

Business Guides on the Go

Josef Baker-Brunnbauer

---

# Trustworthy Artificial Intelligence Implementation

---

Introduction to the TAI  
Framework

 Springer

## Business Guides on the Go

“Business Guides on the Go” presents cutting-edge insights from practice on particular topics within the fields of business, management, and finance. Written by practitioners and experts in a concise and accessible form the series provides professionals with a general understanding and a first practical approach to latest developments in business strategy, leadership, operations, HR management, innovation and technology management, marketing or digitalization. Students of business administration or management will also benefit from these practical guides for their future occupation/careers.

These Guides suit the needs of today’s fast reader.

Josef Baker-Brunnbauer

# Trustworthy Artificial Intelligence Implementation

Introduction to the TAll Framework

 Springer

Josef Baker-Brunnbauer   
Leonding, Austria

ISSN 2731-4758

ISSN 2731-4766 (electronic)

Business Guides on the Go

ISBN 978-3-031-18274-7

ISBN 978-3-031-18275-4 (eBook)

<https://doi.org/10.1007/978-3-031-18275-4>

© The Editor(s) (if applicable) and The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

# Prelude

This book addresses the management awareness about ethical and moral aspects of Artificial Intelligence (AI). It is a general trend to speak about AI and many start-ups and established companies are communicating about the development and implementation of AI solutions. There are many events and marketing workshops about AI solutions that are often driven by the development of the technology. AI systems have immense potential to change many domains like the education system, legal and law system, retail market, whole industries, and societies. Therefore, it is important to consider different perspectives besides the technology. As data is one key element for AI, data protection and legal implementation will change. The way in which societies are interacting and organizing themselves will be affected. Such changes require a multi-perspective of the humanity for shaping the future. Will AI systems promote the high-developed nations, or will it take care of the poorest? The development of AI systems is not only a technical discipline: it requires an interaction between multi-professions. This book is aiming to overcome those barriers with the results of a fundamental literature and empirical research to answer the question: *What kind of awareness does the management have about the social impact of their AI product or service?*

This question is divided into five sub-questions that will be answered by a fundamental literature and an empirical research study. This covers the management understanding of the terms moral, ethics, and artificial

intelligence, the internal company prioritization of moral and ethics, the involved stakeholders in the AI product or service development and it will analyze the known and used ethical AI guidelines and principles. In the end, the social responsibility of the management regarding their AI system is analyzed and compared. This research has not the aim to discuss AI on a technical level, it is analyzing the management awareness about their social impact to shape the future.

Organizations and companies need practical tools and guidelines to kick-off the implementation of Trustworthy Artificial Intelligence (TAI) systems. AI development companies are still in the beginning of this process or have not even started yet. The findings of the research address to decrease the entry level barrier for AI ethics implementation by introducing the Trustworthy Artificial Intelligence Implementation (TAII) Framework. The outcome is comparatively unique given that it considers a meta perspective of implementing TAI within organizations. As such, the framework aims to fill a literature gap for management guidance to tackle trustworthy AI implementation while considering ethical dependencies within the company. The TAII Framework<sup>\*</sup> takes a holistic approach to identify the systemic relationships of ethics for the company ecosystem and considers corporate values, business models, and common good aspects like the Sustainable Development Goals and the Universal Declaration of Human Rights. The TAII Framework creates guidance to initiate the implementation of AI ethics in organizations without requiring a deep background in philosophy and considers the social impacts outside of a software and data engineering setting. Depending on the legal regulation or area of application, the TAII Framework can be adapted and used with different regulations and ethical principles.

# Acknowledgement

I highly appreciate the support of all participants and companies in the empirical research study and all people who shared their thoughts, opinions, and time with me. Moreover, I want to thank Prof. Dr. Joanna Bryson, Prof. Dr. Virginia Dignum, and Prof. Dr. Sandra Wachter for inspiration, professional feedback, and motivation. Ethics of Artificial Intelligence is and will be a multi-profession topic that is generating strong impact for societies.

Finally, I also want to thank my wife Nathalie for her support and understanding of being short of time.



# Contents

|          |                                  |    |
|----------|----------------------------------|----|
| <b>1</b> | <b>Introduction</b>              | 1  |
| 1.1      | The Challenge                    | 3  |
| 1.2      | Course of Action                 | 4  |
|          | References                       | 5  |
| <b>2</b> | <b>Artificial Intelligence</b>   | 7  |
| 2.1      | Terms and Definitions            | 7  |
| 2.2      | Understanding of the Technology  | 10 |
| 2.3      | Current and Near Future State    | 12 |
| 2.4      | Challenges                       | 13 |
| 2.5      | Opportunities and Risks          | 15 |
| 2.6      | Goals                            | 24 |
| 2.7      | Human Life and Behavior          | 25 |
| 2.8      | Chapter Summary                  | 26 |
|          | References                       | 27 |
| <b>3</b> | <b>Ethics and Moral</b>          | 31 |
| 3.1      | How to Train Ethics?             | 35 |
| 3.2      | Influence on Product Development | 37 |
| 3.3      | Ethical and Moral Challenges     | 39 |

