3233. https://doi.org/10.3390/SU10093233

7

Artificial Intelligence and the Food Value Chain

Stefan Wendt and Throstur Olaf Sigurjonsson

7.1 Introduction

The production, processing, and distribution of food face and, in turn, create various environmental and social challenges. For example, the use of fertilizers in the cultivation of plants and the use of antibiotics and other drugs in livestock raising are seen as problematic due to, for example, toxic chemicals in various fertilizers and the development of antibiotic-resistant bacteria as a result of widespread use of antibiotics. Increasingly large areas of arable land face the risk of becoming non-arable due to climate change. Livestock farming and the worldwide transportation of food substantially contribute to emissions, and the conditions of livestock farming are often considered to be in violation of animal

S. Wendt (✉)
Department of Business, Bifröst University, Borgarnes, Iceland
e-mail: stefanwendt@bifrost.is

T. O. Sigurjonsson
University of Iceland, Reykjavík, Iceland
e-mail: olaf@hi.is

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_7

rights. Food needs are increasing globally, and food shortage is a problem in some parts of the world, while food waste is prominent in others.

This chapter aims to examine how the application of artificial intelligence (AI) can make the food value chain more sustainable. Specifically, we aim to answer how AI can help make food production, aggregation, processing, and distribution more environmentally friendly and socially responsible. Amongst other things, we address how AI can help optimize the use of resources and reduce the use of fertilizers and drugs, as well as how AI can help mitigate climate risks, e.g., in agriculture.

Analyzing the food value chain, this means all value-adding activities when making food products (FAO, 2014), is particularly important because of its relevance in food security and food safety (Sturludottir et al., 2021) and to the sustainability challenges it exacerbates regarding the use of water, fertilizers, animal welfare, food waste, and so forth (FAO, 2014). Nearly one-third of global greenhouse gas emissions result from food systems. In this sense, the "Farm to Fork Strategy" of the European Commission, as part of the European Green Deal, aims to make food systems fair, healthy, and environmentally friendly (European Commission, 2020). A sustainable food value chain would contribute to many of the United Nations Sustainable Development Goals, most notably to Goal 2 "Zero Hunger", Goal 12 "Responsible Consumption and Production", and Goals 14 and 15 "Life below Water" and "Life on Land" (United Nations, n.d.). Karanth et al. (2023) argue that the impacts of climate change are most visible in the food ecosphere, impacting the yield, quality, and safety of food.

The chapter explains the implications for food producers, processing companies, and distributors. Despite the increasing digitalization and application of AI in the food industry (Kler et al., 2022; Mavani et al., 2022), to the best of our knowledge, the combination of AI, the food industry, and sustainability has only rarely been addressed in the literature (see Di Vaio et al., 2020). In particular, the role of AI in sustainability and the role of business models and multi-stakeholder approaches have been neglected (D'Amore et al., 2022). However, cooperation, collaboration, and knowledge sharing between stakeholders are essential for AI's safe and sustainable use (Hassoun et al., 2022; Ryan, 2022). Our research is also relevant to policymakers with respect to establishing corresponding regulatory and supportive frameworks for these topics (D'Amore et al., 2022).

In the next section, we present background information on the sustainability of value chains, the food value chain, and AI. Subsequently, we explore the sustainability benefits of applying AI in the food value chain. We then discuss the main results and conclude in the last section of this chapter.

7.2 Background

7.2.1 The Sustainability of Value Chains

Not only do individual companies' activities and processes need to be sustainable, but cross-sector, upstream and downstream sustainability are equally important, if not more so. This means that the sustainability of a company's suppliers and sourcing (upstream), as well as sustainable distribution channels and customer behavior (downstream), are essential to making the entire value chain sustainable (Wagner, 2021). The topic of sustainability across the supply chain has attracted increased attention in recent years, even though considerations and assessments of the subject primarily focus on individual companies and less so on the entire value chain. Reasons for such a narrow focus are manifold and include, for example, the lack of standardization and insufficient quality of information related to sustainability (Jonsdottir et al., 2022), differences in legal requirements and jurisdiction along the value chain, particularly in the case of international value chains, or opaque value chains and governance structures along the value chain.

However, there have been efforts to broaden the scope of studies and to facilitate an assessment of value chains as a whole. For example, companies within the value chain, environmental rating agencies and other organizations use concepts such as Scope 1, 2, and 3 emissions to capture the environmental sustainability of value chains. Scope 1 greenhouse gas emissions result from sources owned and controlled by a company, Scope 2 emissions refer to the production of the energy purchased and consumed by the company, while Scope 3 covers emissions that occur upstream or downstream of the value chain (WRI & WBCSD, 2004). This means that Scope 3, and to some extent Scope 2, cover value chain effects even though emissions are not the only relevant issues relating to sustainability along the value chain. Sustainability assessments and ratings have only just begun to integrate value chain effects.

7.2.2 Sustainable Food Value Chains

The core of the food value chain consists of food production, aggregation, processing, and distribution. Food production includes farming, fishing, forest harvesting, and agroforestry. Aggregation includes the collection of food from multiple sources, such as small local producers, to generate a more consistent food supply. Food processing can come in various forms, including one or more steps from the raw material and original produce to the products sold in retailing or wholesale. Wholesale and retail distribution cover various distribution channels. Financing and supply of inputs and services, such as seeds, fertilizers or transportation, support the core value chain (FAO, 2014).

The sustainability dimensions of the food value chain include environmental and societal/social aspects jointly with the governance structure (see the sustainable food value chain network of the Food and Agriculture Organization of the United Nations (FAO, 2014)). The environmental perspective relates to, for example, carbon footprint, plant and animal health (such as husbandry conditions), water footprint, soil conservation, biodiversity, food loss and waste, as well as toxicity. It is also important to note that this does not only cover the impact of the value chain on the natural environment, but also the impact of the natural environment on the value chain and their interdependence. Societal elements and their role in the sustainable food value chain framework relate to sociocultural, organizational, institutional, and infrastructural elements and include, for example, the distribution of added value, cultural traditions (such as traditional food or culturally accepted food processing), nutrition and health, workers' rights and safety, and animal welfare. Governance structure relates to, for example, ownership structures and contractual arrangements between actors at the different stages of the chain, and organizational structures such as cooperatives at the production level. A number of these aspects also relate to economic and financial factors, such as jobs and corresponding income, tax revenues, and profits (FAO, 2014). For a graphical representation of the sustainable food value chain framework, see Fig. 7.1.

The sustainability of food value chains has also attracted the attention of national and international policymakers. For example, the aims of the European Commission's "Farm to Fork Strategy" are to accelerate the transition to food systems that have a neutral or positive environmental impact, support climate change mitigation and adaptation, reverse the loss of biodiversity, ensure food security, improve nutrition and health, preserve the affordability of food, generate fair economic returns, foster competitiveness, and promote fair trade (European Commission, 2020). In the same vein, Marvin et al. (2022) emphasize the importance of improving efficiency, quality, and resilience to disruptions, as well as

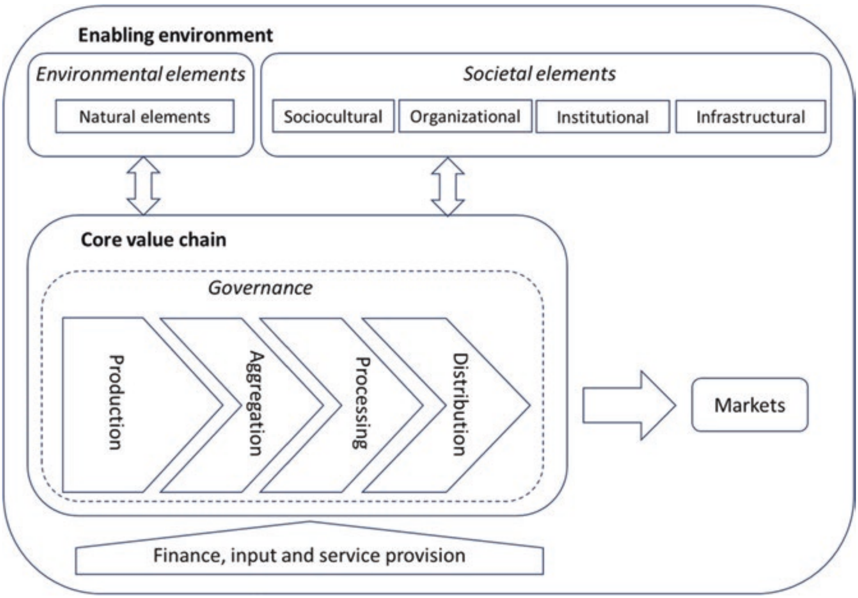

Fig. 7.1 Sustainable food value chain framework. (Source: FAO (2014), adapted by the authors)

reducing the use of pesticides and fertilizers, food waste and food loss, and environmental footprints.

7.2.3 Artificial Intelligence (AI)

Due to the extremely rapid development of AI, definitions of AI also need to be adjusted and updated frequently to keep pace. The umbrella term of artificial intelligence might be considered too broad when it comes to the intricacies of its applications in companies and strategies. Therefore, smaller and more homogenous subcategories of AI, such as speech or object recognition, image classification, robotic perception, language processing, knowledge engineering and machine learning, are considered to better grasp the underlying ideas, concepts, and mechanisms at play (e.g., Sigurjonsson & Wendt, 2023). As it stands, lists or overviews of AI solutions cannot be exhaustive, but new or more advanced and sophisticated solutions can be expected in the near future.

For the purpose of this chapter, we understand AI as the ability of computers and robots to collect, store and transfer, process and analyze, and visualize data (e.g., Zhou et al., 2022). AI enables computers and robots to do tasks that were traditionally done by humans, mimicking human intelligence or that humans are incapable of doing (Raghavendra et al., 2022). Data gathering typically happens via sensors, scanners, cameras, and various other such technologies. The processing and analyzing of information occur via computer algorithms programmed by humans. Automated decisions based on the outcome of the information processing and analyzing shall also be considered part of AI solutions. Such decisions can result in physical actions by robots, including humanoid robots, or trigger adjustments in systems, such as heating or cooling, or changes in production machinery. The analysis can also lead to the creation of reports or trigger warning signals.

More recently, machine learning and deep learning have enhanced AI. Machine learning is based on the application of algorithms and statistical techniques to train machines to recognize patterns and make predictions. Machine learning goes beyond traditional programming for specific tasks and focuses on broader applications (Hassoun et al., 2022). Deep learning via neural networks has proven effective in extracting patterns from big data for, e.g., image and voice recognition, text summarization and classification (Misra et al., 2022).

In the business context, the fast development of AI has led to increased opportunities for applications that support business operations and efficiency. AI has been successfully employed to improve efficiency and profitability, most notably through the automation of tasks that humans previously performed, by more accurately forecasting things such as sales, or by diagnosing when maintenance is needed for machinery and managing inventory. Challenges regarding implementing AI arise from potentially high costs, particularly for small companies in the value chain, as well as from potential resistance if AI is considered a threat to existing business models (Mahroof et al., 2022).

7.3 Sustainability Benefits from Artificial Intelligence in the Food Value Chain

Potential and actual applications of AI in the food value chain exist both along the different stages of the value chain and within the single stages of the chain. Kler et al. (2022) categorize these into smart farming, smart transportation, smart processing, and smart distribution and consumption (see also Marvin et al., 2022). The benefits of applying AI to the entire value chain include improved transparency, traceability, efficiency, and quality monitoring to safeguard food quality and safety, as well as increased ease in providing stakeholders with relevant information (Di Vaio et al., 2020; Karanth et al., 2023; Mahroof et al., 2022). Karanth et al. (2023) and Marvin et al. (2022) underline the use of prediction models along the value chain, for example, to forecast climate change-induced food safety risks and demand patterns. In the following section, we describe the potential sustainability benefits of applying AI at different stages of the value chain and address further benefits of its application beyond single stages (i.e., to the process as a whole).

7.3.1 Food Production

In food production, AI can be applied for crop, greenhouse, and livestock monitoring, as well as towards optimizing the use of fertilizers and water, resulting in what can be called data-driven and precision agriculture (Di Vaio et al., 2020; Marvin et al., 2022). Permanent real-time monitoring via cameras and other sensors of, for example, soil humidity or temperature and crop conditions can help adjust parameters for more consistent and best possible results (Hassoun et al., 2022). Combining this information with weather information and forecasts allows for even more precise identification of needs such as irrigation (Hachimi et al., 2023). Based on this information, irrigation can be largely automatized and adjusted to the needs of different parts of arable land. Similarly, fertilizer use can be optimized and targeted to specific needs; based on the crop conditions, AI identifies the optimal time for harvesting. Subsequently, storage conditions can be monitored, and suboptimal conditions automatically be reported and measured to adjust the conditions triggered. In livestock and fish farming, automated monitoring and data analysis can help optimize feeding patterns and monitor the health of animals via weight and behavior analysis. Automated monitoring allows for the identification of early signs of diseases (both in crops and livestock), and the application of blockchain solutions can help to trace outbreak paths, e.g., when food products or livestock are transported between players in the value chain (Kamath, 2018; Misra et al., 2022). Moreover, urban agriculture can benefit from AI applications, as it often requires an even more controlled environment (Shamshiri et al., 2018).

7.3.2 Food Aggregation

When it comes to optimizing the aggregation of supplies from different producers, combining and analyzing information from various producers or farms within and across regions with AI solutions can help predict availability and timing. The information can be used to steer harvesting times and coordinate storage use and delivery from different producers (Marvin et al., 2022). As a result, supply can be coordinated in such a way as to make food processing less dependent on fluctuations, avoid food spoilage, and result in a more stable flow of food products overall.

7.3.3 Food Processing

In the processing stages, AI can be used for process control and monitoring to reduce food waste and loss (Marvin et al., 2022). One such benefit, for example, follows from the more precise cutting of meat off bones via the automated measuring of each individual animal and a corresponding analysis to identify the optimal cuts, which would reduce meat waste from inaccurate cuts. Based on data analysis, advanced machinery can be automatically fine-tuned and adjusted according to the characteristics of each animal (Sigurjonsson & Wendt, 2023).

Another benefit of using AI is non-destructive quality monitoring via sensors and automated image analysis. Damage, defects, and other irregularities, such as foreign objects, and abnormal color and texture, can be detected. Thanks to the high computing capacity of AI, each unit can be quality checked instead of the small sample that could be manually checked. Furthermore, the quality check does not impact the product itself, i.e., the product is not destroyed by the quality check as would often be the case in manual quality checks (El-Mesery et al., 2019; Hassoun et al., 2023; Sigurjonsson & Wendt, 2023).

For optimal use of machinery and fine-tuning based on the characteristics of the raw material, information flow along the value chain is of essential importance in the processing stages. Based on continuous information flow from farms, fishing vessels and so forth, processing can be automatically planned and coordinated. Similarly, information in (nearly) real-time on customer behavior, such as purchasing choices, swings in demand or demand shocks (as was observed at the height of COVID-19 when restaurants closed), can be immediately considered and food processing adjusted accordingly (Sigurjonsson & Wendt, 2023).

In this case, AI makes the food value chain more sustainable by reducing food loss and improving food safety. A better match with customer demands helps avoid food waste; for the same amount of output less input is needed, which means that fewer animals need to be slaughtered.

Whether or not efficiency gains result in reduced use of resources or in higher productions remains to be seen (Sigurjonsson & Wendt, 2023). In various instances, especially in regions experiencing a food shortage, increased production based on the same input amount might be a socially desired outcome.

7.3.4 Food Distribution

Distribution channels and systems for food products can benefit from the application of AI in various ways. First, monitoring conditions during transportation, such as temperature and humidity, becomes available without interruptions thanks to sensors that collect the corresponding data and an automated analysis via algorithms. Based on defined thresholds, air conditioning systems in containers and trucks as well as in warehouses and shops can use this information to adjust automatically, which means that temperature control becomes much more precise. The application of AI during transportation therefore ensures proper transport conditions and reduces perishability as well as the risk of spoilage, which represents both environmental and social benefits in the sense of reduction of food waste and improved food safety (e.g., Mahroof et al., 2022; Marvin et al., 2022; Vernier et al., 2021). AI-based routing, scheduling, and tracking of delivery performance can help optimize transportation routes and further reduce transportation times with corresponding benefits for not only food quality, but also fuel consumption (Hassoun et al., 2022).[1] Given the enormous quantities of fossil fuels used for food transportation, with transport accounting for nearly 20% of total food-system emissions (Li et al., 2022), reducing the consumption of fossil fuels by optimizing transportation routes would help reduce greenhouse gas emissions significantly. This would be particularly impactful, as renewable energy has not yet replaced fossil fuels in transportation via planes, vessels, trucks, etc.

At various stages during transportation, such as loading, unloading, and arrival, image recognition and analysis can be used to identify

[1] See also Oehler et al. (2013) on smart transportation systems.

damaged products or packaging (El-Mesery et al., 2019; Hassoun et al., 2023). The corresponding benefits with regard to product quality represent social benefits in terms of food safety. Features such as the automated scanning of bar codes upon arrival to, or when leaving, warehouses or shops, combined with the automated tracking of the location where goods are stored and an analysis of the amount in stock, allow for automated ordering processes and prevent stock misplacement. Moreover, the visual supervision of shelves in shops and automated analysis of corresponding pictures can identify empty shelves and trigger either the replenishment from the warehouse or the ordering of additional supply (Sigurjonsson & Wendt, 2023). While these benefits initially appear to be primarily gains in terms of efficiency, in the sense of optimized product flow and inventory management, these improvements also contribute to a more sustainable value chain by reducing excess stock, which in turn reduces food waste. Furthermore, when considering food products that cover basic needs, automated processes can also help secure continuous or regular availability, which can be categorized as a social benefit. A further social benefit of automated warehouses results from the avoidance of dangerous and unfavorable work conditions, such as extremely low temperatures in cold storages, and thus a reduction in the number of workplace accidents.

When it comes to the final stages of distribution, customer behavior plays an essential role. AI-based systems can analyze and predict customer behavior by, for example, anticipating seasonal variations in demand or shoppers' in-shop behavior. Moreover, smart home devices, such as smart refrigerators, can help optimize purchase behavior, avoid overbuying and, in turn, reduce food waste. These decision support systems can help make inventory management, distribution, and consumption more sustainable (e.g., Kumar et al., 2021; Mahroof et al., 2022; Marvin et al., 2022).

7.4 Discussion and Conclusions

The purpose of this chapter was to examine how the application of AI can make the food value chain more sustainable. We analyze the literature to identify how AI can help make food production, aggregation, processing, and distribution more environmentally friendly and socially responsible. While some of the findings are specific to single stages, various aspects are also relevant along the entire value chain.

We found that AI systems, including continuous monitoring with sensors and cameras, automated information analysis, and the prediction of various conditions, can help implement automatized actions or trigger human intervention to make agriculture more precise and, thus, reduce the use of resources. Moreover, diseases can be detected early on with AI, which facilitates swifter countermeasures and reduces, or even avoids, the use of chemicals. Optimized coordination between producers as well as automated measuring, AI-supported processing lines, continuous quality monitoring, and optimized storage and transportation can reduce food loss and waste and, in turn, reduce the number of resources needed.

It becomes evident that most of the environmental benefits result from the efficiency gains that can be achieved when using AI. Social benefits result from increased food security and safety due to improved quality monitoring and reduced food waste and loss. Moreover, optimized warehouse and inventory management based on advanced demand analysis and predictions can help reduce food spoilage. Automatization can also help avoid dangerous working conditions.

Continuous monitoring of soil and crop conditions, in combination with other data, such as weather data, can help identify early signs of droughts and other climate change-induced weather conditions. Similarly, livestock monitoring can help identify early signs of diseases that might result from extreme heat waves, and farmers and health authorities can implement corresponding countermeasures. Optimizing transportation routes, transportation conditions, and storage conditions, as well as inventory and warehouse management with the help of AI can greatly reduce transportation needs, which can significantly reduce greenhouse gas emissions.

Overall, increased efficiency in the form of reduced water and energy consumption, the reduction of waste through more accurate demand prediction, as well as increased adaptation capacity through forecasting and simulation of future scenarios results in potential environmental and social sustainability benefits. Moreover, the improved sustainability of production and manufacturing can be achieved through more precise monitoring and lower emissions across the supply chain by optimizing routes and transportation options (see also Kumar et al., 2021). However, when implementing AI solutions in companies, the primary focus still seems to lie on financial or economic efficiency gains. In contrast, sustainability benefits are relegated to (positive) side effects (Sigurjonsson & Wendt, 2023).

Potential hindrances to the implementation of AI include a lack of trust, digital infrastructure, and technical skills, as well as the need for substantial investments (e.g., Hassoun et al., 2022; Karanth et al., 2023; Zhou et al., 2022). Aspects such as trust in institutions and persons, risk attitude, knowledge and skills have been shown to influence technology acceptance and/or algorithm aversion (Behrenbruch et al., 2013; Oehler et al., 2022) and, hence, also need to be considered in the context of implementing AI. More advancement in the implementation of AI will require revised educational curricula to ensure the required skill sets in employees entering the job market. Besides the need for experts in programming AI solutions, many employees along the value chain need the corresponding basic understanding and skills to apply AI solutions or to use machinery and software solutions with integrated AI applications without being AI experts themselves. Moreover, many AI solutions are still very expensive and need high investment amounts that might not yet be affordable, in particular, for smaller companies or companies that have not yet embarked on the digitalization journey.

