Wie regulieren wir Künstliche Intelligenz?

AI – Definition und AI Audit

Workshop Deutscher EDG-Gerichtstag, Universität Ulm

15. Dezember 2023

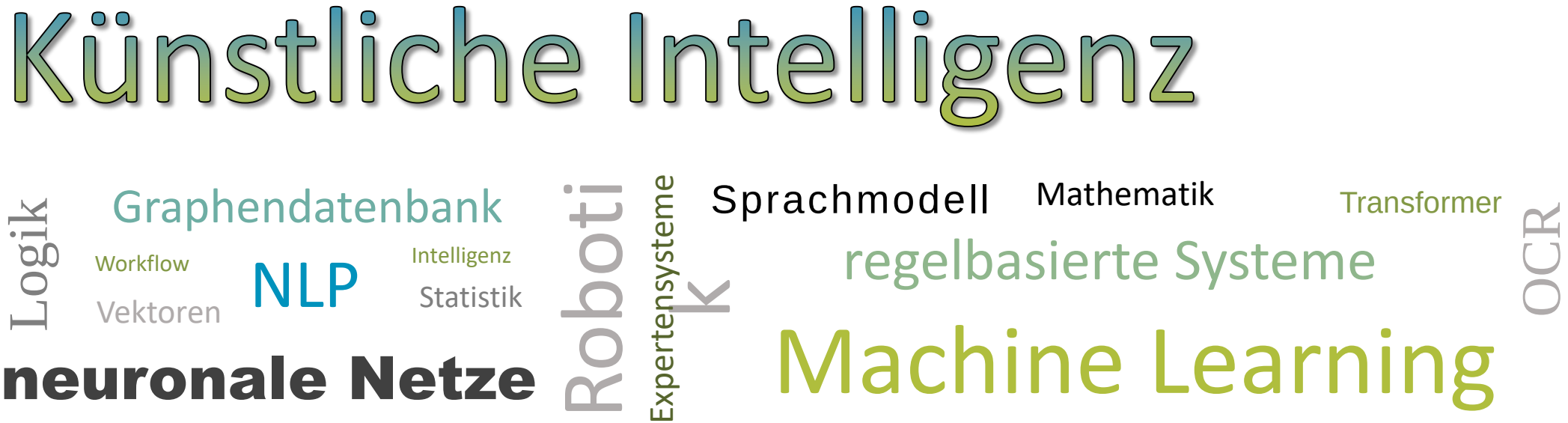
Dr. Stefan Eder

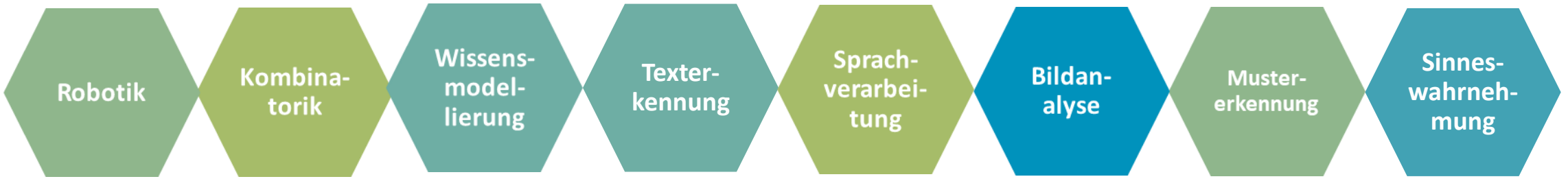
BENN-IBLER

CYBLY

Überblick

KI ist ...





AI Begriffsdefinition

“Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.
AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).”

Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final

EU 2018

EU 2019 überarbeitet

AI Begriffsdefinition

Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

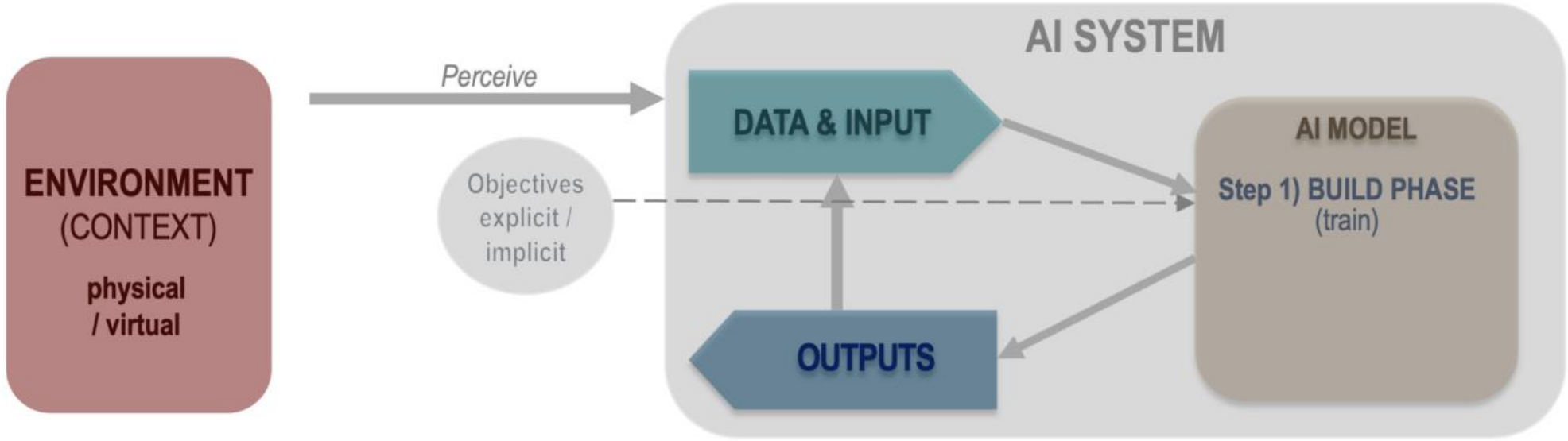
As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).”

INDEPENDENT HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION,

OECD revised

BUILD PHASE:

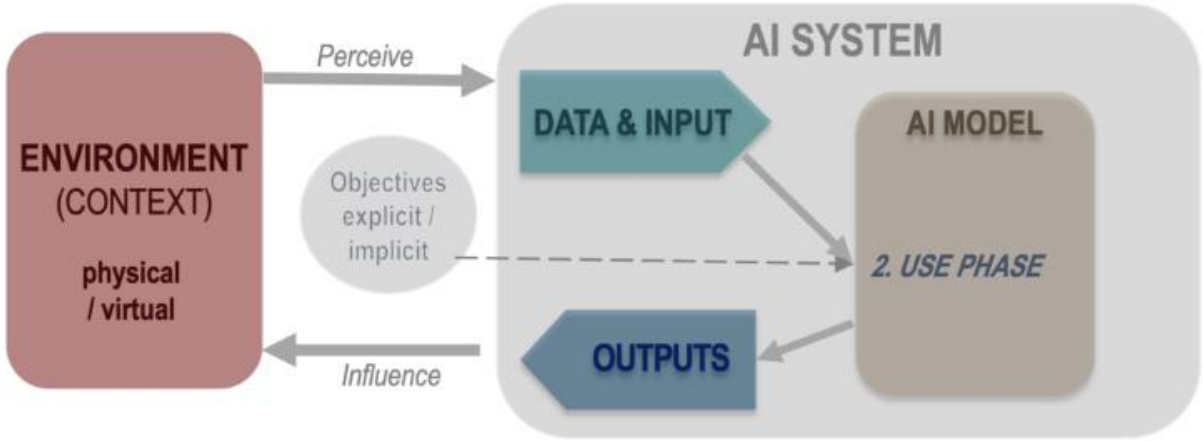
An AI system is a **machine-based** system, that



- for **explicit or implicit objectives**
- **infers**, from the **input** it receives
- How to **generate outputs** such as predictions, content, recommendations, or decisions

OECD revised

USE PHASE (once the model is built):
An AI system is a **machine-based** system, that



- for explicit or implicit objectives
- infers, from the input it receives
- How to generate outputs such as predictions, content, recommendations, or decisions
- **that [can] influence physical or virtual environments;**

Different AI systems vary in their levels of autonomy and adaptiveness [after deployment].