|          |  |           |
|----------|--|-----------|
| 3.4      | Guidelines and Principles                  | 42        |
| 3.5      | Summary                                    | 51        |
|          | References                                 | 51        |
| <b>4</b> | <b>Business and Management Perspective</b> | <b>55</b> |
| 4.1      | Current State of AI Implementations        | 55        |
| 4.2      | Business Strategy                          | 56        |
| 4.3      | Management Decisions                       | 58        |
| 4.4      | Summary                                    | 59        |
|          | References                                 | 60        |
| <b>5</b> | <b>Empirical Study</b>                     | <b>61</b> |
| 5.1      | Research Methodology                       | 62        |
| 5.2      | Definition of the Material                 | 63        |
|          | References                                 | 66        |
| <b>6</b> | <b>Findings</b>                            | <b>69</b> |
| 6.1      | Definitions                                | 69        |
| 6.1.1    | Moral                                      | 70        |
| 6.1.2    | Ethics                                     | 70        |
| 6.1.3    | Artificial Intelligence                    | 70        |
| 6.2      | Prioritization                             | 71        |
| 6.3      | Stakeholders                               | 72        |
| 6.4      | Guidelines and Principles                  | 73        |
| 6.4.1    | Advantages of Ethical AI Guidelines        | 73        |
| 6.4.2    | Doubts About Ethical AI Guidelines         | 74        |
| 6.4.3    | Guideline Proposals                        | 74        |
| 6.4.4    | Ethical AI Guideline Creators              | 74        |
| 6.4.5    | Advantages of AI System Certification      | 76        |
| 6.4.6    | Doubts About AI System Certification       | 76        |
| 6.5      | Social Responsibility                      | 76        |
| 6.5.1    | Ethical Aspects of AI Systems              | 77        |
| 6.5.2    | Environmental Impact                       | 77        |
| 6.5.3    | Social Impact                              | 78        |
| 6.5.4    | Societal Impact                            | 79        |
| 6.5.5    | Legal System Adaption                      | 80        |

|          |                                  |     |
|----------|----------------------------------|-----|
| 6.6      | Responsibility of AI Systems     | 80  |
| 6.6.1    | Transparency                     | 81  |
| 6.6.2    | Robustness                       | 81  |
| 6.6.3    | Bias                             | 82  |
| 6.6.4    | Privacy and Security             | 83  |
| 6.6.5    | Accountability                   | 83  |
| 6.7      | Management Reflection            | 84  |
| 6.8      | Findings and Recommendations     | 85  |
|          | Reference                        | 96  |
| <b>7</b> | <b>TAII Framework</b>            | 97  |
| 7.1      | Introduction                     | 98  |
| 7.2      | Trustworthy AI                   | 101 |
| 7.3      | Overview of the TAI Framework    | 102 |
| 7.4      | Business Model AI                | 108 |
| 7.5      | Implementation Challenges of TAI | 113 |
| 7.6      | TAII Framework Canvas            | 117 |
| 7.7      | Conclusion                       | 119 |
|          | References                       | 121 |

## About the Author



**Josef Baker-Brunnbauer** worked for over 20 years in different industry areas for start-ups and established companies in international projects for market leading clients. He is consulting companies about business model innovation, product innovation, change and digital transformation projects through his consulting company [product-xyz.com](http://product-xyz.com) and leading research projects for [SocialTechLab.eu](http://SocialTechLab.eu). The research focus is on socio-technical systems, trustworthy artificial intelligence, and its social impact. Besides, he is a member of the AI Alliance at the European Commission. He holds a master's degree in international business administration (MBA) from LIMAK Johannes Kepler University Linz/Austria and Tsinghua University Beijing/China, an engineering degree in software engineering (Ing), a diploma in life counseling and coaching (LSB), in enterprise design thinking from IBM, in product management (FH) from Marketing Academy Munich and University of Applied Sciences Upper Austria, and a master's degree for psychosocial counseling (MSc) from the Karl-

**xiv      About the Author**

Franzens University in Graz/Austria. In his leisure time, he likes to travel, discover new opportunities outside his comfort zone, cooking, and endurance sports. Thoughts and feedback are welcome: [josef@bakerbrunnbauer.com](mailto:josef@bakerbrunnbauer.com)

# Abbreviations

|         |  |
|---------|--|
| AAAI    | Association for the Advancement of Artificial Intelligence |
| AI      | Artificial Intelligence                                    |
| AI HLEG | Artificial Intelligence High-Level Expert Group            |
| AGI     | Artificial General Intelligence                            |
| AWS     | Autonomous Weapon System                                   |
| BM      | Business Model   |
| BMI     | Business Model Innovation                                  |
| CE      | Conformité Européenne                                      |
| CHAR    | Character  |
| DL      | Deep Learning  |
| DNA     | Deoxyribonucleic Acid                                      |
| EC      | European Commission  |
| EI      | Emotional Intelligence                                     |
| FDA     | Food and Drug Administration                               |
| FLOPS   | Floating-point operations per second                       |
| fMRI    | Functional Magnetic Resonance Imaging                      |
| FRA     | The European Union Agency for Fundamental Rights           |
| GDPR    | General Data Protection Regulation                         |
| GOF AI  | Good Old-Fashioned AI                                      |
| GPT-3   | Generative Pretrained Transformer 3                        |
| ISS     | International Space Station                                |
| IT      | Information Technology                                     |
| IoT     | Internet of Things   |