Potential negative social and ethical impacts of the implementation of AI result from job displacement due to automation and a loss of privacy as a result of monitoring in the workplace and the tracking of consumer behavior (Ryan, 2022).[2] On a global scale, social inequality might increase when access to AI is limited, particularly in developing countries (Di Vaio et al., 2020; Ryan, 2022). These ethical concerns need to be taken seriously in the public debate, in policymaking for the development and implementation of AI solutions, and when integrating AI in corporate strategy and operations, because the successful implementation of AI needs to be built on trust to foster technology acceptance and overcome algorithm aversion as mentioned above. These ethical concerns relate to

[2] For corresponding cybersecurity risks, see also Marvin et al. (2022).

transparency, traceability, explainability, interpretability, accessibility, accountability, and responsibility (Manning et al., 2022).

Questions have also arisen as to whether the benefits from the use of AI actually lead to more sustainable processes; concerns also exist surrounding the ecological and social impacts of AI itself. Dauvergne (2022) argues that efficiency gains lead to more production instead of less use of resources and that AI has led to more natural resource extraction. Moreover, Ryan (2022) argues that the use of AI can result in increased energy consumption (for AI solutions) and greenhouse gas emissions. These concerns and the continued development of AI, and its implementation in various stages of the food value chain, call for ongoing research to be able to benefit as much as possible from the sustainability benefits of implementing AI. Expanding the use of AI will not solve all sustainability issues in the food value chain. This means other, i.e., non-AI based approaches to make the food value chain more environmentally friendly and socially responsible must also be considered and developed. Such approaches might include restructuring of entire value chains to shorten transportation distances, stronger focus on local sourcing and local distribution, stricter sustainability criteria for selecting suppliers, nudging customers towards more environmentally and socially friendly products, to name a few. These approaches and AI solutions must be analyzed and compared to identify and implement the most effective approaches. In this sense, a combination of AI solutions and non-AI solutions might be a promising approach in order to benefit from both developments.

Acknowledgements The authors would like to thank participants of the conference Vidskipti og Visindi at the University of Iceland in Reykjavik, Iceland, in March 2023 and of the Sustainability Science Days at the University of Helsinki, Finland, in May 2023 for helpful comments. The authors would also like to thank Rannis – The Icelandic Centre for Research for financial support.

References

Behrenbruch, K., Söllner, M., Leimeister, J. M., & Schmidt, L. (2013). Understanding diversity – The impact of personality on technology acceptance. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-computer interaction – INTERACT 2013* (pp. 306–313). Springer. https://doi.org/10.1007/978-3-642-40498-6_23

D'Amore, G., Di Vaio, A., Balsalobre-Lorente, D., & Boccia, F. (2022). Artificial intelligence in the water–energy–food model: A holistic approach towards sustainable development goals. *Sustainability, 14*(2), 867. https://doi.org/10.3390/su14020867

Dauvergne, P. (2022). Is artificial intelligence greening global supply chains? Exposing the political economy of environmental costs. *Review of International Political Economy, 29*(3), 696–718. https://doi.org/10.1080/09692290.2020.1814381

Di Vaio, A., Boccia, F., Landriani, L., & Palladino, R. (2020). Artificial intelligence in the Agri-food system: Rethinking sustainable business models in the COVID-19 scenario. *Sustainability, 12*(12), 4851. https://doi.org/10.3390/su12124851

El-Mesery, H. S., Mao, H., & Abomohra, A. E.-F. (2019). Applications of non-destructive technologies for agricultural and food products quality inspection. *Sensors, 19*(4), 846. https://doi.org/10.3390/s19040846

European Commission. (2020). *A farm to fork strategy for a fair, healthy and environmentally-friendly food system*. European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0381

FAO. (2014). *Sustainable food value chains*. Food and Agricultural Organization of the United Nations. https://www.fao.org/sustainable-food-value-chains/what-is-it/en/

Hachimi, C. E., Belaqziz, S., Khabba, S., Sebbar, B., Dhiba, D., & Chehbouni, A. (2023). Smart weather data management based on artificial intelligence and big data analytics for precision agriculture. *Agriculture, 13*(1), 95. https://doi.org/10.3390/agriculture13010095

Hassoun, A., Jagtap, S., Garcia-Garcia, G., Trollman, H., Pateiro, M., Lorenzo, J. M., Trif, M., Rusu, A. V., Aadil, R. M., Šimat, V., Cropotova, J., & Câmara, J. S. (2023). Food quality 4.0: From traditional approaches to digitalized automated analysis. *Journal of Food Engineering, 337*, 111216. https://doi.org/10.1016/j.jfoodeng.2022.111216

Hassoun, A., Prieto, M. A., Carpena, M., Bouzembrak, Y., Marvin, H. J. P., Pallarés, N., Barba, F. J., Bangar, S. P., Chaudhary, V., Ibrahim, S., & Bono, G. (2022). Exploring the role of green and Industry 4.0 technologies in achieving sustainable development goals in food sectors. *Food Research International, 162*(B), 112068. https://doi.org/10.1016/j.foodres.2022.112068

Jonsdottir, B., Sigurjonsson, T. O., Johannsdottir, L., & Wendt, S. (2022). Barriers to using ESG data for investment decisions. *Sustainability, 14*, 5157. https://doi.org/10.3390/su14095157

Kamath, R. (2018). Food traceability on Blockchain: Walmart's pork and mango pilots with IBM. *Journal of the British Blockchain Association, 1*(1), https://doi.org/10.31585/jbba-1-1-(10)2018.

Karanth, S., Benefo, E. O., Patra, D., & Pradhan, A. K. (2023). Importance of artificial intelligence in evaluating climate change and food safety risk. *Journal of Agriculture and Food Research, 11*, 100485. https://doi.org/10.1016/j.jafr.2022.100485

Kler, R., Elkady, G., Rane, K., Singh, A., Hossain, M. S., Malhotra, D., Ray, S., & Bhatia, K. K. (2022). Machine learning and artificial intelligence in the food industry: A sustainable approach. *Journal of Food Quality, 1-9*, 1. https://doi.org/10.1155/2022/8521236

Kumar, I., Rawat, J., Mohd, N., & Husain, S. (2021). Opportunities of artificial intelligence and machine learning in the food industry. *Journal of Food Quality, 1-10*, 1. https://doi.org/10.1155/2021/4535567

Li, M., Jia, N., Lenzen, M., Malik, A., Wei, L., Jin, Y., & Raubenheimer, D. (2022). Global food-miles account for nearly 20% of total food-systems emissions. *Nature Food, 3*(6), 445–453. https://doi.org/10.1038/s43016-022-00531-w

Mahroof, K., Omar, A., & Kucukaltan, B. (2022). Sustainable food supply chains: Overcoming key challenges through digital technologies. *International Journal of Productivity and Performance Management, 71*(3), 981–1003. https://doi.org/10.1108/IJPPM-12-2020-0687

Manning, L., Brewer, S., Craigon, P. J., Frey, J., Gutierrez, A., Jacobs, N., Kanza, S., Munday, S., Sacks, J., & Pearson, S. (2022). Artificial intelligence and ethics within the food sector: Developing a common language for technology adoption across the supply chain. *Trends in Food Science & Technology, 125*, 33–42. https://doi.org/10.1016/j.tifs.2022.04.025

Marvin, H. J. P., Bouzembrak, Y., van der Fels-Klerx, H. J., Kempenaar, C., Veerkamp, R., Chauhan, A., Stroosnijder, S., Top, J., Simsek-Senel, G., Vrolijk, H., Knibbe, W. J., Zhang, L., Boom, R., & Tekinerdogan, B. (2022). Digitalisation and artificial intelligence for sustainable food systems. *Trends* *in Food Science & Technology, 120*, 344–348. https://doi.org/10.1016/j.tifs.2022.01.020

Mavani, N. R., Ali, J. M., Othman, S., Hussain, M. A., Hashim, H., & Rahman, N. A. (2022). Application of artificial intelligence in food industry—A guideline. *Food Engineering Reviews, 14*(1), 134–175. https://doi.org/10.1007/s12393-021-09290-z

Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2022). IoT, big data, and artificial intelligence in agriculture and food industry. *IEEE Internet of Things Journal, 9*(9), 6305–6324. https://doi.org/10.1109/JIOT.2020.2998584

Oehler, A., Horn, M., & Wendt, S. (2022). Investor characteristics and their impact on the decision to use a Robo-advisor. *Journal of Financial Services Research, 62*, 91–125. https://doi.org/10.1007/s10693-021-00367-8

Oehler, A., Schalkowski, H., & Wendt, S. (2013). Persönliche und automatisierte Kommunikation in KMU: Ein Ansatz zum Management gestiegener Komplexität von Kommunikationsprozessen. In J.-A. Meyer (Ed.), *Jahrbuch der KMU-Forschung und -Praxis 2013, Eul* (pp. 105–121).

Raghavendra, G. S., Mary, S. S. C., Acharjee, P. B., Varun, V. L., Bukhari, S. N. H., Dutta, C., & Samori, I. A. (2022). An empirical investigation in analysing the critical factors of artificial intelligence in influencing the food processing industry: A multivariate analysis of variance (MANOVA) approach. *Journal of Food Quality, 2022*, 1–7.

Ryan, M. (2022). The social and ethical impacts of artificial intelligence in agriculture: Mapping the agricultural AI literature. *AI & Society*. https://doi.org/10.1007/s00146-021-01377-9

Shamshiri, R. R., Kalantari, F., Ting, K. C., Thorp, K. R., Hameed, I. A., Weltzien, C., Ahmad, D., & Shad, Z. M. (2018). Advances in greenhouse automation and controlled environment agriculture: A transition to plant factories and urban agriculture. *International Journal of Agriculture & Biological Engineering, 11*(1), 1–22.

Sigurjonsson, T. O., & Wendt, S. (2023). Meiri sjálfbærni með gervigreind? [More sustainability with artificial intelligence?]. Working Paper, University of Iceland and Bifröst University.

Sturludottir, E., Thorvaldsson, G., Helgadottir, G., Gudnason, I., Sveinbjörnsson, J., Sigurgeirsson, Ó. I., & Sveinsson, T. (2021). Fæðuöryggi á Íslandi. Skýrsla unnin fyrir atvinnuvega- og nýsköpunarráðuneytið, Landbúnaðarháskóli Íslands. https://www.stjornarradid.is/library/01%2D%2DFrettatengt%2D%2D-myndir-og-skrar/ANR/

KThJ/F%C3%A6%C3%B0u%C3%B6ryggi%20%C3%A1%20%C3%8Dslandi%20lokask%C3%BDrsla.pdf
United Nations. (n.d.). *The 17 goals*. United Nations Department of Economic and Social Affairs. https://sdgs.un.org/goals
Vernier, C., Loeillet, D., Thomopoulos, R., & Macombe, C. (2021). Adoption of ICTs in Agri-food logistics: Potential and limitations for supply chain sustainability. *Sustainability, 13*(12), 6702. https://doi.org/10.3390/su13126702
Wagner, S. M. (2021). *Business and environmental sustainability*. Routledge. https://doi.org/10.4324/9781315208275
WRI & WBCSD. (2004). *The greenhouse gas protocol. World Resources Institute and World Business Council for Sustainable Development.*. https://ghgprotocol.org/sites/default/files/standards/ghg-protocol-revised.pdf
Zhou, Q., Zhang, H., & Wang, S. (2022). Artificial intelligence, big data, and blockchain in food safety. *International Journal of Food Engineering, 18*(1), 1–14. https://doi.org/10.1515/ijfe-2021-0299

Part III

The Role of AI in Sustainable Development and the 2030 Agenda

8

Analysis of Smart Meter Data for Energy Waste Management

Djordje Batic, Lina Stankovic, and Vladimir Stankovic

8.1 Introduction

When we think about artificial intelligence (AI), we often think of machines that have the potential to assist in or entirely replace human jobs: personal assistants such as Siri and Alexa; self-driving vehicles developed by Tesla, Mercedes, and Volvo; or perhaps humanoid robots such as those designed by Boston Dynamics. However, very few people realize that one of the more advanced uses of AI may be found in their utility room, closets, or cupboards. These are the places where one might often find a gas or electricity meter, sometimes a "smart meter", which is a small device that collects energy usage data, and, combined with the powerful predictive capabilities of state-of-the-art AI can not only enable accurate billing and encourage energy-efficient behavior, but better manage the electricity grid, power generation and distribution, and facilitate the transition to renewable energy sources and electric vehicles (Gellings, 2020).

D. Batic (✉) • L. Stankovic • V. Stankovic
University of Strathclyde, Glasgow, UK
e-mail: djordje.batic@strath.ac.uk; ina.stankovic@strath.ac.uk; vladimir.stankovic@strath.ac.uk

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_8

This energy transition, characterized by a shift from fossil fuels to renewable energy sources, is a complex and challenging process that requires the cooperation of multiple stakeholders and includes the integration of advanced technologies, innovative business models, effective policy frameworks, and people. The core technology that has the potential to unlock this better, brighter future and enable the integration of these elements is that of smart meters (Farhangi, 2010). Embedding smart metering as the interface between the smart grid and buildings (i.e., between energy generation and consumption) will play a pivotal role in the digitalization of the energy industry and is the key to sustainability and a successful energy transition.

Smart meters are devices that communicate high-frequency, real-time energy consumption (and production) information to utilities, consumers, and other stakeholders. Smart meters have become immensely important in recent years, as an increased number of energy consumers are also becoming producers (e.g., through solar power generation) (Parag & Sovacool, 2016). Furthermore, policies rolled out by many European countries promote the purchase of electric vehicles (EVs) and the installation of EV charging stations in homes, workplaces, and in places of social gathering. As a result, the ongoing transition is bound to cause a dramatic increase in electricity demand, exacerbating the need for the accurate metering of energy consumption and generation.

Smart metering is at the forefront of integration between the energy industry, AI, wireless network systems, and other forms of advanced information technologies. Additionally, the technology underpins the transition towards renewable energy sources and fulfilment of energy efficiency goals as set out by the Paris Agreement. However, the collection, processing, and interpretation of smart meter data has brought forward many challenges. As AI systems that use smart meters grow increasingly more complex, they bring about challenges related to robustness, privacy, and transparency (Kaselimi et al., 2022). Businesses providing feedback using AI-based smart meter analytics are ultimately accountable to their customers. As a result, "Trustworthy AI" has attracted significant interest within the research community and among businesses, policymakers, and individual consumers. The integration of new "trustworthy" AI technologies within systems that analyze smart meter data aims to create systems that are transparent, more accurate, better able to deal with malicious attempts at the destabilization of the system, as well as address issues related to data collection and privacy (European Commission, 2019).

Moreover, the increased rollout of smart meter technologies has the potential to tackle some of the issues we face with climate change. From the 1950s till 2018, the proportion of the global population living in urban areas has experienced a dramatic increase from 30% to 55%, and it is expected to grow to 68% by 2050 (United Nations Department of Economic and Social Affairs, 2018). Urbanization is one of the main drivers of economic activity and growth through the concentration of labor, capital resources, and production. However, the increased demand for goods and services, resulting from economic growth in industrialization and urbanization, has contributed to a rapid increase in the use of non-renewable resources, leading to environmental degradation. The rise in the exploitation of non-renewable resources is the leading cause of climate change, one of the most significant global issues (United Nations Department of Economic and Social Affairs, 2018). In particular, carbon dioxide (CO_2), a by-product of the use of fossil fuels for industry, forestry, and agriculture, constitutes the majority of anthropogenic greenhouse gas emissions and is considered one of the most damaging pollutants. The energy supply sector is the largest producer of CO_2 emissions and it is projected to remain so due to increased electricity use in buildings and industry (Bruckner et al., 2014).

Global warming represents a serious threat to environmental sustainability. Governments, international organizations, and businesses have joined forces to seek methods for reducing the negative environmental impact of the energy sector. Leveraging technological innovations, efforts have been concentrated on using low-emission energy sources, including renewables and nuclear power. Carbon neutrality goals, the implementation of sustainable policies, and increased solar and wind energy cost-effectiveness have contributed to significant changes in the energy sector. As a result, global renewable electricity capacity is projected to reach 4800 GW by 2026, a 60% increase from 2020 levels (IEA, 2021). Furthermore, advances in transportation systems, particularly the electrification of vehicles, have a great potential to reduce greenhouse gases and promote energy efficiency. However, achieving such a shift to a low-carbon economy while assuring energy security involves drastic changes in how energy is supplied and consumed.

This rest of this chapter is organized as follows. Section 8.2 will focus on exploring the sustainability benefits of smart grids and smart meters, drawing attention to how these technologies offer definitive solutions for energy management, particularly in a world increasingly dependent on renewable energy. While these advancements seem promising, they also come with unique challenges, particularly when incorporating AI. In Sect. 8.3, we delve into these challenges, focusing on the issues related to robustness, fairness, transparency, and privacy. A particular emphasis will be placed on Trustworthy AI for Non-Intrusive Load Monitoring (NILM) systems, a subtopic under Sect. 8.3. Lastly, the chapter will offer insights into the future of AI for smart meter data analytics, shedding light on the transformative potential of AI, its role in revolutionizing and decarbonizing the electricity grid, and the prospective challenges that lie ahead.

8.2 Sustainability Benefits of Smart Grids and Smart Meters

As the world moves to more renewable energy sources, the impending energy transition requires significant transformations in the energy sector. Not only will renewables cause major increases in electricity demand, but they will also cause substantial supply fluctuations (IEA, 2021). As a result, there are several challenges in the mass integration of renewables into the existing electricity grids, primarily due to supply variability. Solar photovoltaics (PVs) and wind turbines, the main drivers of renewable technologies, only operate when sunlight or wind is abundant. Because of this, the electricity grid is subjected to massive injections of PV or wind power during sunny or windy days and low inputs during cloudy or calm days, which poses a grid stability and renewable energy integration challenge.

Interestingly, electric vehicles (EVs) can absorb excess energy that would otherwise be wasted, improving the economic benefit of wind and solar power generation (Richardson, 2013). However, this can introduce major changes in the electricity demand, especially in locations with a large number of charging stations, where EVs may be charged at the same time during the day, leading to a very high load on the distribution grid and uncertainty in electricity supply meeting demand (Palensky & Dietrich, 2011). Another important element of modern grids is that an increasing number of energy consumers are becoming producers of energy, i.e., prosumers. As a result, the generation, transmission, distribution, and control operations of the traditional grid need to be able to accommodate the dramatic changes caused by the transition towards renewable energy and the electrification of transport (Parag & Sovacool, 2016). This need has led to the emergence of Smart Grid (SG) systems that aim to facilitate operational efficiency and reliability of the grid using advances in infrastructure, intelligent information collection, automation and knowledge extraction (Fang et al., 2012). SGs enable the decentralization of distribution and communication and are at the forefront of efforts geared towards addressing operational complexity introduced by the increased utilization of renewable energy. It aims to deliver power more efficiently and automatically address any events that may impact the quality or reliability of the power supply and generation (Gellings, 2020).

Despite the apparent benefits, it is important to note that in some instances, energy efficiency efforts lead to rebound effects (a phenomenon where energy efficiency can lead to an increase in the consumption of energy services). Consequently, rebound effects could offset the expected benefits of SGs, where lower energy costs induced by them may stimulate additional energy use. In the context of rebound mitigation, one promising initiative to boost environmental consciousness is raising awareness of households' energy consumption behavior (Font Vivanco et al., 2016). Knowledge extracted from electricity use behavior can be used to better understand the changing needs of consumers and prosumers alike and make sense of complex consumption patterns, and the core technology that enables this is smart metering (Fan et al., 2013).

Successful operation of SG infrastructure depends on the ability to collect and extract knowledge from live load measurements that will enable better monitoring of supply and demand, reduce operational costs, and optimize energy efficiency. Smart meters enable real-time consumption feedback that is communicated remotely to utilities, consumers, prosumers, and other stakeholders with an interest in meeting energy efficiency targets. Unlike traditional meters, which usually provide consumption feedback only once a month, smart meters can provide feedback in real-time. Smart meters represent the core pillar of the smart grid and are the key to the successful realization of future energy management systems that make informed decisions for consumers, electricity producers, and network operators alike. Smart grid technology is essential in lowering carbon emission goals, as it facilitates the broader incorporation of renewable energy sources like solar and wind power. This incorporation enables small-scale electricity production, enhances supply and demand flexibility, assures accurate customer billing, and promotes the decentralization of power generation (Sovacool et al., 2017).

Smart metering deployment has been observed through large-scale national implementations (Zhou & Brown, 2017) and smaller experimental ventures predominantly seen in Europe, Canada, and the USA. Moreover, smart metering benefits multiple actors in the smart grid system (Depuru et al., 2011). Energy consumers benefit from more accurate billing and increased awareness of their energy usage patterns, which can promote energy efficiency behavior. Additionally, smart metering provides further savings in cases of variable tariffs, where users can shift the time of use of certain appliances and get rewarded with lower electricity costs or bonuses. Furthermore, it can lead to a higher quality of service due to increased competition among suppliers who can offer more personalized contracts and various other incentives. Finally, utility companies benefit from smart meters due to an increased understanding of their customer needs, which can drive customized pricing and lower costs due to eliminating the need for physical meter readings.