OECD AI system model: use phase

OECD revised

An AI system is a machine-based system that ~~can~~, for ~~a given set of human-defined~~ ***explicit or implicit*** *objectives,* ***infers, from the input it receives, how to generate outputs such as*** *~~makes~~ predictions,* ***content****, recommendations, or decisions* ***that can*** *influenc****e~~ing~~*** ***physical*** *~~real~~ or virtual environments.* ***Different*** *AI systems ~~are designed to operate with~~ vary~~ing~~* ***in their*** *levels of autonomy* ***and adaptiveness after deployment***

https://oecd.ai/en/wonk/ai-system-definition-update

Science

Artificial Intelligence Definitions

Stanford University
Human-Centered
Artificial Intelligence

Intelligence might be defined as the ability to learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world. A fully pre-programmed factory robot is flexible, accurate, and consistent but not intelligent.

Artificial Intelligence (AI), a term coined by emeritus Stanford Professor John McCarthy in 1955, was defined by him as "the science and engineering of making intelligent machines". Much research has humans program machines to behave in a clever way, like playing chess, but, today, we emphasize machines that can learn, at least somewhat like human beings do.

Autonomous systems can independently plan and decide sequences of steps to achieve a specified goal without micro-management. A hospital delivery robot must autonomously navigate busy corridors to succeed in its task. In AI, autonomy doesn't have the sense of being self-governing common in politics or biology.

Machine Learning (ML) is the part of AI studying how computer agents can improve their perception, knowledge, thinking, or actions based on experience or data. For this, ML draws from computer science, statistics, psychology, neuroscience, economics and control theory.

In **supervised learning**, a computer learns to predict human-given labels, such as dog breed based on labeled dog pictures; **unsupervised learning** does not require labels, sometimes making its own prediction tasks such as trying to predict each successive word in a sentence; **reinforcement learning** lets an agent learn action sequences that optimize its total rewards, such as winning games, without explicit examples of good techniques, enabling autonomy.

Deep Learning is the use of large multi-layer **(artificial) neural networks** that compute with continuous (real number) representations, a little like the hierarchically organized neurons in human brains. It is currently the most successful ML approach, usable for all types of ML, with better generalization from small data and better scaling to big data and compute budgets.

An **algorithm** lists the precise steps to take, such as a person writes in a computer program. AI systems contain algorithms, but often just for a few parts like a learning or reward calculation method. Much of their behavior emerges via learning from data or experience, a sea change in system design that Stanford alumnus Andrej Karpathy dubbed **Software 2.0**.

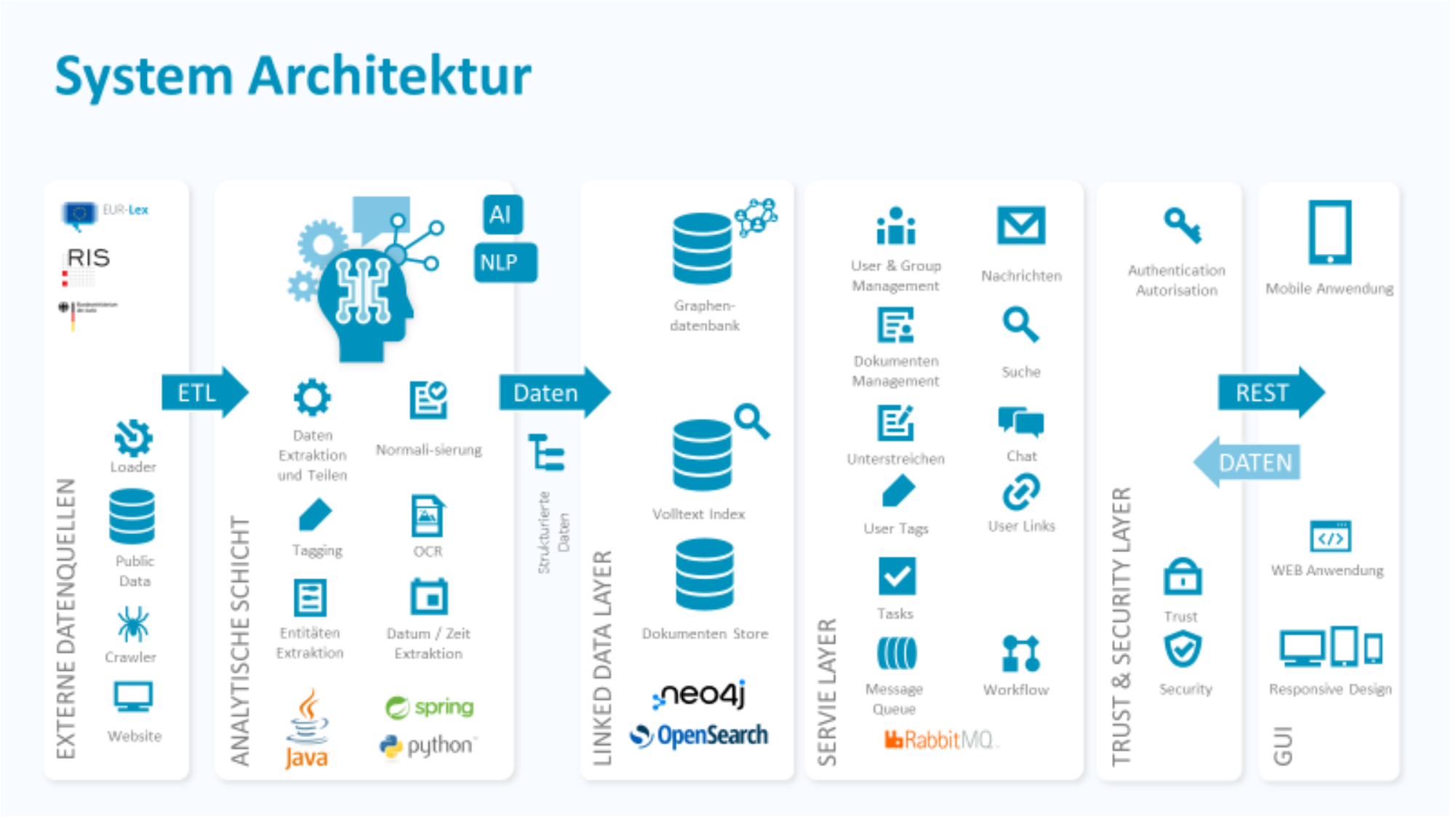
Narrow AI is intelligent systems for one particular thing, e.g., **speech** or **facial recognition**. **Human-level AI**, or **Artificial General Intelligence (AGI)**, seeks broadly intelligent, context-aware machines. It is needed for effective **social chatbots** or **human-robot interaction**.

