

Models for Classifying AI Systems: the Switch, the Ladder, and the Matrix

Jakob Mökander

Oxford Internet Institute, University of Oxford
jakob.mokander@oii.ox.ac.uk

David S. Watson

Department of Statistical Science, University College
London
david.s.watson11@gmail.com

Margi Sheth

R&D Data Office, AstraZeneca plc
margi.sheth@astrazeneca.com

Luciano Floridi

Oxford Internet Institute, University of Oxford, and
Department of Legal Studies, University of Bologna
luciano.floridi@oii.ox.ac.uk

ABSTRACT

Organisations that design and deploy systems based on artificial intelligence (AI) increasingly commit themselves to high-level, ethical principles. However, there still exists a gap between principles and practices in AI ethics. A major obstacle to operationalise AI Ethics is the lack of a well-defined material scope. Put differently, the question to which systems and processes AI ethics principles ought to apply remains unanswered. Of course, there exists no universally accepted definition of AI, and different systems pose different ethical challenges. Nevertheless, pragmatic problem-solving demands that things should be sorted so that their grouping will promote successful actions for some specific end. In this article, we review and compare previous attempts to classify AI systems for the practical purpose of implementing AI governance in practice. We find that attempts to classify AI systems found in previous literature use one of three mental models: the Switch, i.e., a binary approach according to which systems either are or are not considered AI systems depending on their characteristics; the Ladder, i.e., a risk-based approach that classifies systems according to the ethical risks they pose; and the Matrix, i.e., a multi-dimensional classification of systems that take various aspects into account, such as context, data input, and decision-model. Each of these models for classifying AI systems comes with its own set of strengths and weaknesses. By conceptualising different ways of classifying AI systems into simple mental models, we hope to provide organisations that design, deploy, or regulate AI systems with the conceptual tools needed to operationalise AI governance in practice.

ACM Reference Format:

Jakob Mökander, Margi Sheth, David S. Watson, and Luciano Floridi. 2022. Models for Classifying AI Systems: the Switch, the Ladder, and the Matrix. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3531146.3533162>

CCS CONCEPTS

• **Management of computing and information systems;** •
Codes of ethics; • **Socio-technical systems;**

KEYWORDS

Artificial intelligence, Ethics, Governance, Material scope

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

FAccT '22, June 21–24, 2022, Seoul, Republic of Korea

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9352-2/22/06.

<https://doi.org/10.1145/3531146.3533162>