The imminent shift from conventional electricity networks towards a smart grid is expected to bring various opportunities and challenges, and shift towards predominantly low-carbon energy and distribution across the energy generation system is one of the foremost challenges. To balance the inherent variability and unpredictability associated with renewable energy sources such as wind and solar power, proactive operational measures will be vital for maintaining the stability of the system and a balance between production and demand, ensuring the adherence of generation outputs and network transfers to capacity constraints and mandatory security margins. Importantly, demand will be a controllable aspect as end-users transition from passive consumers to active participants in both system and market operations. In this context, smart appliances, electric heaters, and batteries will offer significant demand-side flexibility. In contrast, the electrification of other energy sectors, such as heating/cooling (e.g., via electric heat pumps) and transport (e.g., electric vehicles), is set to broaden further the scope for load control and synchronization with overall system requirements.

The issue of understanding consumption patterns through the analysis of metered data has long been a point of discussion in the scientific community. One of the first implementations of knowledge extraction from metered data of total energy consumption was proposed by (Hart, 1992), who derived a technique that enables the estimation of power usage of individual appliances, given only the total electricity consumption. This technique is known as non-intrusive load monitoring (NILM). The core principle of NILM is modeled as a source separation problem. It aims to understand individual appliances' consumption by modeling the typical energy consumption signals of electrical appliances and extracting them from the total electricity load. This technique effectively enables the "virtual" monitoring of electricity consumption of individual appliances, such as washing machines, refrigerators, kettles, etc., without the need to install separate meters on each appliance. Since the deployment of smart grids, NILM has received an uptake in interest from businesses and researchers due to its ability to understand consumption profiles and the potential to be integrated with demand response programs aimed at achieving energy sustainability and increased efficiency (Zeifman & Roth, 2011). Time and frequency of use of particular appliances, including EV charging, are critical in meeting demand and managing energy consumption and production. Over the years, many algorithmic approaches have been proposed in NILM; however, the most successful and impactful methods have come from the advances in AI and Deep Learning (DL) over the last decade.

8.3 Challenges of Using AI for Smart Meter Data Analysis

The technological progress introduced by the emergence of DL-based AI has been one of the most impressive breakthroughs of the last decade. Historically, significant technological breakthroughs, such as agricultural, industrial, and computational revolutions, have been followed by immense improvements in production and efficiency, lower prices and increased quality of goods, increased wealth, and an overall increase in societal prosperity. Looking ahead, the scale of possibilities introduced by the successful integration of AI has the potential to bring forth a new "AI revolution". AI systems currently in use can already solve complex problems and offer a wide range of valuable services. As a result, it is no surprise that there have been many efforts to apply DL principles to a wide range of problems associated with meeting energy efficiency goals, where NILM is one of the most interesting challenges. The majority of modern NILM systems operate on the backbone of neural networks (NNs). NN represents a system of interconnected artificial neurons that are coherently organized and may be used to discover patterns in a large body of data by an iterative learning process (LeCun et al., 2015). NILM researchers and businesses that incorporate NILM have employed NN architectures to successfully detect signatures of a wide variety of appliances in many countries across the globe, which is a complex task due to variations in climate, building quality standards, and types of appliances commonly used in different areas. However, despite their excellent predictive performance and potential, AI applications have several ethical pitfalls hindering large-scale adoption.

AI algorithms are able to extracting knowledge from abundant data sources, including images, text, and sequential data. Widespread adoption in finance, transportation, entertainment, medicine, social media, and autonomous vehicles has shown that AI can lead to increased economic benefits. However, this rapid integration of AI into the everyday lives of citizens has led to questions about the pervasiveness of AI (Rahwan et al., 2019). This issue is particularly important for application domains such as smart homes, where AI algorithms analyze highly sensitive and personal data. For example, smart meter data can be used to infer the daily routines of a household (Efthymiou & Kalogridis, 2010). While this information is intended to provide energy-saving recommendations, it also has the potential to reveal the social habits of the people in the households, such as when a family member is cooking, exercising, doing laundry, or performing other household activities (Stankovic & Stankovic, 2020).

Consequently, with the help of AI, smart meter data can be used to determine when someone is at home, their sleeping patterns, and household activities. Due to these issues, introducing smart meters has raised serious concerns from the public, governments, regulators and other concerned individuals and organizations. In France, since the introduction of smart meters in 2015, there have been a number of incidents of people refusing installation of smart meters due to concerns centered around privacy, which is a trend reported in many other European countries, such as Netherlands, Canada and USA (Vigurs et al., 2021). As a result, to establish trust in the AI technologies used for smart meter data analysis, it is of utmost importance to ensure that the algorithms involved prevent infringement of user privacy and security and offer transparency of the predictive results to encourage user acceptance.

Additionally, developments in algorithmic approaches in AI have led to improved performance in knowledge extraction tasks. Recent NILM surveys (Angelis et al., 2022; Huber et al., 2021) suggest that the performance of AI models achieves the best predictive performance across a wide range of real-world and synthetic dataset scenarios. However, the rapid growth of AI has led to some emerging concerns surrounding the use of AI models. Particularly, on the fact that focusing solely on predictive accuracy metrics when evaluating the quality of AI (often the case in many real-world applications) neglects other important factors, such as the trustworthiness of the AI system. "Trustworthy AI" refers to AI systems that are reliable, unbiased, and safe (European Commission, 2019). Correspondingly, Trustworthy AI for NILM would imply that the algorithms trained on smart meter data obtain good predictive performance but are also designed to emphasize other important properties such as robustness, fairness, transparency, and privacy. Accounting for these properties is an essential step towards wider adoption of AI and an opening for a new, mature era of AI design and deployment. Furthermore, it is important that these issues become public knowledge, as public awareness will motivate discussion but also drive innovation. Furthermore, it can help open the marketplace to new start-ups centered around realizing Trustworthy AI design and deployment principles.

A. Trustworthy AI for NILM

The remainder of this chapter will focus on presenting approaches that account for each of the core properties of Trustworthy AI for NILM. However, it is important to retain a holistic view, as a breach of trust in any single property can undermine the trust level of the whole system. Furthermore, it is vital to understand that the pursuit of trustworthiness should include awareness of the interactions between the properties of the trustworthy system, which is particularly important when requirements interfere with each other. For example, by ensuring strong data privacy guarantees, the predictive performance of the system may suffer. Additionally, if input data is significantly altered, it might negatively affect the system's transparency. Thus, even though there are significant possibilities for progress in individual instances of trustworthiness, it is equally important to understand the requirements and interactions of each aspect of the trustworthy system.

The robustness of an AI system is reflected in its ability to deal with unseen, noisy, or erroneous data (Kaselimi et al., 2022). Hence, robustness is a property that strongly affects the system's performance in real-world scenarios. Systems that lack robustness may show unpredictable behavior or cause harm to the user and consequently diminish the system's trustworthiness. This trust in the system that a user has is especially important for applications that analyze personal data or have a direct impact on the financial security of users. In the context of NILM, both are true, as smart meter data contains information about consumption patterns and enables billing and energy savings benefits. As a way of addressing such issues, there has been a great deal of work done to evaluate the robustness of AI models, and the proposed techniques may be categorized based on the way they approach the problem: at the data level, through the algorithm or via evaluation metrics. Significant attention has been devoted to evaluating disaggregation performance on unseen data, where unseen data represents data that the AI model has not been trained on. Given a model trained on smart meter data from houses in a distinct dataset, there are two problems that researchers and businesses have identified in such unseen scenarios: generalizability, i.e., the ability to perform well on unseen smart meter data from houses in the same dataset, and cross-domain transferability, i.e., the ability to perform well on unseen smart meter data from houses in other datasets. The distinction between datasets is essential because they may contain information on different types of appliances, and also on different consumption habits, locations, tariffs, household income, etc.

As large amounts of data are often required for the training of AI models, the most critical factors in terms of data quality that can negatively affect robustness are a low sampling rate and an insufficient number of houses in the dataset (Shin et al., 2019). These issues can lead to poor predictive performance in many real-world scenarios due to insufficient data or datasets with imbalanced classes. Thus, one of the ways to increase the data quality and alleviate such problems is to use data augmentation techniques that aim to enlarge the training dataset with high-quality data. Kong et al. (2020) use target appliance activations collected from customers in the database and randomly insert them into the aggregate data profile. In this way, the size of the dataset is significantly enriched with realistic data samples, which is crucial for the robustness of the AI model. On the other hand, Harell et al. (2021) propose a technique based on a Generative Adversarial Network (GAN). A GAN can be considered a NN that aims to create replicas of the data used to train it. Imagine that there is a pianist who seeks to recreate the sound of Artur Rubinstein playing Chopin. Unfortunately, he has never heard any Chopin compositions and only has access to a fellow musician who knows precisely what Rubinstein playing Chopin sounds like. The fellow musician lets his friend play the piano, and every time he sounds similar to Rubinstein gives him a reward. Thus, the musician learns to play Chopin like Rubinstein. Based on this example, GANs used for NILM aims to synthesize realistic appliance power signatures by playing a similar game between imitating the existing data in the dataset.

To determine the performance of a NILM algorithm, Klemenjak et al. (2019) propose various evaluation metrics that quantify the relative difference in predictive performance on seen and unseen houses in the dataset. Similarly, Rafiq et al. (2021) propose an energy-based evaluation metric to quantify the predicted, overlapping, missing, and extra energy percentage. As a way of advancing the architectural design of robust NILM networks, Murray et al. (2019) developed Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) based architectures and conducted experiments to show that their algorithmic approach can achieve favorable generalizability performance across multiple public datasets.

One of the often-overlooked aspects of AI robustness in NILM is the ability of the system to withstand adversarial attacks. Adversarial attacks imply that AI models can be exposed by malicious imputation of small and often unnoticeable changes in the training data. As the use of AI with smart meter data is one of the core elements of smart grids, AI models are vulnerable to adversarial data attacks that aim to compromise the system by injecting data to purposefully cause negative effects on the performance (Goodfellow et al., 2015). Wang and Srikantha (2021) stress the importance of more research in the area of adversarial machine learning for smart grids and show that by training a substitute AI model, it is possible to steer the original model (used in the smart grid) towards incorrect predictions, without any knowledge of original model architecture or training data. One approach that can alleviate such a problem is anomaly detection, or detection of outliers, which is critically important in systems subjected to large amounts of fraudulent data.

Despite being at the forefront of energy efficiency efforts, smart meters have the potential to compromise user privacy. Even though the integration of traditional meters has already raised privacy and security concerns, smart meters have exacerbated this problem due to the high granularity of data collection, as well as the bidirectional, wireless nature of their communication. Many everyday household activities such as washing the clothes or dishes, cooking, heating the boiler, or watching TV have distinct and detectable power consumption signatures, which leads to the ability of AI tools to predict what sort of activities family members are performing at any point in time (McDaniel & McLaughlin, 2009). This information can be used maliciously, such as targeted advertising or burglary. One of the most critical aspects of Trustworthy AI for smart meter analysis systems is deploying AI models that do not infringe on users' privacy. Privacy violation risks encompass the full lifecycle of the AI system. They can be categorized as follows: unauthorized access to raw data through data breaches, unauthorized extraction of sensitive information through AI model inference, and unauthorized access to the AI model itself (Brundage et al., 2020). All of these risks are relevant to smart metering systems and prevention of such risks has attracted researchers and businesses to develop various techniques for privacy protection. Privacy concerns can be addressed during data pre-processing, during model training, or during deployment.

First, AI NILM systems can address privacy risks with techniques for the privacy protection of data used for training. One of the most common approaches in privacy preservation is data anonymization. Data anonymization aims to address privacy concerns by modifying the data so that the knowledge contained in the private data cannot be attached to a person. Popular algorithms include "k-anonymity", which aims to ensure that each user in the anonymized dataset cannot be identified within a set of "k" number of users. Other common approaches include random noise injection, data transformation, and differential privacy. Backes and Meiser (2014) examine a way to achieve different levels of privacy through the use of rechargeable batteries for noise injection, while Cao et al. (2018) propose an approach based on differentially private Fog computing. Fog computing is a system of interconnected edge devices that provide a substantial amount of communication, storage and computation over an internet-based service. The authors use the Fog computing backbone for NILM and introduce differential privacy by obfuscating user behavior through the addition of noise to switch states of appliances. Differential privacy (DP) is one of the most widely adopted frameworks for data privacy, and leading technology companies such as Google, Meta, and Apple all have versions of the DP framework. The core mechanism of DP is the notion of a "privacy budget", which quantifies the greatest possible information gain in case of malicious data access.

Another way privacy preservation can be achieved is during model training. The most successful approach used for this is Federated Learning

(FL). Zhang et al. (2022) propose an FL approach for NILM, where personalized models can be trained privately and locally, without collecting user data from the main server (i.e., utility company). As opposed to the traditional AI learning paradigm that requires the collection and storage of data on a central server where the model is subsequently trained, FL is a framework that enables edge devices (e.g., smart meter) to collaboratively train an AI model without the need for sharing of user data to the central server. In every iteration, the central server collects the parameters of AI models trained on each edge device and aggregates them to update the global model. Furthermore, FL also enables personalization through options to customize the global model to improve the performance on individual edge devices. It is important to note that FL facilitates user privacy because there is no communication of smart meter data to the central server; instead, only AI model parameters are shared.

Finally, user privacy can be preserved during the deployment of AI models. Hardware attacks represent important risks, as the data can be subjected to malicious purposes. As a result, various approaches have been proposed, and one of the most widely adopted hardware security protocols for the privacy of user data that can be applied to data collected from smart meters is Trusted Execution Environment (TEE) (Karopoulos et al., 2017). TEE protocols support data encryption and secure storage, while applications and software executed in a protected environment prevent unauthorized access, tampering, or reverse engineering of the security mechanism.

The rapid success of modern AI algorithms, such as deep neural networks (DNNs), also presents enormous challenges related to system transparency. An increasing number of governments, researchers, businesses and users have voiced concerns over the lack of transparency in AI algorithms. The inherent complexity of DNNs makes understanding the process behind the computation of their decisions very difficult, which is why they are often called “black box” algorithms. As an example of lack of transparency, difficulty in explaining the mechanisms behind how Netflix recommends a movie to watch might not be worrying; however, transparency of systems that analyze personal data such as personal activity, health, and gender, as well as systems that have a direct impact on the financial security of users is a much more serious concern. The link between transparency and trust has long been an active area of research interest, and studies suggest that systems that do not display physical characteristics (e.g., robot) or have no identity (e.g., chatbot or avatar) rely primarily on aspects of robustness and transparency for establishing trust (Glikson & Woolley, 2020). Hence, in recent years, there has been a strong push toward developing practical tools that enable a better understanding of how AI models arrive at their decisions. These tools can broadly be categorized under the field of "explainable AI" (XAI). XAI aims to develop techniques for deriving understandable human explanations from specific predictions or decisions of an AI algorithm. XAI techniques have been shown to facilitate a better understanding of AI systems and increase trustworthiness. However, adoption in certain fields has been limited. This limited adoption is especially true of smart grid systems, including NILM applications.

The first approach by Murray et al. (2021) provides analysis of predictions of a typical NILM DNN by generating visual heatmaps that point to areas of the highest importance and by occluding certain parts of the signal and then examining the impact on the derived prediction. The analysis provides a solution that helps in better understanding the AI-based NILM system. Machlev et al. (2022) propose an XAI methodology for the binary classification of appliances in an aggregate signal window and subsequent visualization of relevant parts of the signal for the classified appliance. Finally, Batic et al. (2023) propose a modification to the training mechanism of certain types of AI models used for deployment on edge devices that utilizes XAI mechanisms to improve the model's performance. Models are trained using the principle of knowledge distillation, where the behavior of a large and complex AI model is used to train a simpler model; one that is suitable for use on edge devices with limiting computational or storage capacity (like smart meters). Despite the importance of transparency in the context of Trustworthy AI for smart meter data analysis, the current scope of research is still in its infancy. To deliver better service and foster trust, researchers and industries in this sector need to actively work to promote the establishment of a transparent system design.

8.4 The Future of AI for Smart Meter Data Analytics

The rapid advancement of AI technologies, coupled with developments in how energy is generated, supplied, and consumed, has brought various products and services that aim to facilitate a greener and more sustainable future. This intersection of AI and smart grid technologies has united businesses, researchers, and policymakers in the critical goal of reducing the harmful effects of climate change and has enormous potential for growth in the coming decades. However, despite the transformative impact of AI in the smart grid, the important issue of the trustworthiness of AI remains. Smart meter analytics, usually customer-facing, are on the leading edge of AI integration in the smart grid, and NILM presents an important application of AI using smart meter data. Looking ahead, developments in the trustworthy aspects of AI used for NILM have promising potential for further research and implementation.

First, the robustness of AI can be achieved by incorporating various domain-agnostic and domain-specific evaluation metrics, as well as developing models that can deal with the issues of generalizability. Another essential requirement of smart metering technology is its ability to deal with adversarial attacks, which is a domain that has so far seen limited interest, despite its importance. Second, AI used for smart meter analytics needs to provide privacy guarantees. The most promising direction is the further development of federated learning (FL) approaches to ensure algorithmic personalization to improve service and benefit users. Third, AI used on smart meter data needs to ensure transparency by providing explanations for its derived predictions. With the rapid development of the field of XAI, available techniques are abundant for generating visual and humanly understandable explanations. However, there is a significant limitation in the reliability of such techniques, which calls for a quantitative evaluation of their quality. Providing metrics suited for smart meter data types might be one of the most important elements for safely integrating XAI techniques in transparent smart grid systems. Finally, the relationship between robustness, privacy and transparency of AI used on smart meter data needs to be carefully examined, and businesses interested in providing such services need to address how they can be combined and successfully implemented to create an efficient, sustainable, and trustworthy system.

In conclusion, the potential for smart meters to drive energy efficiency and reduce emissions, combined with the application of AI in the smart grid, underscores the critical role of innovations in the fight against climate change. However, fostering trust in these technologies among consumers is a prerequisite for their widespread adoption and success. Implementing safeguards for system robustness, ensuring privacy, and providing transparent and understandable AI operations are integral for building this trust. By having an inclusive approach and placing the consumer at the center of this AI transformation, we can harness the full potential of smart meters and AI to pave the way towards a long-term sustainable practices and climate-resilient future.

References

Angelis, G.-F., Timplalexis, C., Krinidis, S., Ioannidis, D., & Tzovaras, D. (2022). NILM applications: Literature review of learning approaches, recent developments and challenges. *Energy and Buildings, 261*, 111951. https://doi.org/10.1016/j.enbuild.2022.111951

Backes, M., & Meiser, S. (2014). Differentially private smart metering with battery recharging. In J. Garcia-Alfaro, G. Lioudakis, N. Cuppens-Boulahia, S. Foley, & W. M. Fitzgerald (Eds.), *Data privacy management and autonomous spontaneous security* (pp. 194–212). Springer. https://doi.org/10.1007/978-3-642-54568-9_13

Batic, D., Tanoni, G., Stankovic, L., Stankovic, V., & Principi, E. (2023). Improving knowledge distillation for non-intrusive load monitoring through explainability guided learning. *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023)*.

Bruckner, T., Bashmakov, I., Mulugetta, Y., Chum, H., Navarro, A., Edmonds, J., Faaij, A. P. C., Fungtammasan, B., Garg, A., Hertwich, E., Honnery, D., Infield, D., Kainuma, M., Khennas, S., Kim, S., Nimir, H., Riahi, K., Strachan, N., Wiser, R., & Upadhyay, J. (2014). Energy systems. *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group*

III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 511–598). https://doi.org/10.1017/CBO9781107415416.013.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., Yang, J., Toner, H., Fong, R., Maharaj, T., Koh, P. W., Hooker, S., Leung, J., Trask, A., Bluemke, E., Lebensold, J., O'Keefe, C., Koren, M., … Anderljung, M. (2020). *Toward trustworthy AI development: Mechanisms for supporting verifiable claims* (arXiv:2004.07213). arXiv. https://doi.org/10.48550/arXiv.2004.07213.

Cao, H., Liu, S., Wu, L., Guan, Z., & Du, X. (2018). *Achieving differential privacy against non-intrusive load monitoring in smart grid: A fog computing approach* (arXiv:1804.01817). arXiv. http://arxiv.org/abs/1804.01817

Depuru, S. S. S. R., Wang, L., & Devabhaktuni, V. (2011). Smart meters for power grid: Challenges, issues, advantages and status. *Renewable and Sustainable Energy Reviews, 15*(6), 2736–2742. https://doi.org/10.1016/j.rser.2011.02.039

Efthymiou, C., & Kalogridis, G. (2010). *Smart grid privacy via anonymization of smart metering data* (pp. 238–243). 2010 First IEEE International Conference on Smart Grid Communications. https://doi.org/10.1109/SMARTGRID.2010.5622050

European Commission. (2019). *Ethics guidelines for trustworthy AI*. Publications Office. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Fan, Z., Kulkarni, P., Gormus, S., Efthymiou, C., Kalogridis, G., Sooriyabandara, M., Zhu, Z., Lambotharan, S., & Chin, W. H. (2013). Smart grid communications: overview of research challenges, solutions, and standardization activities. *IEEE Communications Surveys & Tutorials, 15*(1), 21–38. https://doi.org/10.1109/SURV.2011.122211.00021

Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Smart grid — The new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials, 14*(4), 944–980. https://doi.org/10.1109/SURV.2011.101911.00087

Farhangi, H. (2010). The path of the smart grid. *IEEE Power and Energy Magazine, 8*(1), 18–28. https://doi.org/10.1109/MPE.2009.934876

Font Vivanco, D., Kemp, R., & van der Voet, E. (2016). How to deal with the rebound effect? A policy-oriented approach. *Energy Policy, 94*, 114–125. https://doi.org/10.1016/j.enpol.2016.03.054

Gellings, C. W. (2020). The smart grid: Enabling energy efficiency and demand response. *River Publishers.* https://doi.org/10.1201/9781003151524

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). *Explaining and harnessing adversarial examples* (arXiv:1412.6572). arXiv. http://arxiv.org/abs/1412.6572

Harell, A., Jones, R., Makonin, S., & Bajić, I. V. (2021). TraceGAN: Synthesizing appliance power signatures using generative adversarial networks. *IEEE Transactions on Smart Grid, 12*(5), 4553–4563. https://doi.org/10.1109/TSG.2021.3078695

Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE, 80*(12), 1870–1891. https://doi.org/10.1109/5.192069

Huber, P., Calatroni, A., Rumsch, A., & Paice, A. (2021). Review on deep neural networks applied to low-frequency NILM. *Energies, 14*(9) Article 9. https://doi.org/10.3390/en14092390

IEA. (2021). *Renewables 2021*. IEA.