Human-Centered Artificial Intelligence is AI that seeks to augment the abilities of, address the societal needs of, and draw inspiration from human beings. It researches and builds effective partners and tools for people, such as a robot helper and companion for the elderly.

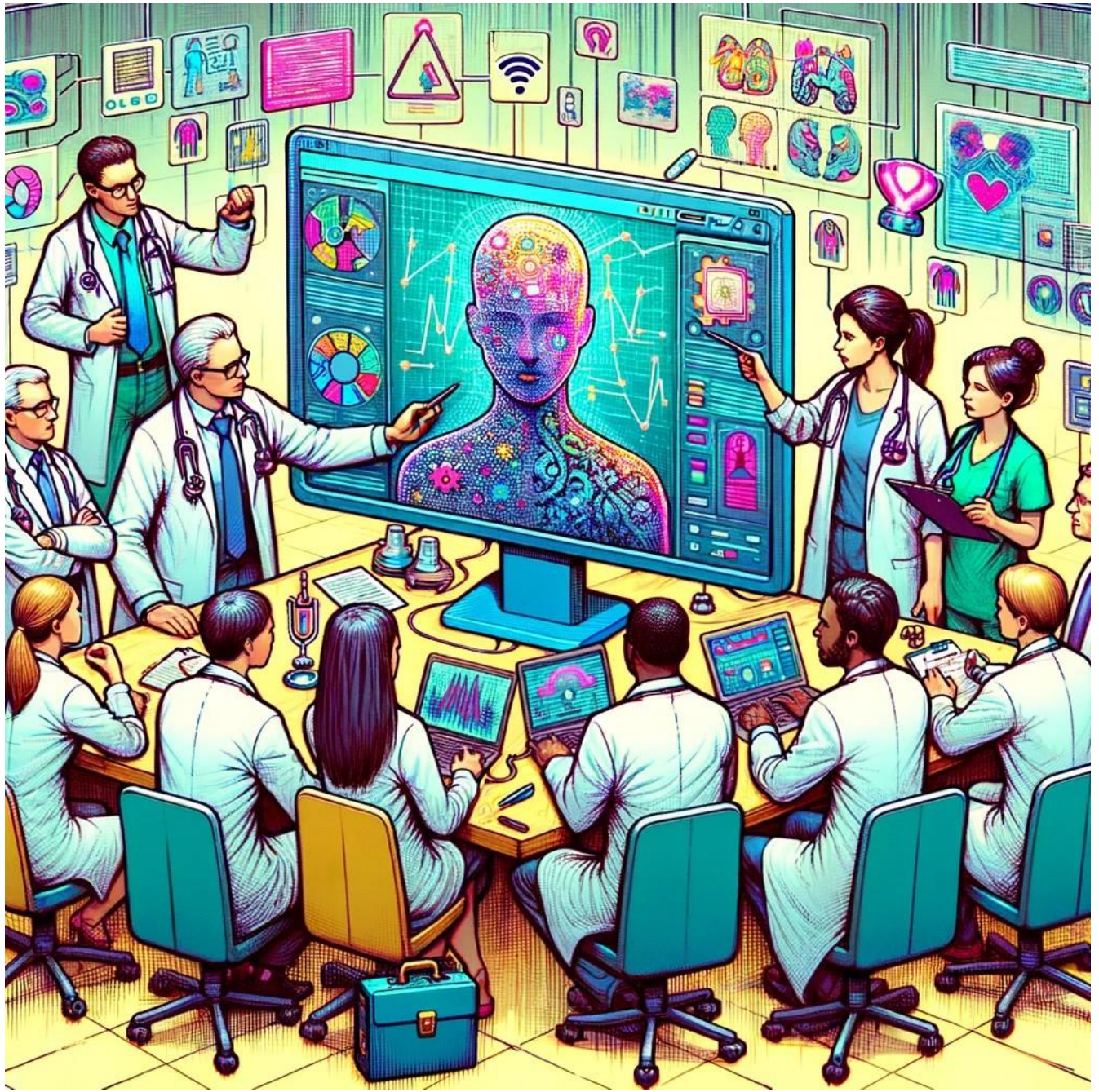
Text by Professor Christopher Manning, September 2020

Datenextrahierung

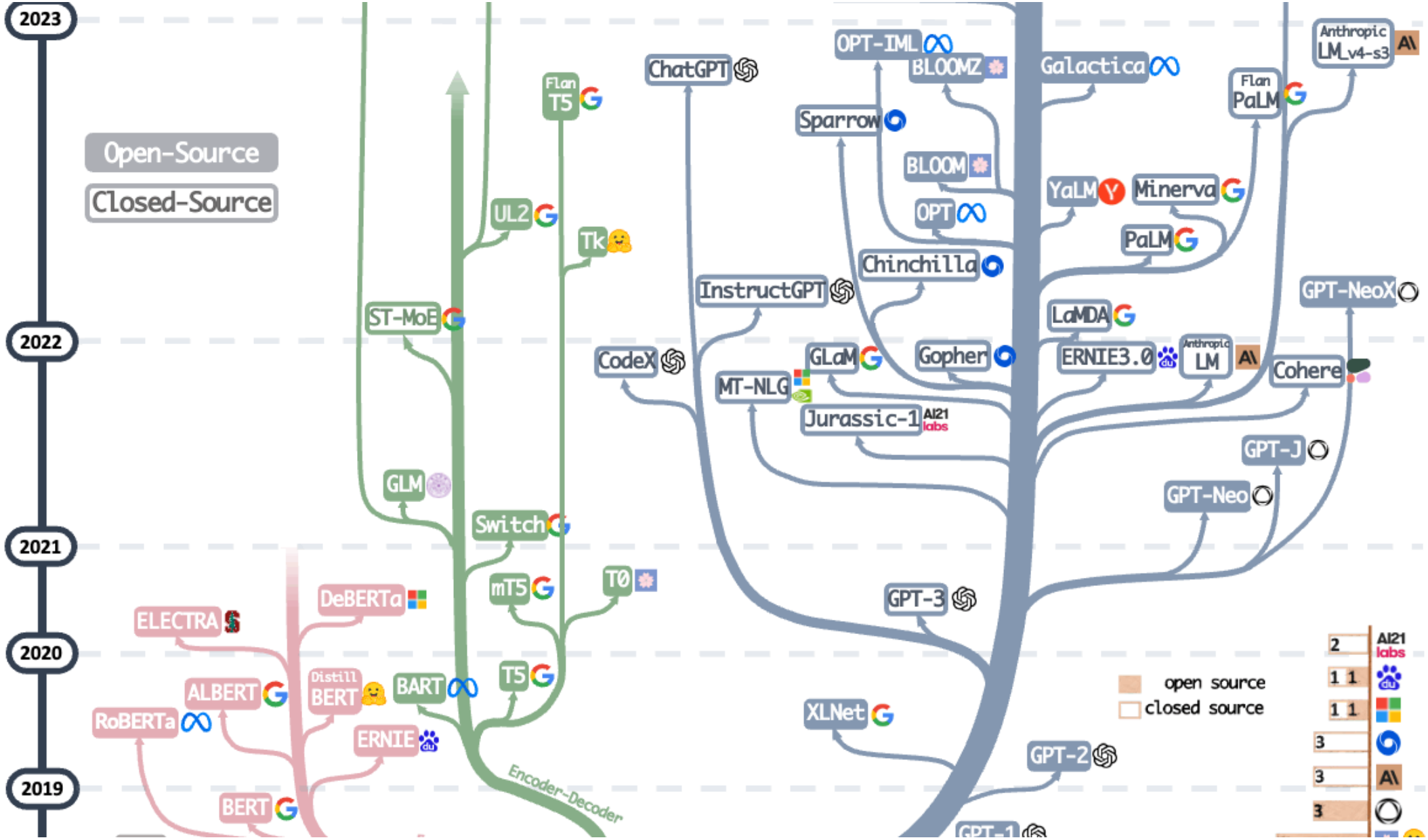
und was braucht es dazu – Beispiel Cybly