Karopoulos, G., Xenakis, C., Tennina, S., & Evangelopoulos, S. (2017). Towards trusted metering in the smart grid. *2017 IEEE 22nd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 1–5. https://doi.org/10.1109/CAMAD.2017.8031643.

Kaselimi, M., Protopapadakis, E., Voulodimos, A., Doulamis, N., & Doulamis, A. (2022). Towards trustworthy energy disaggregation: A review of challenges, methods, and perspectives for non-intrusive load monitoring. *Sensors, 22*(15) Article 15. https://doi.org/10.3390/s22155872

Klemenjak, C., Faustine, A., Makonin, S., & Elmenreich, W. (2019). *On metrics to assess the transferability of machine learning models in non-intrusive load monitoring* (arXiv:1912.06200). arXiv. http://arxiv.org/abs/1912.06200

Kong, W., Dong, Z. Y., Wang, B., Zhao, J., & Huang, J. (2020). A practical solution for non-intrusive type II load monitoring based on deep learning and post-processing. *IEEE Transactions on Smart Grid, 11*(1), 148–160. https://doi.org/10.1109/TSG.2019.2918330

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

Machlev, R., Malka, A., Perl, M., Levron, Y., & Belikov, J. (2022). *Explaining the decisions of deep learning models for load disaggregation (NILM) based on XAI* (pp. 1–5). 2022 IEEE Power & Energy Society General Meeting (PESGM). https://doi.org/10.1109/PESGM48719.2022.9917049

McDaniel, P., & McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *IEEE Security & Privacy, 7*(3), 75–77. https://doi.org/10.1109/MSP.2009.76

Murray, D., Stankovic, L., Stankovic, V., Lulic, S., & Sladojevic, S. (2019). *Transferability of neural network approaches for low-rate energy disaggregation* (pp. 8330–8334). ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://doi.org/10.1109/ICASSP.2019.8682486

Murray, D., Stankovic, L., & Stankovic, V. (2021). *Transparent AI: Explainability of deep learning based load disaggregation* (pp. 268–271). Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. https://doi.org/10.1145/3486611.3492410

Palensky, P., & Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE Transactions on Industrial Informatics, 7*(3), 381–388. https://doi.org/10.1109/TII.2011.2158841

Parag, Y., & Sovacool, B. K. (2016). Electricity market design for the prosumer era. *Nature Energy, 1*(4) Article 4. https://doi.org/10.1038/nenergy.2016.32

Rafiq, H., Shi, X., Zhang, H., Li, H., Ochani, M. K., & Shah, A. A. (2021). Generalizability improvement of deep learning-based non-intrusive load monitoring system using data augmentation. *IEEE Transactions on Smart Grid, 12*(4), 3265–3277. https://doi.org/10.1109/TSG.2021.3082622

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A., et al. (2019). Machine behaviour. *Nature, 568*(7753), 477. https://doi.org/10.1038/s41586-019-1138-y

Richardson, D. B. (2013). Electric vehicles and the electric grid: A review of modeling approaches, impacts, and renewable energy integration. *Renewable and Sustainable Energy Reviews, 19*, 247–254. https://doi.org/10.1016/j.rser.2012.11.042

Shin, C., Rho, S., Lee, H., & Rhee, W. (2019). Data requirements for applying machine learning to energy disaggregation. *Energies, 12*(9) Article 9. https://doi.org/10.3390/en12091696

Sovacool, B. K., Kivimaa, P., Hielscher, S., & Jenkins, K. (2017). Vulnerability and resistance in the United Kingdom's smart meter transition. *Energy Policy, 109*, 767–781. https://doi.org/10.1016/j.enpol.2017.07.037

Stankovic, L., & Stankovic, V. (2020). *The risks and benefits of AI smart meters*. https://apolitical.co/en/solution_article/the-risks-and-benefits-of-ai-smart-meters

United Nations Department of Economic and Social Affairs. (2018). *2018 revision of world urbanization prospects*. United Nations Department of Economic and Social Affairs.

Vigurs, C., Maidment, C., Fell, M., & Shipworth, D. (2021). Customer privacy concerns as a barrier to sharing data about energy use in smart local energy systems: A rapid realist review. *Energies, 14*(5) Article 5. https://doi.org/10.3390/en14051285

Wang, J., & Srikantha, P. (2021). Stealthy black-box attacks on deep learning non-intrusive load monitoring models. *IEEE Transactions on Smart Grid, 12*(4), 3479–3492. https://doi.org/10.1109/TSG.2021.3062722

Zeifman, M., & Roth, K. (2011). Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics, 57*(1), 76–84. https://doi.org/10.1109/TCE.2011.5735484

Zhang, Y., Tang, G., Huang, Q., Wang, Y., Wu, K., Yu, K., & Shao, X. (2022). FedNILM: Applying federated learning to NILM applications at the edge. *IEEE Transactions on Green Communications and Networking, 1*, 857. https://doi.org/10.1109/TGCN.2022.3167392

Zhou, S., & Brown, M. A. (2017). Smart meter deployment in Europe: A comparative case study on the impacts of national policy schemes. *Journal of Cleaner Production, 144*, 22–32. https://doi.org/10.1016/j.jclepro.2016.12.031

9

Leveraging AI to Map SDG Coverage and Uncover Partnerships in Swiss Philanthropy

Maria Cristiana Tudor, Lucia Gomez,
Camilla Della Giovampaola, Hubert Halopé,
and Giuseppe Ugazio

9.1 Overview

In 2015, all United Nations (UN) member states ratified the Sustainable Development Goals (SDGs), which define the core issues to be addressed and targets to be achieved by 2030 to ensure a sustainable future (UN, 2015). The recent rise in armed conflicts, increasing inequalities within and between countries, extreme weather events, and the COVID-19 pandemic significantly impacted SDGs' advancements. For example, it is

M. C. Tudor (✉) • L. Gomez • H. Halopé • G. Ugazio
Geneva Finance Research Institute, University of Geneva, Geneva, Switzerland
e-mail: m.tudor@alumni.ucl.ac.uk

C. D. Giovampaola
Geneva Finance Research Institute, University of Geneva, Geneva, Switzerland
The Graduate Institute, Geneva, Switzerland

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_9

estimated that the pandemic caused four years of regression on SDG 1 – No Poverty, pushing 93 million people back into extreme poverty (UN, 2022). Such setbacks strongly challenge the possibility of SDG attainment by 2030, highlighting the need for a faster and more resilient recovery (*Nature*, 2020).

Novel technological resources can provide unique means for a more innovative and efficient approach to accelerating SDG advancements. Among others, Artificial Intelligence (AI) presents powerful tools to activate solutions. Focusing on SDG 17 – Partnership for the Goals, this chapter applies Natural Language Processing (NLP) and network mapping tools to uncover areas of untapped partnership potential in the Swiss nonprofit sector (see Fig. 9.1). We first present an overview of the SDG framework and the value of partnerships, followed by an outline of AI's role in the SDG landscape. We then illustrate philanthropy's crucial role in this context. The following section integrates these subjects by introducing an NLP-based network mapping tool for uncovering SDG partnerships potential in the Swiss nonprofit ecosystem. Lastly, we discuss some of the general risks of incorporating AI in philanthropic organizations[1] and for SDGs.

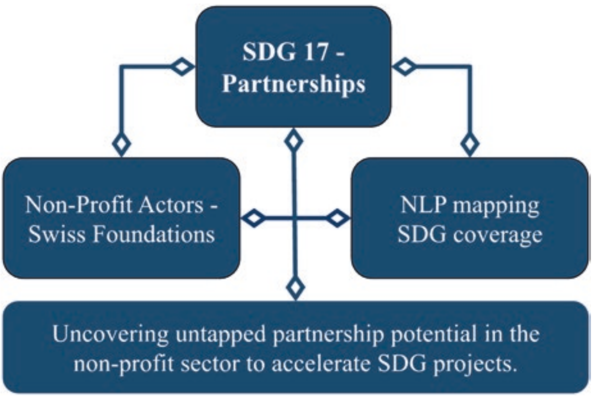

Fig. 9.1 Chapter overview

[1] Interchangeably referred to as Non-Profits (NPs) throughout the chapter.

9.2 Contextualizing Sustainability – The SDG Framework

A. A Mindful Approach to SDG Implementation

Despite being non-binding, the SDGs showcase a cooperative effort to acknowledge systemic problems while promoting global awareness. As of 2022, public awareness of the SDGs stands around 50% globally (Kleespies & Dierkes, 2022), a stark increase from the 28% identified shortly after their launch in 2016 (Miller, 2016). In the private sector, 83% of companies express their support for SDGs, while only 40% set measurable targets for their contribution (GRI, 2022). These numbers provide encouraging evidence of the SDGs' utility as a unifying structure to enhance awareness and subsequently streamline efforts and equally reveal the remaining gap to be addressed between expressed support and measurable targets.

However, the framework simplifies a highly complex system, leaving room for inconsistencies (Neubauer & Calame, 2017). Several core concerns are worth mentioning to remain aware of shortcomings and inform a mindful approach to implementation. Specifically, the economic growth-oriented perspective (e.g., SDG 8 – Decent Work and Economic Growth) can propel the same unsustainable systems driving overconsumption and exploitation that other SDGs (e.g., SDG 12 – Responsible Consumption and Production) are required to tackle (Hickel, 2019; Hickel & Kallis, 2020). Moreover, there are concerns on whether green growth is indeed possible (Hickel, 2019), suggesting the relevance of considering other models, e.g., degrowth models[2] and circular economy[3] (Belmonte-Ureña et al., 2021; Schröder et al., 2019). Secondly, it could be argued that the SDGs represent a top-down approach to sustainability, neglecting unique local needs. It is necessary to balance top-down and

[2] Degrowth models refer to focusing on shrinking some aspects of the economy (e.g., excessive consumerism), as to reduce energy and resource use.

[3] A circular economy aims to incentivize reusing, repurposing, and recycling products to extend their utility and lifespan, which reduces overall energy and resource use as it avoids continuous production of novel products. This contrasts with a linear economy approach, which places less emphasis on end-of-life repurposing.

bottom-up approaches, for example, by carrying out field research to specifically address local needs, regardless of the context (i.e., high- or low-income countries). Additional concerns revolve around the validity of the SDGs' metrics, as inconsistencies in data collection and definitions of indicators remain prevalent (Mair et al., 2018; Swain, 2018).[4]

Nonetheless, the introduction of the SDGs has led to significant advancements and success stories. For instance, in the context of SDG 17 – Partnership for the Goals, UN Human Rights, in partnership with the Danish social enterprise Specialisterne, developed an AI algorithm to enhance human rights analysis (OHCHR, 2021). This resulted in the Universal Human Rights Index,[5] a platform providing detailed information on regional human rights issues and their linkages to the 169 targets of the 2030 Agenda, which allows actors to find concrete and locally specific solutions for SDG projects. This example showcases both the utility of AI and a successful partnership within the SDG framework.[6]

B. The Synergistic Value of Partnerships

Multi-stakeholder partnerships represent a key resource for effective SDGs implementation and resilient post-pandemic recovery. This is captured by SDG 17, the UN acknowledging that "[t]op-down, short-term, single-sector approaches generally cannot deliver long-lasting impact – the system is too complex" (UN DESA, 2020). Specifically, target 17.17 addresses multi-stakeholder partnerships, strongly encouraging cross-sector and cross-environment collaborations. At a macro level, multi-stakeholder partnerships represent a structural backbone of the general SDGs framework (see Fig. 9.2) and provide an underlying structure to advance all other SDGs. At a micro-level, the goals enhance action coherence in a given domain (see Fig. 9.2). Merging the macro and micro levels, it becomes evident that the added value of partnerships is

[4] For further reading on ideological perspectives on the SDG framework, see Bendell, 2022; Neubauer & Calame, 2017.

[5] Universal human rights index – human rights recommendations (n.d.). Available at: https://uhri.ohchr.org/en/

[6] More SDG success stories can be found in UN's Bringing Data to Life 2022 report. Available at: https://unstats.un.org/sdgs/report/2022/SDG2022_Flipbook_final.pdf

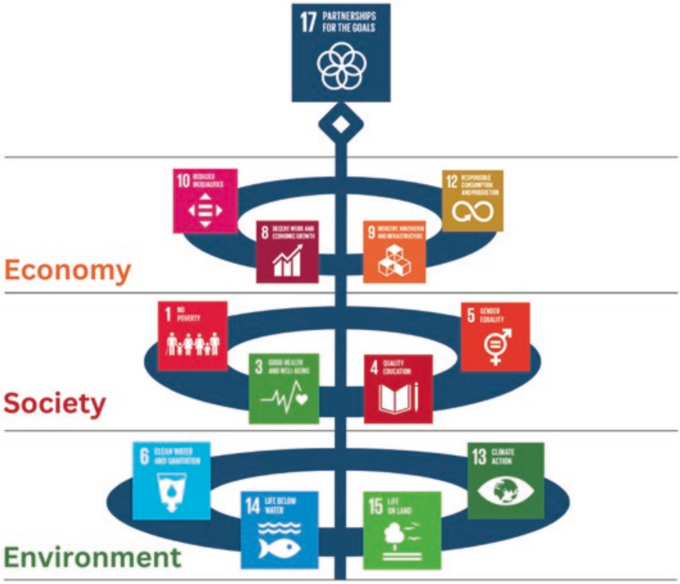

Fig. 9.2 Overview of the role of sustainable development goal 17 as the structural backbone of the sustainable development goals. (Note: adapted from UN DESA (2020, p. 10))

synergistic: the result is greater than the sum of the parts, a concept captured as "Collaborative Advantage" and "Partnership Delta" in the UN's Partnership Guidebook which can be visualized in Fig. 9.3 (UN DESA, 2020, p. 34). The same report highlights ten benefits of partnerships (see Fig. 9.4 for a summary of each). Finally, with an estimated funding gap of 2.5 trillion per year (in addition to current expenditure) to meet the SDGs by 2030 standing at USD (SECO, 2020), unlocking areas of untapped potential in terms of novel partnerships and their pooled finances is imperative.

C. System Mapping

A prerequisite to partnership formation is having a clear map of actors' aligned interests. This alignment relies on the condition that all stakeholders involved clearly communicate their activities, interests, and

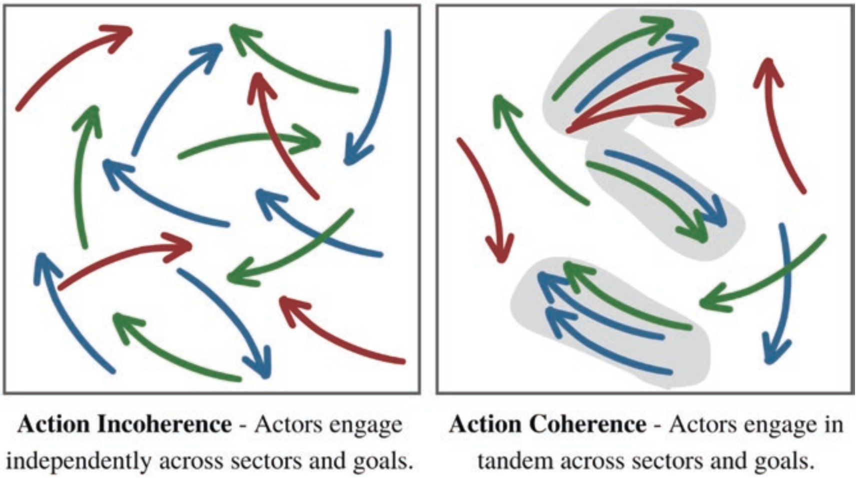

Fig. 9.3 Action incoherence versus action coherence. (Note: adapted from (UN DESA, 2020, p. 12))

1 Complementarity	2 Standards	3 Innovation	4 Critical Mass	5 Holism
Bringing together essential complementary resources.	Creating collective procedures and norms; disseminating standardised knowledge.	Combining diverse resources, thinking, approaches; novel solutions.	Collectively providing sufficient weight of action and emboldening of actors.	Holistic, locally informed, and context-appropriate nuanced approaches.
6 Shared Learning	**7 Shared Risk**	**8 Synergy**	**9 Scale**	**10 Connection**
Mechanisms for ongoing and adaptable collective learning and capability-building.	Co-investments; collectively sharing thus reducing financial or other forms of risk.	Aligning programmes and sharing resources to leverage synergies.	Scaling up delivery-capacity across geographical or cultural barriers.	Networking, connecting, building relationships; pro-social interactions.

Fig. 9.4 Overview of the benefits of partnerships. (Note: adapted from (UN DESA, 2020, p. 35))

objectives through unified SDG terminology, subsequently allowing for easy identification of and connection with similar actors. However, lack of visibility (especially for smaller agents), inconsistencies in the ways in which SDGs activities are communicated about (e.g., vocabulary used), and reliance on human-based search and connection-making constrain the system's ability to find aligned partners**.** To reduce current inefficiencies, data-driven methods can improve network visibility, diminish communication inconsistencies, and forego reliance on human-initiated searches, which are limited by the system's complexity. Such methods can also circumvent limitations caused by geographical disconnections and language barriers (e.g., across countries or within a multilingual country), further expanding the pool of possibilities for partnership formation. However, for AI tools to work in this context, access to big digital data – which is not uniform across the globe, a fact referred to as the global digital divide (i.e., between data-rich and data-poor countries) – is necessary. This divide implies that data-driven approaches to finding aligned actors currently work only in countries with extensive data availability and a solid digital infrastructure.

9.3 The Landscape of AI for SDGs

AI for social good and sustainability is a rapidly expanding field providing algorithms for aiding, accelerating, and monitoring progress toward sustainable goals (Palomares et al., 2021). AI's contribution to the SDGs (AI4SDG) can be categorized into two overarching strategies:

- *AI4SDG-A*[7]*:* using AI to develop solutions addressing SDG targets (e.g., problem-solving, measuring impact),
- *AI4SDG-B:* using AI to track and map SDGs coverage and progress in a certain sector.

The first strategy (AI4SDG-A) benefits from extensive research efforts to map various use cases for dissemination. McKinsey Global Institute

[7] For easier reference later in the text, we labelled the two distinct approaches with -A and -B.

(2018) found 160 use cases across 10 SDG-related social impact domains (e.g., crisis response, health, justice), yet not all domains receive equal AI support. For instance, more than 29 use cases were found for SDG 3 – Good Health and Well-being, while only four were related to SDG 16 – Peace, Justice and Strong Institutions (for a non-exhaustive list of examples, see Fig. 9.5). Importantly, with AI4SDG-A implementations rising and expanding across the three major SDG domains (Economy, Society, Environment), emerging evidence suggests that using AI for SDGs is not always beneficial. Vinuesa et al. (2020) document that AI can positively enable progress towards 134 out of 169 SDG targets while simultaneously inhibiting the remaining 59 targets (primarily those belonging to SDG 1 – No Poverty, 4 – Quality Education, and 10 – Reduced Inequalities). Environment is the most positively impacted category, with 93% of targets being enhanced by AI tools. In contrast, society could see the most harmful effects (38% of targets exhibiting negative interactions), partly due to biases in datasets (e.g., different social groups being represented by more or less complete data). Both the unbalanced coverage across sectors and the contrasting impact across targets are partially caused by the unequal data availability for each segment, which also explains why Health and Environment are more covered compared to Society (Google, 2019). As such, both because of biases and

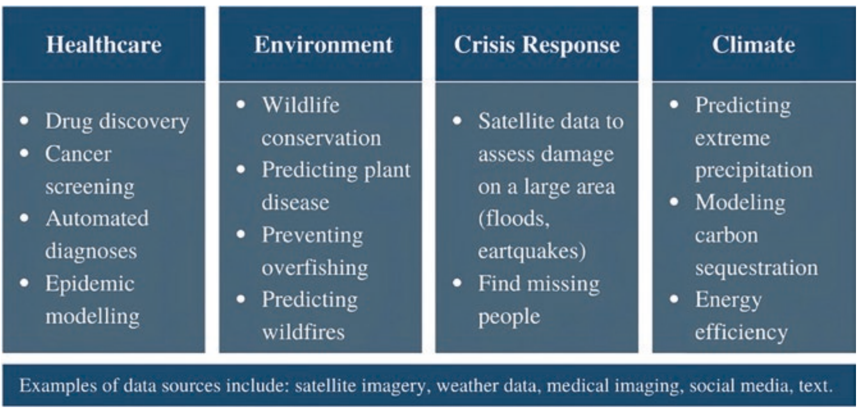

Fig. 9.5 Examples of AI4 sustainable development goal use cases

because of unbalanced data availability, AI is not currently adequate for all SDG targets.[8]

Alongside AI4SDG-A, efforts are being made to track and map SDG coverage and progress (strategy AI4SDG-B), providing a data-driven panoramic view of SDG advancements, critical for monitoring and strategic planning purposes. Across sectors and goals, the consistency and reliability of data on measured targets are uneven, creating a scarcity of consistent information and impairing SDGs-tracking (Dang & Serajuddin, 2020). Instead, what is common to all SDG projects are their reliance on human-produced communication and reporting, generating large quantities of easily traceable text. Natural Language Processing (NLP) – a type of AI – facilitates large-scale text processing to extract patterns of information, thus representing an ideal methodology for the AI4SDG-B strategy to improve the traceability and quantifiability of SDG activities greatly. It is important to mention that NLP AI4SDG-B methods are not free of pitfalls and biases, given that they depend on vocabulary specificity, which is not always precise nor consistent, in turn impacting tracking reliability for different goals.

Regarding current state-of-the-art NLP methods for SDG tracing, Wulff et al. (2023) proposed a robust labelling system that merges and fine-tunes seven other systems trying to address their biases and false positives.[9] Today, this tool represents the most comprehensive SDGs text processing method and stands behind the analyses presented in Sect. 9.5.

A. AI 4 SDG17

The first 16 SDGs are traceable to different extents with NLP algorithms across various bodies of text regardless of the text's source (Wulff et al., 2023). Despite advancements in mapping SDG coverage and a good overall assessment of the first 16 SDGS, SDG 17 lacks specificity and traceability. A potential cause is a lack of specific vocabulary used to describe SDG 17, which impacts its classification. Parallelly, in the context of AI4SDG-A (projects targeting SDG 17 itself), less than 5% of

[8] For more information on the general role of AI in sustainability, alongside opportunities and challenges, see Kar et al., 2022; Nishant et al., 2020.

[9] This labelling system is openly accessible through the text2sdg R package (Meier et al., 2021).

projects in Oxford's AI4SDGs database included a tag for SDG 17 (Nasir et al., 2023), suggesting this goal receives less attention compared to others. In sum, SDG 17 faces two challenges: (1) data-driven insights addressing its advancement remain elusive due to insufficient language specificity, and (2) emerging evidence suggests it is less prioritized (at least explicitly) in projects compared to other goals.

A novel AI approach used to map the interactive SDGs landscape and propose optimized data-driven partnerships can potentially address both issues. Despite the lack of an AI-driven strategy for SDG 17, efforts to track and enhance partnerships do exist. Examples include the UN Partnership Platform[10] – the most notable global registry of voluntary commitments of multi-stakeholder partnerships, the SDG Platform, and the AIforGood Network, among others.[11] These platforms and registries offer valuable resources for understanding existing types of partnerships, insights into implemented projects, and the different contributions of various stakeholders. Nonetheless, we identify limitations as well. The UN's Partnership Platform requires manual registration of already established partnerships (thus being a retroactive registry) and does not feature future-oriented recommendations regarding advantageous potential partnerships. The reliance on human input and lack of future-oriented applicability could be overcome through automated tools.

As discussed in Sect. 9.2, partnerships represent both the backbone and the catalyzer for SDGs advancement through action coherence and collaborative advantage. The lack of traceability, quantifiability, and optimization surrounding SDG 17 impedes both progress assessment and collaborative strategic planning. Consequently, an effort to quantifiably measure multi-stakeholder partnerships needs to be developed in addition to already-existing NLP approaches, with implications for enhancing system mapping capabilities. This approach contrasts with a manual and retroactive human registry of already established partnerships (e.g., the case of the UN Partnership Platform), expanding the possibilities to identify relevant partners.

[10] https://sdgs.un.org/partnerships/browse

[11] Examples inclusive Sustainable Development Solutions Network, World Business Council for Sustainable Development, SDG Impact Finance Platform, SDG Hub, SDG Pioneer Programme, SDG Collaboration Platform.

9.4 Philanthropy for the SDGs

Since the SDGs' adoption, Philanthropic Organizations (POs) have been playing an increasingly important role in supporting the advancement of the 2030 Agenda, aligning their objectives, operations, and grant strategies to the goals with the intent of accelerating their implementation. The SDGs Philanthropy Platform (SDGPP) aims to help the philanthropic sector meaningfully engage with the SDGs. It estimates the potential to unlock USD 651 billion in philanthropic giving by 2030, contributing to the current annual funding gap of USD 2.5 trillion (SECO, 2020).

Philanthropy's relationship with the SDGs goes beyond financial contributions (Charities Aid Foundation, 2019). While the SDGs provide an important framework through which POs can align their work with civil society, governments, and the private sector, POs' unique flexibility, value-driven approach, and independence allows them to contribute to the 2030 Agenda on multiple fronts (UN DESA, n.d.-a). As advocates, POs can raise awareness of the importance of integrating SDGs in programming efforts and the benefits of adopting a long-term focus. As innovators and risk takers, they can test new ideas and attempt innovative approaches, which is often not possible through traditional funding sources (e.g., the case of private businesses), yet crucial to address the interconnected and complex challenges of SDGs. Lastly, as impact drivers, they can catalyze change in spaces unreachable by governments, incentivize collective action, and lead the creation of multi-stakeholder partnerships by providing funds to kick-start initiatives with the potential of attracting market capital.

As Sect. 9.2 exemplifies, SDG 17 embodies the international recognition of partnerships as a tool vital to accelerating SDG implementation and highlights the urgent need to overcome barriers to collaboration. POs, with their multi-level engagement and privileged position as independent entities at the intersection of industry, government, and academia, are uniquely positioned to lead the formation of innovative partnerships for the attainment of the SDGs. Whether partnering with nonprofit organizations (NPOs), fellow POs, corporations or government agencies, the act of merging resources to advance social well-being is a central feature of the history of POs and one increasingly recognized as essential for effective philanthropy (Stanford Social Innovation Review, 2015). From the Rockefeller and Ford Foundations partnership to fuel the Green Revolution in the twentieth century to the Vaccine Alliance's key partnerships with the Bill and Melinda Gates Foundation to support mass vaccination programs and the more recent World Health Organization (WHO) Foundation's "Go Give One" campaign, examples showcasing the leveraging power of philanthropic partnerships for development are numerous. The SDGs, in particular SDG 17, have only reinvigorated this long-standing philanthropic tradition of collaboration (United Nations Partnerships for SDGs platform, n.d.); after all, the notion of partnership is not new to the philanthropic sector. Nonetheless, among the multiple actors involved in advancing the SDGs, POs are those leveraging the power of partnerships the least. Today, only 0.7% of UN-registered SDG partnerships involve POs (UN DESA, n.d.-b).

Moreover, multiple efforts have facilitated philanthropic partnerships. An example is the SDGPP 2.0, a partnership between the UN Development Program (UNDP) and Worldwide Initiatives for Grantmaker Support (WINGS), which aims to revitalize efforts to "cataly[z]e and unlock philanthropy's potential through multi-stakeholder partnerships" (SDG Philanthropy Platform, n.d.). While valuable in mobilizing big philanthropic capital, these large-scale initiatives do little to leverage smaller POs' potential adequately.[12] Specifically, while well-established foundations (e.g., the Bill and Melinda Gates Foundation) possess the financial resources, visibility, and network connections necessary to leverage partnerships' synergistic power successfully, smaller non-profit actors often do not. As a result of this gap, the potential of these smaller POs remains untapped. This untapped potential particularly impacts the implementation of local-level SDG targets and minority causes, as these smaller actors are characterized by a unique local perspective and bottom-up approach, embodying the very essence of the SDGs'

[12] Community foundations (CF) represent an example of a small philanthropic actor. Despite the different operational approaches taken by CF in different geographies, what unites their efforts is the proactive stance to bring the SDGs to the local level. Specifically, when it comes to the SDGs, the work of community foundations revolves around three core elements: alignment (local and national), dissemination of education and tools, and fostering community engagement.

mantra "leave no one behind." It is crucial to allow these actors to capitalize on "collaborative advantage" and "partnership delta," with the NLP network mapping approach presented in Sect. 9.5 being an essential step in this direction.

Notably, POs are not the principal stakeholders advancing sustainable development. On the contrary, they are part of a wide ecosystem comprising private and public companies, governments, international organizations, and civil society organizations (CSOs). Indeed, the SDGs have enabled diverse stakeholders to act, leading to success stories and projects contributing to the estimated annual USD 3.3 to 4.5 trillion needed to achieve the SDGs by 2030 (UNSDG, 2018). Despite the "noise" surrounding many of these SDGs stakeholders, examples of and insights on the philanthropic sector's targeted efforts towards the SDGs remain limited, despite POs representing a valuable player in the SDGs ecosystem.

A. Case Study on Swiss Philanthropy

Swiss philanthropy represents an ideal context to study the use data-driven methods for SDGs-partnership recommendations. Home to a philanthropic sector that generates a positive annual net growth of approximately 1% and counting over 10,000 active foundations within its territory, Switzerland has been building a strong philanthropic tradition (Eckhardt et al., 2020). As rankings demonstrate, the country's strong economy, political stability, and legal frameworks for foundations contribute to the sector's growth. Within the European context, in 2021, Switzerland ranked 6th in the number of foundations and 5th in terms of POs by population (Philea, 2021). At the international level, metrics have indicated how Swiss philanthropy extends its action well beyond its borders, scoring 4.75/5 in the Global Philanthropy Environment Index (Indiana University Lilly Family School of Philanthropy, 2018) and 4.5/5 in the Philanthropic Environment Cross-Border Flows score (Indiana University Lilly Family School of Philanthropy, 2020). Notably, Switzerland also hosts multiple UN offices, financial institutions, and other nonprofit organizations, creating an environment favorable to forming multi-sector partnerships for SDG acceleration.

Despite Switzerland's vibrant philanthropic sector and commitment to the SDGs, creating successful philanthropic partnerships is a complex task, particularly for smaller POs. Failure to align priorities, limited knowledge of the market and area of intervention, and mismatched communication[13] are only some of the elements hindering the philanthropic sector's ability to fully capitalize on the benefits of partnerships. To this end, our work develops an open-access digital tool that relies on NLP techniques allowing for a dynamic inspection of the multi-level complexities of the Swiss philanthropic landscape.

9.5 Mapping SDG-Alignment of the Swiss Philanthropic Ecosystem

Aiming to develop a novel AI methodology tailored explicitly to uncovering POs' SDG 17 potential, we matched text data from 10,755 Swiss philanthropic organizations (equally referred to as nonprofits or NPs hereafter) to SDG lexicons. Additionally, we performed the same for Swiss public companies (for-profits, FPs hereafter) listed on the Swiss Stock Exchange to compare NPs and FPs ecosystems and investigate whether there are any structural differences between the connectivity of the two sectors. To this end, we fed mission statements (hereafter MSs) of NPs and annual reports (hereafter ARs) of FPs into an NLP algorithm for SDG labelling (Meier et al., 2021), then we connected organizations to SDGs based on matching text using network-based Machine Learning. Given their shared SDG alignment, we identify potential connections between within-sector actors. Overall, this approach constitutes a new AI4SDG-B methodology to paint a landscape of sector-specific SDG alignments and strategies. It serves both as an intuitive partnership matcher within current sector dynamics and as a potent tool for proposing alternative partnership modalities to align heterogeneous stakeholders based on their SDG relationships.

[13] The presence of three official languages in Switzerland - German, French, and Italian - may represent a language barrier limiting communication amongst different cantons.

A. Results

SDG labelling results indicate that only 4% of Swiss NPs use SDG vocabulary in their MSs (see Fig. 9.6a), while 60% of public FPs communicate with SDG-aligned vocabulary in their ARs (see Fig. 9.6b). For those NPs and FPs using SDG vocabulary, only 2.2% of the entire word sample for each sector matched SDG language (see Fig. 9.6a, b). These observations reveal that, in Switzerland, although the philanthropic sector is less SDG-centric than public companies, sustainability-oriented messages represent a small fraction of their reporting (i.e., vocabulary coverage) for both types of organizations.

As previously mentioned, AI efforts to track and map SDG coverage and progress (strategy AI4SDG-B) depend on NLP tools, which have inherent limitations when vocabulary coverage is low. Mapping SDG 17 proved particularly challenging in previous efforts (Wulff et al., 2023) as well in our sample. We find that, out of the matched SDG words (535 for NPs and 478 for FPs), a staggering minority were connected to SDG 17 (2.4% for NPs and 2.9% for FPs). Thus, instead of mapping language specific to SDG 17 itself (which would show which organizations already consider partnerships in their activities), we sought to understand how the remaining SDGs interact as a proxy for mapping possible partnerships.

To obtain a landscape view of how each sector is organized around SDG achievement and discover to which extent analyzing shared SDG communication could constitute a fruitful strategy for finding potential partnership alignments, we built bipartite networks for each sector. This task entailed connecting each actor (NPs and FPs independently) to one

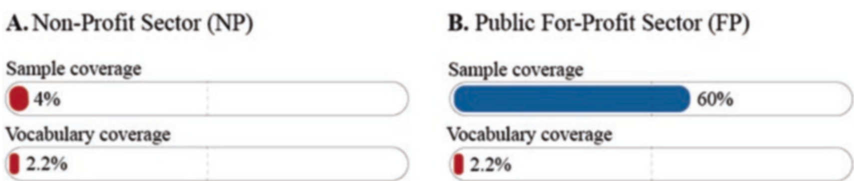

Fig. 9.6 Sample and vocabulary coverage in the non-profit and for-profit sectors

or multiple SDGs based on the words they use in their MSs or ARs, mapping all existing links (see Fig. 9.7a, b).

This approach proved both visually insightful and analytically promising for understanding the status of SDG alignment and the potential for SDG partnerships in both sectors. The NP network presents less SDG connectivity than the FP network, as shown by network density (level of network interconnectivity) (0.08 and 0.19, respectively, per Fig. 9.7a, b), highlighting the two sectors' different sustainability approaches. More precisely, NPs are preferentially connected to one specific SDG (cluster organization), as opposed to FPs, which tend to form intricate inter-SDG webs (network organization). Concerning links to SDG 17 itself, very few NPs and FPs are directly connectable to it (1.1% and 3.5% respectively), gesturing concomitantly at two explanations: (1) the challenge to track partnerships solely using NLP; and, more importantly, (2) that the present stakeholders might not value nor consider the importance of partnership in their activities in the first place (and thus not use words related to it in their communication).

To further understand the nature of each sector's SDG engagement, we explore the potential of each (NP and FP) to engage in uni- or

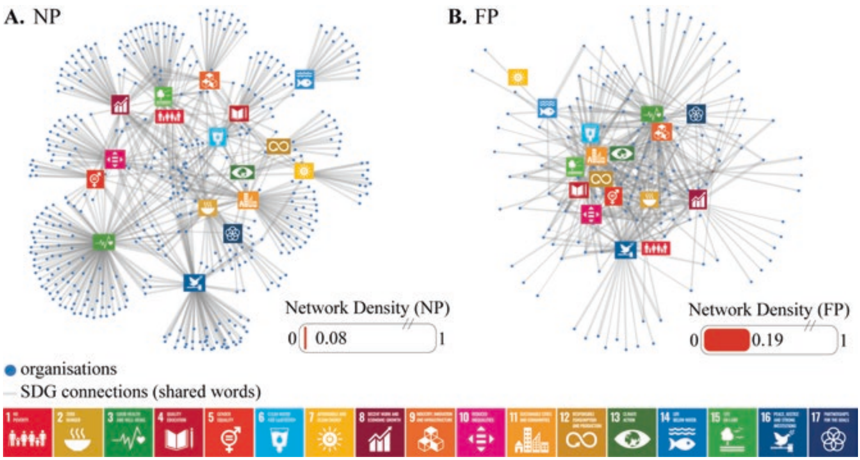

Fig. 9.7 Sustainable development goals network connections of the non-profit and for-profit sectors

multi-dimensional SDG contributions. The data showed that 73.6% of analyzed NPs align to a single goal (see Fig. 9.8a), while the inverse is true in the FP sector, with 71.4% of organizations being connected to multiple SDGs (see Fig. 9.8b), reflecting the resulting pattern already shown by network density analysis. Interestingly, the single-SDG nature of NPs could indicate that POs have narrower purpose schemes, but these do not necessarily hamper their potential for SDG partnership formation. Indeed, two types of partnerships are needed to attain SDG 17: intra- and inter-SDG, i.e., within single-SDG collaboration and between multiple SDG collaborations. As Figs. 9.7 and 9.8 show, NPs seem to be better suited for intra-SDG partnerships, allowing for focused and deep SDG action and collaboration (based on cluster organization). In contrast, FPs appeared to adopt a more inter-SDG approach and could be better at designing work plans and engaging in partnerships that tackle several SDGs simultaneously (based on network organization). Importantly, the uneven SDG alignment profiles and connectivity strategies of NPs and FPs, likely reflective of the inherent characteristics of each sector (resources, target populations, etc.), suggest the existence of different strategies afferent to each sector for tackling sustainability matters, also implying that different stakeholders can fulfil multiple SDG roles. Understanding the SDG strategies of each sector is fundamental to effectively leveraging SDG 17. As such, the insights derived from this approach can guide each sector to establish partnerships in accordance with their

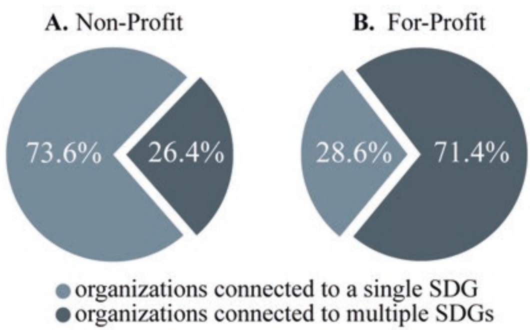

Fig. 9.8 Aggregate percentage of single- vs multi-sustainable development goals connection for each organization by sector

current SDG strategies and propose alternative ways of collaboration different from each sector's default. In both scenarios, network analysis could be instrumental in identifying advantageous connections.

Focusing hereafter on POs, we illustrate our proposed methodology's capacity for identifying precise partnership suggestions. For simplicity, and to demonstrate the point, we chose SDG 3 – Good Health and Well-being as our primary focus, given it is the most richly connected goal.

Consistent with previous findings, an SDG 3-focused analysis indicated that most NPs are single SDG-oriented (67.5%, Fig. 9.9a). The network structure visually showcases the SDG 3 community and potential paths for collaboration (see Fig. 9.9b). The light-green network represents those organizations that only connect to SDG 3 and could join forces for a unique purpose. In contrast, the dark-green network portrays actors who could collaborate on multiple targets given their multi-SDG connectivity. Importantly, similarly with the higher network density of the FP sector, the SDG 3 NP landscape seems to be in the developing stages of higher inter-network ramifications compared to other goals.

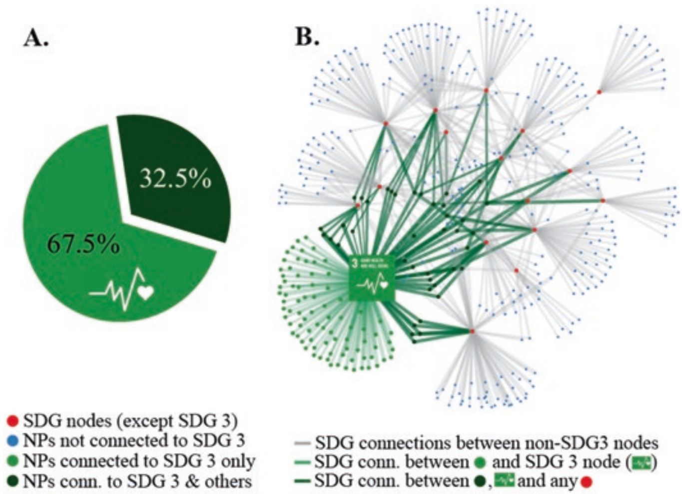

Fig. 9.9 Intra- vs inter-sustainable development goal 3 nonprofit network connections. (Notes: A. Proportion of NPs aligning to SDG 3 both uniquely (light green) or in combination with other SDGs (dark green); B. Network highlighting connections of SDG3 aligned NPs)

This result implies that NPs could adapt their activity and partnership strategies to increase overall network connectivity.

While the network view facilitates a rapid overview of the SDG 3 centric NP landscape, to further identify precise multi-SDGs partnerships with SDG 3 as a common goal, we analyzed the most frequent co-occurrences of SDGs combinations (see Fig. 9.10). The results show that common co-goals include, for example, SDG 2 – Zero Hunger and SDG 16 – Peace, Justice, and Strong Institutions, expanding the possibility of engaging in inter-goal sustainability projects.

B. Discussion

This methodological and analytical approach is a starting point for considering the role of AI beyond SDG solutions (i.e., approach AI4SDG-A) and into AI solutions for SDG coverage (AI4SDG-B). We

Fig. 9.10 Schematic of most abundant non-profit sustainable development goal 3-centric potential partnerships links. (Notes: 1 – intra-SDG potential connections, i.e., NPs connected to just SDG3 (from Fig. 9.9 – light green). 2 – dual inter-SDG connections, e.g., between SDG 2 and 3. The text in line exemplifies a few of the many foundations that share SDG 2 and 3 in their vocabulary. 3, 4, 5 – inter-SDG connections between more than two goals (3, 4, 5 respectively). Horizontal lines – connections between SDG 3 and other SDGs. Vertical lines – foundations that are connected to SDG 3 and share another SDG (e.g., SDG 2) also present partnership potential (grey dashed line))

present an AI approach for discovering new action plans for Partnership Formation (SDG 17) within a single goal (intra-SDG) or multi-goal (inter-SDG).

Mapping the SDG alignment of the Swiss philanthropic ecosystem led to several insights into its status with the broader philanthropic network and within the for-profit sector. These insights stand to inform partnership formation towards SDG achievement both by POs independently and in collaboration with FPs. We find that the philanthropic ecosystem's SDG network connectivity is low, making space for beneficial data-driven partnership suggestions. Additionally, Swiss NPs appear to have action strategies mainly focusing on a single SDG (see Fig. 9.8a), although inter-SDG foundations exist and could act as bridges. Finally, focusing our data exploration on understanding SDG 3 philanthropic alignment, we discovered network patterns indicating potential bridging points for uniting forces on SDG 3 and with other SDGs (see Fig. 9.10).

Of particular interest, the NP sector rarely describes itself in the sustainability lexicon, given that only a minority of organizations use SDG vocabulary in their MSs (4%, Fig. 9.6a), and even when they do, only 2.2% of the total text represents SDG words. This lack of SDG terminology use could constitute an entry-level problem for foundations to discover each other's SDG alignments and impede partnership formation.[14] Consequently, AI systems based on NLP mapping on their main descriptions and MSs might not always be the optimal approach for accurate mapping. However, it still represents an improvement compared to the baseline.

Future AI directions should go beyond NPs' MSs and explore, for example, the specific missions they are executing by considering their specific activities. This higher resolution of specific activity domains would lead to project-specific partnership suggestions. However, few POs systematically report their specific activities, even less so in a vocabulary that matches the SDG lexicon. A change in communication paradigm and efforts to align and standardize communications, both within and

[14] It is also possible that the Swiss philanthropic ecosystem isn't necessarily geared towards sustainability more generally. However, in a parallel analysis we ran on NP and FP communication content, generally the primary message they deliver centers on an SDG-aligned purpose.

beyond philanthropy, would greatly enhance NLP efficiency; a necessary first step would be to foster digitization and action reporting among POs (as of today, only 30% of POs have websites[15]). Lastly, POs could adopt similar strategies as FPs, not just by increasing SDG-specific communication, but also by incorporating SDG target metrics (e.g., as up to 40% of FPs provide those (GRI, 2022), whereas the same cannot be stated for POs).

Overall, this NLP approach represents a cost-efficient and automated system mapping of SDG stakeholders (in this case, applied in the context of Swiss nonprofits), based on which finding and building multi-stakeholder partnerships with aligned actors can be greatly facilitated. Whereas this chapter maps the NP and FP sectors independently for comparison purposes, it can be equally applied at the intersection of multiple sectors to find cross-sector collaborations, fully embodying target 17.17's aim.

A. Methods and Limitations

The data sources for the analysis[16] consist of mission statements (MSs) from 10,755 NPs and annual reports (ARs) from 231 FPs listed on the Swiss stock exchange. Text was collected from organizations' websites or online databases. To map SDG content in text data, we used the text2sdg R package (Meier et al., 2021). This approach successfully mapped 4% of 10,755 NPs (432) and 60% of 231 FPS (140), sample used for subsequent analyses. The innovative aspect of this approach to specifically aid SDG 17 through AI is brought by network analytics (see Figs. 9.7, 9.9, and 9.10), which allows for SDG partnership potential system mapping and intuitive visualizations of SDG-matched actors.

While bypassing several methodological limitations brought by SDG classification systems, namely for tracking SDG 17, this proposed analytical approach inherits the SDG mapping and classification biases intrinsic to the training data used in text2sdg (Meier et al., 2021). In

[15] Analysis performed on our scraped data sample, from the Stiftungschweiz repository (https://stiftungschweiz.ch/)

[16] The full methodology and its limitations, alongside the code used and output data, can be found at https://osf.io/vr5j7

addition, the datasets for NPs and FPs are unequal – there is a bigger NP sample size and there are differences in text length between MSs and ARs. However, our results do not reflect potential biases caused by such dataset imbalances given that the larger sample of NPs did not result in a denser network, and the smaller size of MSs' text length did not result in a different vocabulary coverage. Secondly, using MSs could bias results due to their permanent nature – they are formulated at the creation point of a NP and are rarely updated. To account for this possible lexicon limitation between pre- and post-2015 MSs (timestamp of SDG formulation), we compared the sample coverage of foundations created only after 2015 with the entire sample without finding significant differences (Sample Coverage of 5.4% post-2015 compared to 4% for the whole sample), indicating the overall reliability and utility of this approach independent of the permanent MSs' nature. Lastly, the current approach does not account for all fundamental factors determining successful partnerships, e.g., human dynamics or value structures. Overall, this research constitutes proof-of-concept work laying the foundation for AI-based partnership-recommendation systems.

9.6 Risks of AI in Philanthropy and SDGs

Despite POs' early and continued investment in AI,[17] the relationship between the two is somewhat paradoxical. While POs are important funders of AI, with contributions to Artificial Intelligence, Machine Learning, and Data Science technology (AIMS) philanthropy amounting to USD 2.6 billion in 2021, the extent to which POs adopt the technology remains limited (Herzog et al., 2021). The underuse of AI in the nonprofit sector is widespread, as the sector continues to possess one of the lowest AI usages (Google, 2019; Herzog et al., 2021). The limited knowledge and use of AI, especially amongst smaller POs, translates to

[17] Philanthropy played a key role in the initial development of AI. In 1956 the Rockefeller foundation gave a grant of USD 7'500 to support a conference in Dartmouth, US, now widely recognized as the birthplace of modern AI. Moreover, the first recorded usage of the term "Artificial Intelligence" can be found in a grant application that was made for the conference (Rockefeller Philanthropy Advisors, 2019).

these actors operating below their full potential, including supporting the SDGs. AI can play an essential role in enhancing POs' contribution to the SDGs, both in terms of SDG targeting (AI4SDG-A) and SDG coverage (AI4SDG-B). As such, uncovering why POs remain hesitant to adopt AI and providing tangible risk-mitigation efforts is necessary to contribute to the sector's digitalization and the overall SDGs advancement. For instance, in the Swiss philanthropic ecosystem, over twelve different concerns were raised by POs, among which data privacy, AI explainability, corruption detection, and job displacement (Della Giovampaola et al., 2023), however most of those concerns remain without adequate risk mitigation plans in place.

Multiple issues are halting POs' decision to integrate AI into their systems and operations. Firstly, given the largely for-profit nature of the tech sector, most AI software is designed to extract maximum profit from digital data. Often this entails the collection and the long-term holding and handling of digital data. Such practices, particularly in vulnerable humanitarian settings, can be dangerous and lead to discrimination, polarization, and what has been referred to as "techno-colonialism" (Madianou, 2019). Buolamwini and Gebru (2018) and Whittaker et al. (2019) show, respectively, how commercially available facial recognition algorithms are less accurate in their recognition of women with darker skin tones and have greater difficulty in the identification of people with disabilities when these individuals use assistive technologies, such as wheelchairs. The lack of diversity in training datasets is part of the cause. This lack of diversity, in turn, can increase inequality within societies, having a large impact on peace (SDG 16) and stability (SDG 17) (Gupta et al., 2021). Cyber-attacks and data breaches, which are often directed at humanitarian organizations (such as the 2022 cyber-attack conducted against the International Committee of the Red Cross (ICRC, 2022)), further heighten these risks. As such, "collect little and destroy as soon as possible" is seen as a far more suitable modus operandi for safeguarding sensitive data (Bernholz & Reich, 2017). However, creating and managing such tailored infrastructure requires vast resources mostly unavailable to POs.

Available data quality is another main issue. Databases of POs' activities and strategies are frequently unavailable, incomplete, inaccurate, and contain irrelevant data (known as "data deserts"). These shortcomings can severely hinder POs' ability to leverage AI's power and can heighten bad practices in worst-case scenarios (Kanter & Fine, 2020). Decision-making informed by ill-adapted AI technologies can dangerously magnify and reinforce pre-existing inequalities and biases (O'Brien, 2022). This increase in inequality can heighten exclusion and discrimination, particularly against minority groups, countering the overall objective of the 2030 Agenda of a more equitable and sustainable world. Uneven data availability also widens the gap between unequal levels of support provided to different SDGs. For example, the increased use of AI to address health (SDG 3) and climate issues (SDG 13) is, in part, due to the vast amount of data available, as opposed to Peace and Justice (SDG 16), an issue area more complex to capture with data (Google, 2019). Given this SDG data gap, POs may focus primarily on issues benefiting from sufficient data availability to ensure adequate impact. Overall, data collaboratives and comprehensive outcomes represent timely yet efficient solutions to address the problem of "data deserts" (Kanter & Fine, 2020).

Even before being treated, digital data presents multiple difficulties. Data ownership, data quality, and data type (i.e., sensitive or personal data) are some of the most recurring issues. Philanthropy is centered around the voluntary giving of private resources, whose ownership is largely clear and undisputed. By contrast, digital data is contested property (Bernholz & Reich, 2017). Who is the "true" owner? The person whose information is involved, the company providing the software collecting the data, or the platform on which the data is collected? Despite the question of data ownership remaining unanswered, the phenomena of "data philanthropy"[18] – the donation of data from private companies for socially beneficial purposes – and "data-raising" – efforts to get people to give their data for a cause – are gaining traction (Bernholz, 2021; Lev Aretz, 2019; Taddeo, 2016, 2017).

Even if unintentional, divulgating misinformation is another important yet often overlooked risk that may derive from using AI to address social problems. Many factors may fuel the production of misinformation,

[18] The term was reportedly coined by World Economic Forum CTO Brian Behlendorf during a spontaneous conversation at the 2011 World Economic Forum (Lev Aretz, 2019).

including language and biases, particularly as these technologies are primarily developed and fabricated in the "Global North." Misinformation generated by human-machine communication can be at odds with the widely respected humanitarian imperative of "do not harm" and raises important issues around responsibility and supervision. Moreover, one must not forget that AI needs to experiment to learn. In the humanitarian sphere, this may again be severely at odds with the "do no harm" principle, leaving the use of such technology at a crossroads (Coppi et al., 2021).

In addition to these more tangible risks, POs also raise ethical concerns when discussing their reluctance to incorporate AI. Despite the multiple ethical frameworks and principles for AI proposed in the last decade, much disillusionment appears to surround these frameworks, and POs question their adequacy in catering to the needs of the nonprofit sector. One important step to boost the sector's confidence in AI is greater "explainability" of algorithms, also known as "explainable AI", to target algorithmic opacity (Coppi et al., 2021). Additionally, incorporating AI in a sector that, since inception, revolves around human welfare may have a vaster impact. The sector's resistance to AI may not only derive from the limited knowledge and resources available to operate this technology, but also from a fear of alienation. The delegation of tasks from humans to machines can be seen as unnatural and inadequate by many sector professionals, who may view it as diluting their efforts. Sharing successful case studies and establishing adequate frameworks to strengthen transparency and accountability in the use of AI in humanitarian contexts might attenuate this fear.

It is hard to predict the positive and negative impact that technology can and will have on philanthropy. However, what cannot be ignored is how technology democratizes philanthropy. By increasing the connectivity between the receiver and the giver, improving efficiency and transparency, and widening operations reach, AI has much to offer to the philanthropic sector and vice versa. Moreover, when considering the risks that may accompany the incorporation of AI in POs, one must not forget that the underuse of AI can also be considered a misuse in itself (Floridi et al., 2018). The tool presented in this chapter aims to provide a risk-free alternative on how technology can be an SDG achievement enhancer. Specifically, it seeks to initiate and enhance an agent's ability to identify and visualize possible aligned actors to enhance partnership visibility. The tool itself does not present any inherent risks or dangers, as it does not handle sensitive data and does not aim to provide an AI4SDG-A type solution with ethical implications for human or planetary well-being (e.g., an algorithm predicting cancer subtypes with evident consequences for human health). Instead, it makes use of public data, and no inherent risks caused by possible data biases can be identified, with the main consequence of biased "word-classifiability" being that some potential partnership connections will not be identified.

9.7 Conclusion

This chapter approaches sustainability using the SDG framework, engaging in a balanced discourse of the framework itself, followed by the crucial yet underleveraged role of SDG 17. We present the landscape of AI for SDGs, focusing on AI for SDG 17. We also highlight the crucial role of POs as value-driven players engaging in promoting activities geared towards SDG attainments and emphasize small POs as underrepresented actors. A conclusion of the chapter's findings is presented in Fig. 9.11.

Together, an underleveraged goal and a set of underrepresented actors present significant growth potential. We focus on Swiss philanthropic organizations to illustrate how an AI NLP tool can be used to uncover potential partnerships among NPs and FPs, illustrating the role of AI for SDG 17 itself – method which is translatable in other national or international contexts both in non-profit and for-profit ecosystems. Overall, the untapped growth potential can be uncovered through this network mapping approach, which helps connect stakeholders with SDGs. This cost-efficient approach increases the visibility of aligned actors in a complex system, for which partnership potential would otherwise be more difficult to trace. Lastly, we present a general overview of possible risks associated with AI in the context of POs, while noting that this specific AI4SDG-B approach we propose does not present any inherent risks. Future work could examine how action purposes, moral values, or target populations influence SDG alignment, which could inform potential partnerships based not only on shared interest alignment but also on other factors, leading to a multi-dimensional integrative approach.

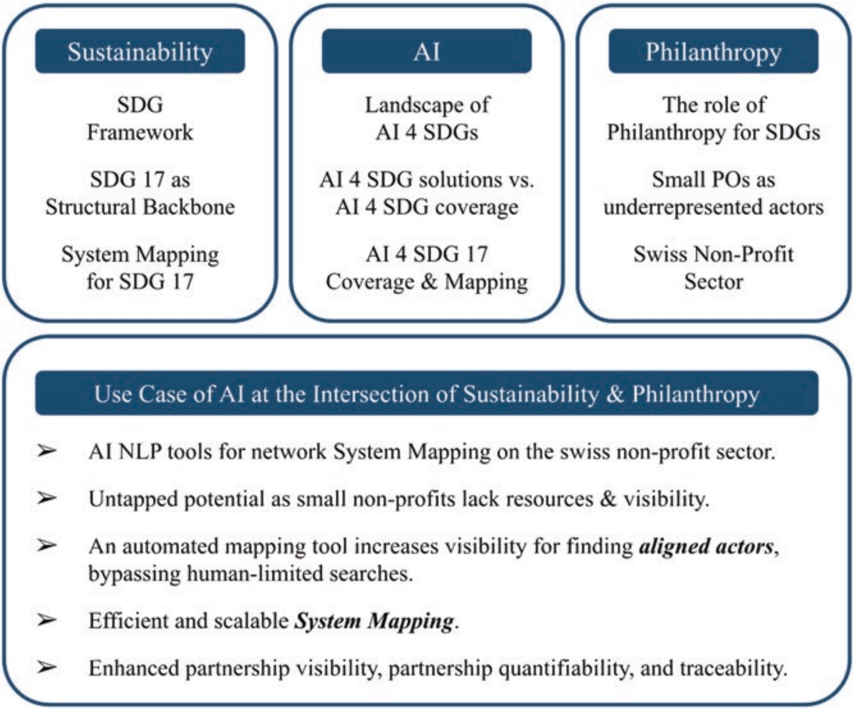

Fig. 9.11 Visual chapter summary

References

Belmonte-Ureña, L. J., Plaza-Úbeda, J. A., Vazquez-Brust, D., & Yakovleva, N. (2021). Circular economy, degrowth and green growth as pathways for research on sustainable development goals: A global analysis and future agenda. *Ecological Economics, 185*, 107050. https://doi.org/10.1016/j.ecolecon.2021.107050

Bendell, J. (2022). Replacing sustainable development: Potential frameworks for international cooperation in an era of increasing crises and disasters. *Sustainability, 14*(13) Article 13. https://doi.org/10.3390/su14138185

Bernholz, L. (2021). *Philanthropy and digital civil society: Blueprint 2022*. Stanford PACS. https://pacscenter.stanford.edu/publication/philanthropy-and-digital-civil-society-blueprint-2022/

Bernholz, L., & Reich, R. (2017). *Nonprofit data governance*. Stanford PACS. https://pacscenter.stanford.edu/publication/nonprofit-data-governance/

Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. Proceedings of machine learning research: Conference on Fairness, Accountability and Transparency, 81. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Charities Aid Foundation. (2019). *The role of philanthropy for the SDGs | CAF online*. https://www.cafonline.org/about-us/blog-home/giving-thought/the-role-of-giving/the-role-of-philanthropy-for-the-sdgs-is-not-what-you-expect.

Coppi, G., Moreno Jimenez, R., & Kyriazi, S. (2021). Explicability of humanitarian AI: A matter of principles. *Journal of International Humanitarian Action, 6*(1), 19. https://doi.org/10.1186/s41018-021-00096-6

Dang, H.-A. H., & Serajuddin, U. (2020). Tracking the sustainable development goals: Emerging measurement challenges and further reflections. *World Development, 127*, 104570. https://doi.org/10.1016/j.worlddev.2019.05.024

Della Giovampaola, C., Tudor, M. C., Gomez, L., & Ugazio, G. (2023). *Current and Potential AI use in Swiss Philanthropic Organizations—Survey Results.* SwissFoundations. https://www.swissfoundations.ch/fr/actualites/current-and-potential-ai-use-in-swiss-philanthropic-organizations-survey-results/

Eckhardt, B., Jakob, D., & von Schnurbein, G. (2020). Rapport sur les fondations en Suisse. *CEPS Forschung und Praxis, 21*.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Google. (2019). *Accelerating social good with artificial intelligence: Insights from the Google AI Impact Challenge*. https://services.google.com/fh/files/misc/accelerating_social_good_with_artificial_intelligence_google_ai_impact_challenge.pdf

GRI. (2022). *GRI – Most companies align with SDGs – But more to do on assessing progress*. https://www.globalreporting.org/news/news-center/most-companies-align-with-sdgs-but-more-to-do-on-assessing-progress/

Gupta, S., Langhans, S. D., Domisch, S., Fuso-Nerini, F., Felländer, A., Battaglini, M., Tegmark, M., & Vinuesa, R. (2021). Assessing whether artificial intelligence is an enabler or an inhibitor of sustainability at indicator level. *Transportation Engineering, 4*, 100064. https://doi.org/10.1016/j.treng.2021.100064

Herzog, P. S., Naik, H. R., & Khan, H. A. (2021). *AIMS philanthropy project: Studying AI, machine learning & data science technology for good [technical report]*. Indiana University Lilly Family School of Philanthropy and Indiana University School of Informatics and Computing. IUPUI. https://scholarworks.iupui.edu/handle/1805/25177

Hickel, J. (2019). The contradiction of the sustainable development goals: Growth versus ecology on a finite planet. *Sustainable Development, 27*(5), 873–884. https://doi.org/10.1002/sd.1947

Hickel, J., & Kallis, G. (2020). Is green growth possible? *New Political Economy, 25*(4), 469–486. https://doi.org/10.1080/13563467.2019.1598964

ICRC. (2022). *Cyber attack on ICRC: What we know (Europe and Central Asia/Switzerland)*. https://www.icrc.org/en/document/cyber-attack-icrc-what-we-know

Indiana University Lilly Family School of Philanthropy. (2018). *Global philanthropy environment index: Switzerland*. Global Philanthropy Indices. https://globalindices.iupui.edu/index.html.

Indiana University Lilly Family School of Philanthropy. (2020). *Global philanthropy tracker*. Global Philanthropy Indices. https://globalindices.iupui.edu/tracker/index.html

Kanter, B., & Fine, A. (2020). *AI4Giving: Unlocking generosity with artificial intelligence: The future of giving*. https://ai4giving.org

Kar, A. K., Choudhary, S. K., & Singh, V. K. (2022). How can artificial intelligence impact sustainability: A systematic literature review. *Journal of Cleaner Production, 376*, 134120. https://doi.org/10.1016/j.jclepro.2022.134120

Kleespies, M. W., & Dierkes, P. W. (2022). The importance of the sustainable development goals to students of environmental and sustainability studies—A global survey in 41 countries. *Humanities and Social Sciences Communications, 9*(1), 1. https://doi.org/10.1057/s41599-022-01242-0

Lev Aretz, Y. (2019). Data philanthropy. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3320798

Madianou, M. (2019). Technocolonialism: Digital innovation and data practices in the humanitarian response to refugee crises. *Social Media & Society, 5*(3) https://journals.sagepub.com/doi/10.1177/2056305119863146

Mair, S., Jones, A., Ward, J., Christie, I., Druckman, A., & Lyon, F. (2018). A critical review of the role of indicators in implementing the sustainable development goals. In W. Leal Filho (Ed.), *Handbook of sustainability science and research* (pp. 41–56). Springer International Publishing. https://doi.org/10.1007/978-3-319-63007-6_3

McKinsey Global Institute. (2018). Applying AI for social good | McKinsey. https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good.

Meier, D., Mata, R., & Wulff, D. (2021). *text2sdg: An open-source solution to monitoring sustainable development goals from text*.

Miller, R. (2016, November 8). *Awareness of sustainable development goals (SDGs) vs millennium development goals (MDGs)*. GlobeScan | Know Your World. Lead the Futures https://globescan.com/2016/11/08/awareness-of-sustainable-development-goals-sdgs-vs-millennium-development-goals-mdgs/

Nasir, O., Javed, R. T., Gupta, S., Vinuesa, R., & Qadir, J. (2023). Artificial intelligence and sustainable development goals nexus via four vantage points. *Technology in Society, 72*, 102171. https://doi.org/10.1016/j.techsoc.2022.102171

. (2020). Time to revise the sustainable development goals. *Nature, Nature, 583*(7816), 331–332. https://doi.org/10.1038/d41586-020-02002-3

Neubauer, C., & Calame, M. (2017, May 8). *Global pressing problems and the sustainable development goals*. Guni Network. https://www.guninetwork.org/articles/global-pressing-problems-and-sustainable-development-goals

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management, 53*, 102104. https://doi.org/10.1016/j.ijinfomgt.2020.102104

O'Brien, C. (2022). *Big data and A.I. for the SDGs: Private corporation involvement in SDG data-driven development, policy and decision-making*. United Nations Department of Economic and Social Affairs Sustainable Development. https://sdgs.un.org/documents/big-data-and-ai-sdgs-private-corporation-involvement-sdg-data-driven-development-policy

OHCHR. (2021). *Artificial intelligence ensuring human rights at the heart of the sustainable development goals*. OHCHR. https://www.ohchr.org/en/stories/2021/03/artificial-intelligence-ensuring-human-rights-heart-sustainable-development-goals

Palomares, I., Martínez-Cámara, E., Montes, R., García-Moral, P., Chiachio, M., Chiachio, J., Alonso, S., Melero, F. J., Molina, D., Fernández, B., Moral, C., Marchena, R., de Vargas, J. P., & Herrera, F. (2021). A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: Progress and prospects. *Applied Intelligence, 51*(9), 6497–6527. https://doi.org/10.1007/s10489-021-02264-y

Philea. (2021). *About philanthropy*. Philea. https://philea.eu/philanthropy-in-europe/about-philanthropy/

Rockefeller Philanthropy Advisors. (2019). *Philanthropy and the SDGs: practical tools for alignment*. https://www.rockpa.org/project/sdg/

Schröder, P., Bengtsson, M., Cohen, M., Dewick, P., Hofstetter, J., & Sarkis, J. (2019). Degrowth within – Aligning circular economy and strong sustainability narratives. *Resources, Conservation and Recycling, 146*, 190–191. https://doi.org/10.1016/j.resconrec.2019.03.038

SDG Philanthropy Platform. (n.d.). *Engaging philanthropy in response to COVID-19 and the SDGs*. Retrieved February 22, 2023, from https://www.sdgphilanthropy.org/home

SECO. (2020). *New public-private partnership to promote innovative development finance*. State Secretariat for Economic Affairs (SECO). https://www.seco.admin.ch/seco/en/home/seco/nsb-news/medienmitteilungen-2021.msg-id-86208.html

Stanford Social Innovation Review. (2015). *The power of philanthropic partnerships*. https://stanford.ebookhost.net/ssir/digital/34/ebook/1/index.php?e=34&user_id=241847&flash=0.

Swain, R. B. (2018). A critical analysis of the sustainable development goals. In W. Leal Filho (Ed.), *Handbook of sustainability science and research* (pp. 341–355). Springer International Publishing. https://doi.org/10.1007/978-3-319-63007-6_20

Taddeo, M. (2016). Data philanthropy and the design of the infraethics for information societies. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2083), 20160113. https://doi.org/10.1098/rsta.2016.0113

Taddeo, M. (2017). Data Philanthropy and individual rights. *Minds and Machines, 27*(1), 1–5. https://doi.org/10.1007/s11023-017-9429-2

UN. (2015). *THE 17 GOALS | Sustainable Development*. https://sdgs.un.org/goals

UN. (2022). *The Sustainable Development Goals Report 2022 | DISD*. https://www.un.org/development/desa/dspd/2022/07/sdgs-report/

UN DESA. (2020). *The SDG Partnership Guidebook | Department of Economic and Social Affairs*. https://sdgs.un.org/publications/sdg-partnership-guidebook-24566.

UN DESA (n.d.-a). *SDGs. Make the SDGs a reality*. Retrieved February 22, 2023, from https://sdgs.un.org/

UN DESA (n.d.-b) *Partnerships Registry | Sustainable Development*. United Nations. https://sdgs.un.org/partnerships/browse

United Nations Partnerships for SDGs platform. (n.d.). SDG philanthropy platform: Helping philanthropy engage in the global development agenda. Retrieved 22 February 2023, from https://sustainabledevelopment.un.org/partnerships/SDGphilanthropy

UNSDG. (2018). *UNSDG | Unlocking SDG Financing: Findings from Early Adopters*. https://unsdg.un.org/resources/unlocking-sdg-financing-findings-early-adopters

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications, 11*(1), Article 1. https://doi.org/10.1038/s41467-019-14108-y

Whittaker, M., Alper, M., Bennett, C., Hendren, S., Kaziunas, L., Mills, M., Ringel Morris, M., Rankin, J., Rogers, E., Marcel, S., & Myers West, S. (2019). *Disability, bias, and AI*. AI Now Institute. https://ainowinstitute.org/publication/disabilitybiasai-2019

Wulff, D. U., Meier, D. S., & Mata, R. (2023). *Using novel data and ensemble models to improve automated labeling of sustainable development goals* (arXiv:2301.11353). arXiv. Doi:https://doi.org/10.48550/arXiv.2301.11353

10

The Potential Role of Artificial Intelligence in the Commercialization of Traditional Medicines in Tropical Regions

Robert Brian Smith, Mark Perry, and Darryl Robert Smith

10.1 Introduction

Indigenous and other local marginalized communities could sustainably harvest thousands of species of tropical plants for incorporation into pharmacological, traditional, and complementary medicines (T&CM). In 2019, the annual global pharmacological medicine market was

R. B. Smith (✉) • M. Perry
School of Law, University of New England, Armidale, NSW, Australia
e-mail: r.b.smith@unswalumni.com; mperry21@une.edu.au; darryl@radio-active.net.au

D. R. Smith
School of Law, University of New England, Armidale, NSW, Australia

Radioactive Networks Pty Ltd, Sydney, Australia
e-mail: darryl@radio-active.net.au

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1_10

estimated to be worth about USD 1.1 trillion (Calixto, 2019). Around 35% of pharmaceutical medicines were sourced directly or indirectly from natural products. Of these, around 25% are derived from plants, 13% from microorganisms, and roughly 3% from animals. Therefore, naturally derived products are a critical resource for the development of new medicines.

For such an initiative to be effective, the identification of potentially useful plants, their harvesting, and commercialization requires assistance from development partners, universities, and non-governmental organizations (NGOs). The issue is that the information on tropical plants and their efficacy as treatments and cosmetics is not readily accessible. A significant impediment occurs when plants are investigated without any efficacy being found, as there are few incentives to document negative findings. A lack of accurate and readily available information can result in the misallocation of scarce resources and the duplication of efforts.

There are continuing initiatives such as those of the Association of Southeast Asian Nations (ASEAN), including the *ASEAN Centre for Biodiversity Traditional Knowledge Digital Library* (ASEAN Centre for Biodiversity, 2019) and the *ASEAN Herbal and Medicinal Plants* (Ali et al., 2010; Sukmajaya et al., 2017). Other initiatives include *Herbs & Natural Supplements: An Evidence-based Guide* (Braun & Cohen, 2015), *YaTCM (Yet another Traditional Chinese Medicine database)*, a free web-based toolkit (Li et al., 2018), and *Women's Knowledge: Traditional Medicine and Nature – Mauritius, Reunion and Rodrigues* (Pourchez, 2017). However, current global and regional initiatives are hindered by the fact that biodiversity information is held by various institutions and recorded in different structures and dialects.

The World Health Organization (WHO) has noted that developing countries in South and Southeast Asia do not often have the staff needed to carry out pharmacovigilance of pharmacological medicines, let alone traditional and complementary medicine (T&CM) (Suwankesawong, 2019). The European Commission (2023) defines pharmacovigilance as "the process and science of monitoring the safety of medicines and taking action to reduce the risks and increase the benefits of medicines." The process includes collecting and managing safety issues, examining data to detect any new or changing safety issues, evaluating data to identify safety issues, proactive risk management, protecting public health, communicating with stakeholders (including the public), and the audit of the outcomes and processes.

In the countries of South and Southeast Asia, staff shortages also mean that databases are incomplete, and information on the efficacy and contra-indications of medicines is often lacking (Suwankesawong, 2019). Project-based initiatives often have limited information because of issues with project design. At the regional level, information sharing, particularly regarding species location, is often purposely not made available in order to hinder poaching and other illegal activities. Southeast Asia contains 20% of the planet's vertebrate and plant species and is home to the world's third-largest tropical forest. Additionally, this region has some of the highest deforestation rates on the planet and direct over-exploitation of species. Even when forests remain intact, they are steadily emptied of their biodiversity through hunting.

This chapter investigates the issues associated with developing traditional and complementary medicines and the role that AI might play. It addresses the challenges associated with identifying promising medicinal plants to be commercialized in a sustainable fashion. To overcome this challenge, the associated traditional knowledge must be protected, as must forest resources. The chapter then provides examples of existing sources of information for search by AI and how AI has been utilized in the pharmaceutical industry. We then define a role for AI and discuss the proposed data model and its associated obstacles and opportunities. Finally, the chapter concludes with a discussion and recommendations for future studies.

10.2 Sustainability Challenges with Developing Traditional and Complementary Medicines

Artificial intelligence (AI) has the potential to identify promising plants that could be utilized to produce efficacious traditional medicines for commercialization by marginalized communities. With proper support from Universities, NGOs, and other extension services, products could be developed and marketed to the broader community (see, for instance, Smith and Smith (2023); Smith and Perry (2023a)). Marginalized communities would then be able to improve their standard of living and become more sustainable. Community members would be involved in the collection of plants from the forest and, ideally, in developing their own land to cultivate the target plants. They would also be involved in the production and commercialization process.

Preserving traditional knowledge has become a critical sustainability issue. In their study of African Traditional Herbal Medicines (ATHMed), Devine et al. (2022) argue that "knowledge on these African traditional herbal medicines is gradually being lost if not lost already" (Devine et al., 2022, p. 11). Of major concern is that their preparation and administration is often based on the transfer of knowledge which is often word of mouth and poorly documented. This can result in poorly trained ATHMed practitioners, with limited knowledge of the treatment process, administering poor medication. The potential for patient having a negative response to their medication is high (Devine et al., 2022, p. 11). Furthermore, a study of the protection and promotion of traditional knowledge of marginalized communities from a legal perspective by Smith and Perry (2023b), which focused on the countries of Southeast Asia, concluded that the provision of an intellectual property framework must be supported by legislation which protects traditional knowledge and sets the framework for its promotion and commercialization. In addition, regulations and codes of practice must be developed to address issues such as good manufacturing practices (GMP), product safety, advertising, and trade practices.

In terms of establishing a legal framework, the World Intellectual Property Organization (WIPO) (2023) explains that "traditional knowledge as such – knowledge that has ancient roots and is often oral – is not protected by conventional intellectual property (IP) systems." Innovative products based on traditional knowledge may be able to be protected using patents, trademarks, geographical indications, or as a trade secret or confidential information" (WIPO, 2023). We consider geographical indication protection the best option, provided the product has unique properties due to its location.

The third pillar of the sustainability paradigm is the sustainability of forest resources. This issue is far more complex as marginalized communities seek to survive in a competitive world. Major initiatives would need to aim to allow local communities to benefit from their participation in the commercialization process without adding to the global loss of more than 13 million hectares of trees per year (Rainforest Alliance, 2023). For instance, Robinson and Raven (2019) give the example of how the Berber (Amazigh) women in Morocco have developed a "booming" international demand for skin and hair care products containing Argan from the Moroccan Argan tree (*Argania spinosa*) which is endemic to Morocco and has been used for nearly 1000 years. Women's cooperatives have been developed, and the women have benefited from infrastructure development, improvements in product quality, product development, and benefit-sharing agreements with developers from the cosmetic and skin care industries.

The challenges outlined above present a particular set of variables. We question how AI can be leveraged in data collection and analytics to mitigate these challenges and what possible sources of information and data can be collected using AI.

10.3 Examples of Sources of Information on Traditional and Complementary Medicines

There are three primary purposes for the collection of data on traditional and complementary medicines:

(1) pharmaceutical companies seeking to exploit available knowledge to identify plants that could be utilized in the development of pharmaceutical products, providing cures for ailments and diseases in humans and animals which can be commercialized for their financial benefit;
(2) governments and organizations seeking to ensure that traditional and complementary medicine products are accurately documented and registered so that the prior rights of traditional knowledge can be recognized when patent applications are lodged for a particular compound or process; and

(3) organizations collecting and recording traditional knowledge related to T&CM products so that local communities can commercialize products to assist in developing their communities.

The *Traditional Knowledge Data Library (TKDL)* of India was explicitly developed as a tool to prevent the biopiracy of traditional knowledge (Council of Scientific & Industrial Research (CSIR), 2023). The source of this traditional knowledge is existing literature in local languages such as Sanskrit, Urdu, Arabic, Persian, and Tamil. The data, converted into digitized format, is available in English, German, Spanish, French, and Japanese. The information is classified according to a comprehensive Traditional Knowledge Resource Classification (TKRC) developed by the CSIR. The TKDL database comprises around 424,000 formulations and practices. Data can only be accessed by patent assessors worldwide, as they determine whether a patent application covers products or processes in the public domain. The aim is for such patent applications to be denied rather than undertaking the costly process of revoking a patent.

As noted above, the Association of Southeast Asian Nations (ASEAN) actively promotes the protection and promotion of traditional knowledge amongst its ten member states. Significant initiatives include two compendiums describing herbal and medicinal plants within the ASEAN region in which plants are reported by country (Ali et al., 2010; Sukmajaya et al., 2017). Each entry includes a photograph, scientific name, vernacular names, brief plant description, propagation method, geographical distribution/ecology, chemical constituents, reports on medical usage (i.e. uses supported by experimental/clinical data; uses in traditional medicine), contraindications, and a bibliography. Subsequently, ASEAN developed a *Traditional Knowledge Digital Library* "crafted on the codified traditional knowledge, particularly about medicinal plants and formulations" (ASEAN Centre for Biodiversity, 2019). It is a publicly available database and is searchable by ASEAN country and common ailment for which the medicinal plant is used as a treatment. It lists ailments such as asthma, cold, cough, diabetes, diarrhea, fever, headache, painful swellings, skin diseases, and ulcers. The relevant plant species are listed by their botanical name, common name (transliterated into the Latin alphabet as required), all indications for which it is claimed as a treatment, and the ASEAN countries of origin, together with a photograph of the plant and its flower. However, it does not include studies showing the efficacy of the claimed treatments. The TKDL compendium mentioned earlier in this paragraph is superior in this regard.

Regardless of a treatment's "claimed" efficacy, the evidence must support it. For instance, the *Draft ASEAN Agreement on Regulatory Framework for Traditional Medicines* of 2017 proposes a three-tiered level of evidential requirements to substantiate a traditional medicine claim for approval by health authorities (see Table 10.1). In addition, the draft provides a

Table 10.1 Degree of evidence required to support different traditional medicine claims

Type of traditional medicine claim	Level of evidence	Evidence to substantiate claim
Traditional health use	Evidence from documented traditional use and knowledge	Evidence of documented traditional use or history of use may be found in the following: • Classical traditional medicine texts • Pharmacopoeias and monographs • Reference textbooks/journals
Traditional treatment	Evidence from documented traditional use and knowledge	Evidence of the documented history of traditional treatment may be found in the following: • Classical traditional medicine texts • Pharmacopoeias and monographs • Reference textbooks/journals
Scientifically established treatment	Scientific data and traditional medicine principles	*Compulsory evidence*: Substantiation of traditional medicine claims based on scientific data as required by the regulatory authority to be conducted on finished products or ingredient(s). Justification must be provided to the regulatory authority if the evidence provided is based on the ingredients. *At least one piece of additional evidence*: Evidence of the documented history of traditional treatment that may be found in the following: • Classical traditional medicine texts • Pharmacopoeias and monographs • Reference textbooks/journals

Source: ASEAN (2017) Annex 7 Table 2

comprehensive regulatory process to ensure efficacy before the product can be placed on the market.

The conventional pharmaceutical industry provides examples of how AI might be utilized for the traditional and complementary medicine industry, which is discussed in the next section.

A. Examples of the Use of AI in the Pharmaceutical Industry

In North America, researchers used artificial intelligence to screen roughly 7500 molecules and discovered a new antibiotic. The screening revealed that "the antibacterial compound *abaucin*, which was originally touted as an anti-diabetes drug, could inhibit the growth of *Acinetobacter baumannii* in vitro" (Inside Precision Medicine, 2023). This finding is potentially highly significant as the bacterium can display multidrug resistance and "may lead to pneumonia and meningitis, among other serious conditions, and was a leading cause of infection among soldiers wounded in Iraq and Afghanistan" (Inside Precision Medicine, 2023, p. 1).

Workman et al. (2019) consider that "oncology probably represents the therapeutic area that has most benefited from Big Data, not least due to the amount of data and resources that are now publicly available" (p. 1094). They note, however, that data quality is a continuing issue. The maximum impact and benefits were considered to occur where the benefits of Big Data and AI were utilized throughout the drug discovery journey and continued into the clinical development and routine care phases. Tripathi et al. (2021) have come to a similar conclusion and consider that "the knowledge gleaned from applying these AI algorithms in big data has provided a stimulus to design and discover novel molecules and their further optimization" (p. 1054). AI can better identify preclinical candidates by rapidly screening virtual compound libraries, representing a significant advantage over traditional methods.

10.4 Potential Role for AI in Developing a Comprehensive Process to Identify Efficacious Traditional Medicines

Much of the traditional knowledge associated with traditional and complementary medicines is not well documented and often undocumented. As discussed above, a major issue that leads to this is the lack of staff to undertake pharmacovigilance. In addition, databases are often very incomplete and, critically, information on the efficacy and contraindications of medicines is often lacking. In some cases, the acceptance of oral advice on the efficacy of traditional treatments leads to the misapplication of treatments. In extreme cases, this could lead to dire consequences for the patient. Whilst there are many potential challenges, AI could contribute to the process by collecting and consolidating data, particularly in the collection of the efficacy of treatments using traditional and complementary medicines.

A. Proposed Data Model

As discussed above, the preferred model would resemble that of the two compendiums prepared by the ASEAN (i.e. (Ali et al., 2010; Sukmajaya et al., 2017)) with two additional fields. The proposed fields for the model are:

(a) Country;
(b) Photograph;
(c) Botanical name;
(d) Vernacular names;
(e) Brief plant description;
(f) Propagation method;
(g) Geographical distribution/ecology;
(h) Chemical constituents;
(i) Reports on medical usage

  (i) uses supported by experimental/clinical data;
  (ii) uses in traditional medicine);

(iii) contraindications;
(iv) bibliography;

(j) Type of traditional medical claim and evidence to support that claim as specified in Table 10.1 above (i.e. as described in ASEAN (2017) Annex 7 Table 2).

We recognize that not all of this data may be available, and compromises will need to be made.

B. AI Process

We propose implementing a circular business process for plant investigations, copying the concepts of a traditional manufacturing-based circular economy model towards a business process model for investigations. One of the critical departures in this model from a traditional pharmaceutical R&D model is the informational output from the process. In a traditional process, some limited results are published, and pharmaceuticals are produced. Understandably, the publication of results is limited in scope and frequency due to commercial reasons; thus, we are not proposing this change. We, therefore, propose encouraging the publication of investigations where the outcome is not ideal. The result could be that the investigation produces no benefits, the side effects were too significant, or the research was orphaned due to changed priorities or lack of funding. We propose that researchers be actively encouraged to publish such outcomes to be fed back as input into the business process flowchart. Additionally, we propose that a similar business process flowchart be implemented for medicinal plant investigations, copying the concepts from a traditional manufacturing-based circular economy model towards a business process model for investigations.

A circular process flowchart for plant investigations is shown in Fig. 10.1.

In the Circular Economy, products are used and reused (Masterson & Shine, 2023). The concept is to reduce as far as possible the waste products associated with producing, reusing, and repairing after the completion of primary use, and then recycling in a manner that feeds back into

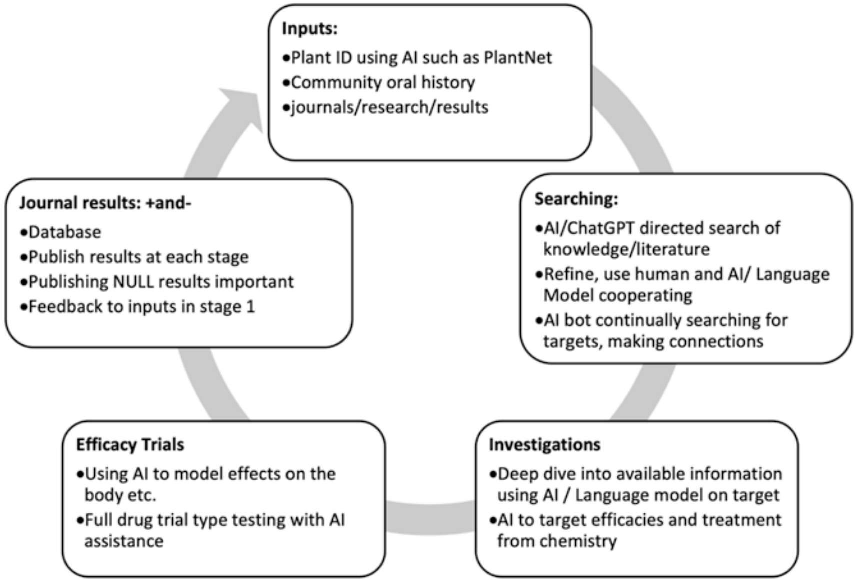

Fig. 10.1 Circular process flowchart for plant investigations. (Source: Author's own elaboration)

the raw materials required for producing new products. This model assumes that any transformation of materials is "expensive" with a cost, so reducing transformations reduces that expense, be it in time, energy, CO_2, pollutants, or any other consumable. The model also assumes that reusing and repairing items generally costs less than manufacturing, either from new or recycled raw materials.

With precious resources, ensuring the best possible return on investment is essential. Although there is little literature on the subject, it might be assumed that worldwide there is a significant waste of resources examining avenues for investigation and experimentation that other researchers have discounted. That is not to say that failed avenues should not be re-examined with a fresh eye. It is more to suggest that researchers should go down such a path armed with the knowledge that others have found it unsatisfactory. The researcher may then choose to verify the results themselves or find other areas for investigation. One issue with the publication of null results is how the publication of those results is funded, as costs of publishing in Open Access journals are being pushed onto researchers and their funding agencies, with funding agencies often paying a fee for manuscript publication. This pressure provides another incentive for researchers to publish null results.

A study of 96 European researchers by Echevarría et al. (2021) revealed that 79% of respondents considered that finding negative or unexpected results is not undesirable, and 82% thought that they should be shared among laboratories. Nevertheless, only 14% of them had tried to publish negative results, and interestingly, most had managed to publish their work. In a separate study, it was found that:

> [there is] a bias toward the publication of positive results. Not only were positive results more likely to be published, but studies that were not positive […] were often published in a way that conveyed a positive outcome. We analyzed these data in terms of the proportion of positive studies and in terms of the effect size associated with drug treatment. Using both approaches, we found that the efficacy of this drug class is less than would be gleaned from an examination of the published literature alone (Turner et al., 2008, p. 256).

In recent years there has been a move toward making it acceptable for researchers to publish non-significant results (Faculty of Science, University of Sydney, 2021). Unfortunately, the move towards this being the norm in scientific research is slow. Even websites promoting publications of null results, such as *alltrials.net*, appear to be infrequently updated.

Resources are often limited in developing and least developed countries, be it financial resources, lack of facilities such as laboratories, reliable transport for laboratory supplies, reliability of utilities such as water and electricity, availability of trained and skilled researchers, availability for service personnel for laboratory equipment, and even placements in tertiary training. Due to a lack of local educational opportunities, some developing countries select researchers to be trained outside the country using precious foreign exchange funds or borrowing funds such as from the Asian Development Bank.

With a circular process, reproducibility of results, positive or negative, is critical. This reproducibility, in turn, requires that data is available and shows the experimental setup, the data that was collected, and the analysis of the collected data.

C. Obstacles and Opportunities

The potential of AI in consolidating data on traditional and complementary treatments is significant, particularly if it can source reliable data on the efficacy of such treatments. To achieve this outcome, several obstacles need to be overcome. These include the ability to collect reliable information on the efficacy of treatments (including correct citation of the data), project creep which results in data overload, data being published in any number of languages, and intellectual property rights over the data sources and products developed from traditional knowledge.

(a) *Limitations of AI Technology in Drug Discovery*

One of the most critical pieces of information in drug discovery is efficacy data from reliable sources. To date, there are several known cases where the references cited by AI are incorrect and significant human intervention is required. We do not consider this an inherent flaw in the AI process. Instead, it shows the need for researchers to be diligent in their verification and information cross-checking, regardless of the source of the data.

One example of how this verification and research diligence can be done was illustrated by Blanco-González et al. (2022). In their research, the authors used AI to generate a paper on "*The role of AI in drug discovery: Challenges, opportunities, and strategies*" and then assessed the outcome. They found that ChatGPT requires strong human intervention to be able to write reliable scientific texts. It was found that it does not have the knowledge and expertise to be able to evaluate the veracity and reliability of the information it collects as part of its processes. They proposed three possible solutions for mitigating the risk associated with AI being used in the production of scientific publications. They suggested that the veracity of the information obtained by the AI algorithms could be enhanced by training them on large datasets of high-quality peer-reviewed research. They also suggested that the algorithms could be programmed to flag potentially problematic information that would alert researchers to undertake further review and verification. Particular attention should be paid to references to unreliable sources. Finally, "[a]nother approach would be to develop AI systems that are better able to evaluate the authenticity and reliability of the information they process" (Blanco-González et al., 2022, s 10).

Schneider (2018) considers that the combination of "big data" and "deep learning" per se does not solve these problems. Instead, it is the responsibility of the researchers "who devise appropriate representations of chemistry and biology for computational analysis" (Schneider, 2018, p. 109). The author goes on to mention, for the foreseeable future, the design, operation, and maintenance of such discovery platforms require an integrated team of skilled scientists, technicians, and engineers.

Moreover, Chu et al. (2022) claim that novel technologies in complementary and alternative medicines (CAM) may fail to be adopted without concrete evidence. It is vital that those undertaking research in CAM "verify the generalizability and validate the effectiveness and the reliability for the betterment of the promotion of CAM globally" (p. 13). They consider that it is vital for researchers who plan to conduct research using AI methods with CAM to verify the generalizability and validate the effectiveness and reliability for improving the promotion of CAM globally.

The challenge to using AI to discover efficacious traditional medicines is ensuring that the algorithms are as well trained as possible. Indeed, at the current state of development of AI, there must be an oversite of the outputs and an independent assessment of the results. What is being sought initially is potential efficacious targets and their related literature. Also, if localities where these plants grow can be identified, action can be taken to protect their source. In addition, the information can be used to prevent the patenting of traditional knowledge, as is one of the aims of the Traditional Knowledge Data Library (TKDL) of India.

(b) Potential Scope Creep

Scope creep occurs when the proponents continue to add additional requirements to the original design of a project. In this case there is a high potential for scope creep resulting in users being overwhelmed by data. This scope creep can occur by researchers increasing the number of data fields "because they can" rather than identifying data fields relevant to the task. In the case of the large traditional knowledge database developed by the CSIR in India, described above, the data checking and translation task would have required a significant resource allocation. Rather, the search should focus on critical elements. In the current proposal, the preferred scope is to extend the framework developed by Ali et al. (2010) and Sukmajaya et al. (2017) to cover the world's tropical regions. The output should highlight any discrepancies in the studies for further assessment by researchers in traditional and complementary medicines.

(c) Language

Language is a significant issue as the local knowledge is in the lingua franca of the local community. In addition, researchers are able to publish in reputable journals in their mother tongue. For instance, this is an issue amongst the ten members of the Association of Southeast Asian Nations (Smith, 2019). Five nations: Cambodia, Lao PDR, Myanmar, Thailand, and Vietnam, use unique language systems using a non-Latin script. Transliteration into English can also be, at times, problematic. Indonesia has its own Latin-script language, Bahasa Indonesia, which is the same as Bahasa Malaysia. The other four nations use English as well as their national language. In fact, in Brunei Darussalam, Malaysia, and Singapore, English is the language of the law. The working language of the ASEAN is English. The trend is for at least some academic/professional journals in each country to publish exclusively in English, with the exception of Singapore, where journals are usually published in English exclusively. The uniform naming of plants is covered by the Shenzhen Code (Turland et al., 2018). Even when a publication is published in the national language, an English abstract is often provided. References are, of course, in the language of the original publication. For example, the

Songklanakarin Journal of Plant Science publishes in Thai with a bi-lingual title and a bi-lingual abstract. An article on guava leaf extracts includes the Latin script botanical name in both the English and Thai abstracts and a keyword for each abstract (Keawdoung & Boonprakob, 2023). While most may not read a paper, it could be included in the literature associated with that species.

The process used for the Traditional Knowledge Data Library (TKDL) of India is far too complex and labor-intensive (Council of Scientific & Industrial Research (CSIR), 2023) for the minimal benefit of what the current project is seeking to achieve. It requires that each text is read with the medicinal formulations and practices are identified and converted into a structured language by subject matter experts in Ayurveda, Unani, Siddha, Sowa Rigpa or Yoga. It is a knowledge-based conversion into English, German, French Japanese and Spanish. As part of the process, traditional terminology is converted into modern terminology.

Even translation with software such as Google Translate is problematic, particularly with non-Latin scripts. A simple test was carried out using a common medicinal product, "neem oil." In Thai, it translated as น้ำมันสะเดา (N̂ảmạn s̄adeā), and when translated back from Thai to English, it came back as "neem oil"; however, translations into Khmer and Lao failed.

(d) Intellectual Property

Copyright is an area that is fraught with difficulties, and protections vary by jurisdiction. Two articles of the *WIPO Copyright Treaty (WCT)* (1996) are particularly relevant:

> Compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such. This protection does not extend to the data or the material itself and is without prejudice to any copyright subsisting in the data or material contained in the compilation (WCT art 5).

> Contracting Parties shall provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures that are used by authors in connection with the exercise of their rights

> under this Treaty or the Berne Convention and that restrict acts, in respect of their works, which are not authorized by the authors concerned or permitted by law (WCT art 11).

It is reasonable to assume that the sourced material is copyrighted unless explicitly stated otherwise. For example, jurisdictions like Australia do not require a copyright statement or logo to be provided for material to be copyrighted. Countries usually include a fair dealing clause in their copyright law. The clause would be similar to that in the Australian *Copyright Act* 1968, namely:

> A fair dealing with a literary, dramatic, musical or artistic work, or with an adaptation of a literary, dramatic or musical work, for the purpose of research or study does not constitute an infringement of the copyright in the work (Copyright Act s 40(1)).

While traditional knowledge per se cannot be protected, it is crucial that outside parties do not attempt to patent the knowledge. As discussed above, products derived from the commercialization of traditional knowledge may be protected by patents, trademarks, and geographical indications.

(e) Compliance and Regulation

Compliance and regulation of complementary medicines can become quite challenging. For instance, in Australia, Complementary Medical manufacturers often must provide "mandatory data" to the regulator for pre-market approval or post-market compliance assessments. In numerous regulatory circumstances, an applicant may be required by law to give information to support the pre-market (marketing approval) or post-market (compliance assessment) context (Complementary Medicines Australia, 2023). For example, an indication and claim that "an herbal ingredient has a tradition of use in Western Herbal Medicine for symptomatic cough relief" may be based on research reported in books or textbooks; published online; or possibly in monographs published by Government or private bodies in Australia or overseas (Complementary

Medicines Australia, 2023, p. 2). Many of the sources would be copyright material and cannot be legally shared without the approval of the copyright owner. As data collection in our case would be multi-jurisdictional, it would be prudent to seek appropriate legal advice before embarking on a comprehensive data mining exercise.

10.5 Discussion and Recommendations

The loss of traditional knowledge of plants and their medicinal properties would be a global tragedy as it covers indigenous and local knowledge and knowledge in the broader community. While artificial intelligence is still in its infancy, "big data" has existed for much longer. The tools are available for AI to identify efficacious traditional medicines, but several caveats exist, with the availability of data being the most pressing issue. Currently, much information is in the informal sector and may or may not be formally documented. This lack of documented information is also the case in the broader agricultural sector. To collect and codify that knowledge is a daunting exercise that requires a step-by-step approach, and the initiatives described above by ASEAN are a good start. However, the two compendia only include a sample of the plant sources for traditional medicine. Once the data is available, it should be actively made available to the marginalized communities, with advice on how it may be commercialized to their benefit. This knowledge-sharing should be a role for governmental organizations, NGOs, universities and the like rather than for-profit organizations (see Smith and Smith (2023) and Smith and Perry (2023a)).

An even more pressing issue is the availability of information that is not open-sourced. This lack of information is an ethical issue: should a researcher who is collecting and using traditional knowledge, which by its very nature cannot be patented or covered by copyright, be able to then publish the information as copyright material? Universities and research institutions should promote free access to information, especially when the research is publicly funded. Moreover, publications of negative results should also be encouraged so that future researchers can focus on areas more likely to provide positive outcomes. There would be incredible benefits if there could be an international consensus on collecting and storing traditional knowledge, including the sources of that knowledge, especially if it comes from an oral tradition. In particular, it would be invaluable if researchers and international organizations could reach a consensus on the critical data elements to be collected and how they are stored. Researchers, wherever their country of residence, should be strongly encouraged to include the botanical name of a plant in the Latin script regardless of the language of publication to enable more accessible access to the outcomes of their research.

Lastly, AI is not infallible, and detailed data quality checking is essential. This data quality checking is particularly important in the case when data sources are cited. As can readily be seen from a general search using a search engine such as Google, an error or misunderstanding on one site is often replicated across several sites. Artificial intelligence has great potential to identify efficacious traditional medicines for commercialization by marginalized communities in tropical regions, which would provide many benefits to these communities. Regardless of the sophistication of the AI algorithms, it is incumbent on the user to undertake detailed quality checks to ensure that the data is reliable and that the source is correctly identified.

Acknowledgements This research is supported by an Australian Government Research Training Program (RTP) Scholarship to the first author.

References

Ali, R. M., Samah, Z. A. S., Mustapha, N. M., & Hussein, N. (2010). *ASEAN herbal and medicinal plants*. ASEAN Secretariat.

ASEAN. (2017). *Draft ASEAN agreement on regulatory framework for traditional medicines*, https://asean.org/wp-content/uploads/2017/09/ASEAN-Guidelines-on-Claims-Claims-Substantiation-TM-V2.0-with-discla….pdf

ASEAN Centre for Biodiversity. (2019). *Traditional knowledge digital library*. Biodiversity Information Management Unit, ASEAN Centre for Biodiversity, Manila, tkdl.aseanbiodiversity.org

Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, Á., & Garcia-Fandino, R. (2022). The role of AI in drug discovery: Challenges, opportunities, and strategies, *arXiv*, https://arxiv.org/abs/2212.08104

Braun, L., & Cohen, M. (2015). *Herbs & natural supplements: An evidence-based guide* (Vol. 2, 4th ed.). Churchill Livingstone.

Calixto, J. B. (2019). The role of natural products in modern drug discovery. *Anais da Academia Brasileira de Ciências, 91*(Suppl 3). https://doi.org/10.1590/0001-3765201920190105

Chu, H., Moon, S., Park, J., Bak, S., Ko, Y. K., & Youn, B.-Y. (2022). Use of artificial intelligence in complementary and alternative medicine: A systematic scoping review. *Frontiers in Pharmacology, 13*(826044). https://doi.org/10.3389/fphar.2022.826044

Complementary Medicines Australia. (2023). *Submission,* https://consultations.ag.gov.au/rights-and-protections/copyright-enforcement-review/consultation/view_respondent?uuId=602126708

Copyright Act. (1968). *(Cth) (as amended and in force on 23 June 2017)* (Australia), https://www.legislation.gov.au/Details/C2017C00180/Download

Council of Scientific & Industrial Research (CSIR). (2023). *Bio-piracy of traditional knowledge*, https://www.tkdl.res.in/tkdl/langdefault/common/Biopiracy.asp?GL=Eng

Devine, S. N. O., Kolog, E. A., & Atinga, R. (2022). Toward a knowledge-based system for African traditional herbal medicine: A design science research approach [original research]. *Frontiers in Artificial Intelligence, 5*. https://doi.org/10.3389/frai.2022.856705

Echevarría, L., Malerba, A., & Arechavala-Gomeza, V. (2021). Researcher's perceptions on publishing 'negative' results and open access. *Nucleic Acid Therapeutics, 31*(3), 185–189. https://doi.org/10.1089/nat.2020.0865

European Commission. (2023). *Public health – 'Pharmacovigilance'*, https://health.ec.europa.eu/medicinal-products/pharmacovigilance_en#

Faculty of Science, University of Sydney. (2021, November 22). *Publication bias: Why null results are not necessarily 'dull' results,* https://www.sydney.edu.au/science/news-and-events/2021/11/22/publication-bias-why-null-results-are-not-necessarily-dull-re.html

Inside Precision Medicine. (2023, May 25). "AI reveals new antibiotic", *Genetic Engineering & Biotechnology News*, https://www.insideprecisionmedicine.com/topics/patient-care/therapeutics/antibiotics/ai-reveals-new-antibiotic/

Keawdoung, M., & Boonprakob, U. (2023). Total phenolic content and antioxidant activity of guava leaf extracts from three Psidium species.

Songklanakarin Journal of Plant Science, 10(1) https://www.sjplantscience.com/index.php/ojs/article/view/73/37

Li, B., Ma, C., Zhao, X., Hu, Z., Du, T., Xu, X., Wang, Z., & Lina, J. (2018). YaTCM: Yet another traditional Chinese medicine database for drug discovery. *Computational and Structural Biotechnology Journal, 16*, 600–610. https://doi.org/10.1016/j.csbj.2018.11.002

Masterson, V., & Shine, I. (2023, March 10). *What is the circular economy, and why does it matter that it is shrinking?* World Economic Forum, Geneva, https://www.weforum.org/agenda/2022/06/what-is-the-circular-economy/

Pourchez, L. (2017). *Women's knowledge: Traditional medicine and nature – Mauritius, Reunion and Rodrigues*. Local & Indigenous Knowledge, UNESCO, https://unesdoc.unesco.org/ark:/48223/pf0000235483

Rainforest Alliance. (2023). *Deforestation*, https://www.rainforest-alliance.org/issues/forests/deforestation/

Robinson, D., & Raven, M. (2019). 'Ethical trade in natural products based on traditional knowledge', *Trade for Development News*, https://trade4devnews.enhancedif.org/en/news/ethical-trade-natural-products-based-traditional-knowledge

Schneider, G. (2018). Automating drug discovery. *Nature Reviews, Drug Discovery, 17*, 97–113. https://doi.org/10.1038/nrd.2017.232

Smith, R. B. (2019). 'Harmonization of Laws in ASEAN: The issue of English', *International Seminar on Politics, Administration and Development 2019 (INSPAD2019)*, Walailak University, https://www.researchgate.net/publication/343178725_CODE_026_HARMONISATION_OF_LAWS_IN_ASEAN_THE_ISSUE_OF_LANGUAGE

Smith, R. B., & Perry, M. (2023a). Protection & promotion of traditional knowledge of marginalized communities – A legal perspective. *Proceedings of 24th European Conference on Knowledge Management – ECKM 2023*. (in press).

Smith, R. B., & Perry, M. (2023b). University extension programs to develop intellectual property in the agriculture sector of marginalized rural communities in the ASEAN region. In T. Walker, K. Tarabieh, S. Goubran, & G. Machnik-Kekesi (Eds.), *Sustainable practices in higher education: Finance, strategy, and engagement* (pp. 143–163). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-27807-5_8

Smith, N. N., & Smith, R. B. (2023). *University extension programs to promote the traditional knowledge of marginalised rural communities in the ASEAN region* (Vol. 18, p. 839). Proceedings of the 18th European Conference on Innovation and Entrepreneurship, ECIE 2023 (in press).

Sukmajaya, D., Permatasari, A., Mohtar, M., Kiong, L. S., Badron, U. H., & Ali, R. M. (2017). *ASEAN herbal and medicinal plants* (Vol. II). ASEAN Secretariat.

Suwankesawong, W. (2019). *Pharmacovigilance and traditional and complementary medicine in South-East Asia: A situation review*. World Health Organization, Regional Office for South-East Asia, New Delhi.

Tripathi, M. K., Nath, A., Singh, T. P., Ethayathulla, A. S., & Punit, K. (2021). Evolving scenario of big data and artificial intelligence (AI) in drug discovery. *MoI Divers, 25*(3), 1439–1460. https://doi.org/10.1007/s11030-021-10256-w

Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T. W., McNeill, J., Monro, A. M., Prado, J., Price, M. J., & Smith, G. F. (Eds.). (2018). *International code of nomenclature for algae, fungi, and plants (Shenzhen code) adopted by the nineteenth international botanical congress Shenzhen, China, July 2017*. Koeltz Botanical Books. https://doi.org/10.12705/Code.2018

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine, 358*, 252–260. https://doi.org/10.1056/NEJMsa065779

WIPO Copyright Treaty (WCT). (1996). https://wipolex.wipo.int/en/treaties/textdetails/12740

Workman, P., Antolin, A. A., & Al-Lazikani, B. (2019). 'Transforming cancer drug discovery with big data and AI' [editorial]. *Expert Opinion on Drug Discovery, 14*(11), 1089–1095. https://doi.org/10.1080/17460441.2019.1637414

World Intellectual Property Organization (WIPO). (2023). *Traditional knowledge*, https://www.wipo.int/tk/en/tk/

Index[1]

[1] Note: Page numbers followed by 'n' refer to notes.

T. Walker et al. (eds.), *Artificial Intelligence for Sustainability*,
https://doi.org/10.1007/978-3-031-49979-1

B