Audit AI Modelle?

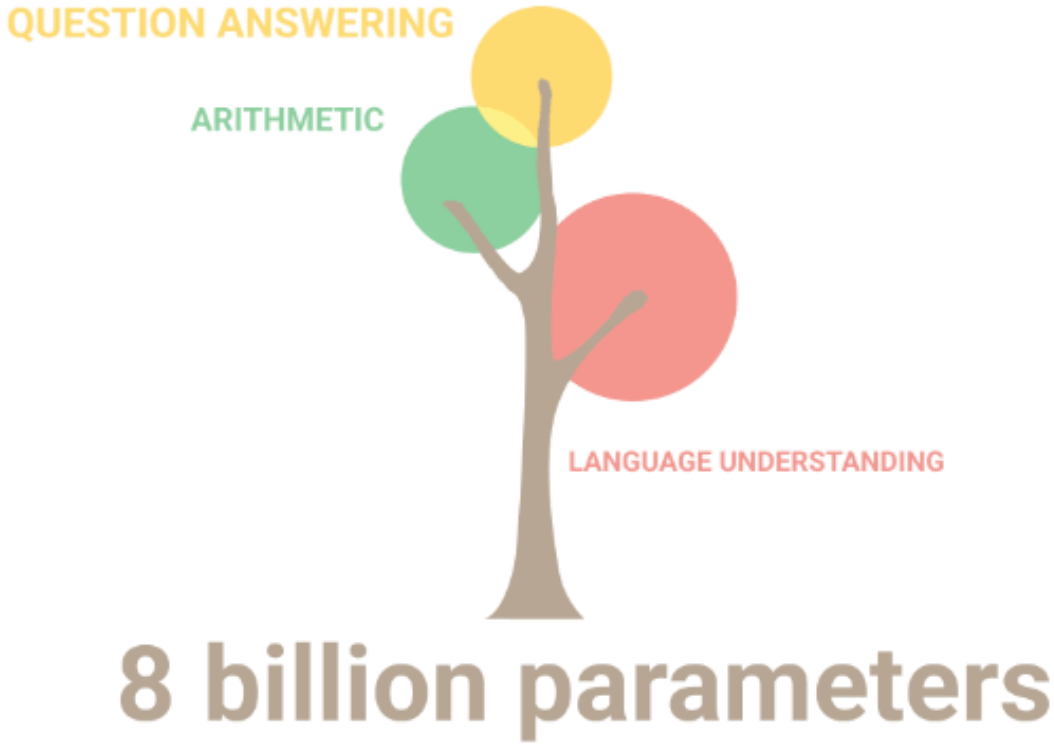


Evolution of LLMs



Quelle: https://github.com/Hannibal046/Awesome-LLM

Parameter LLMs



Quelle: https://github.com/Hannibal046/Awesome-LLM

Audit AI

Audit AI Development – Algorithmus Audit

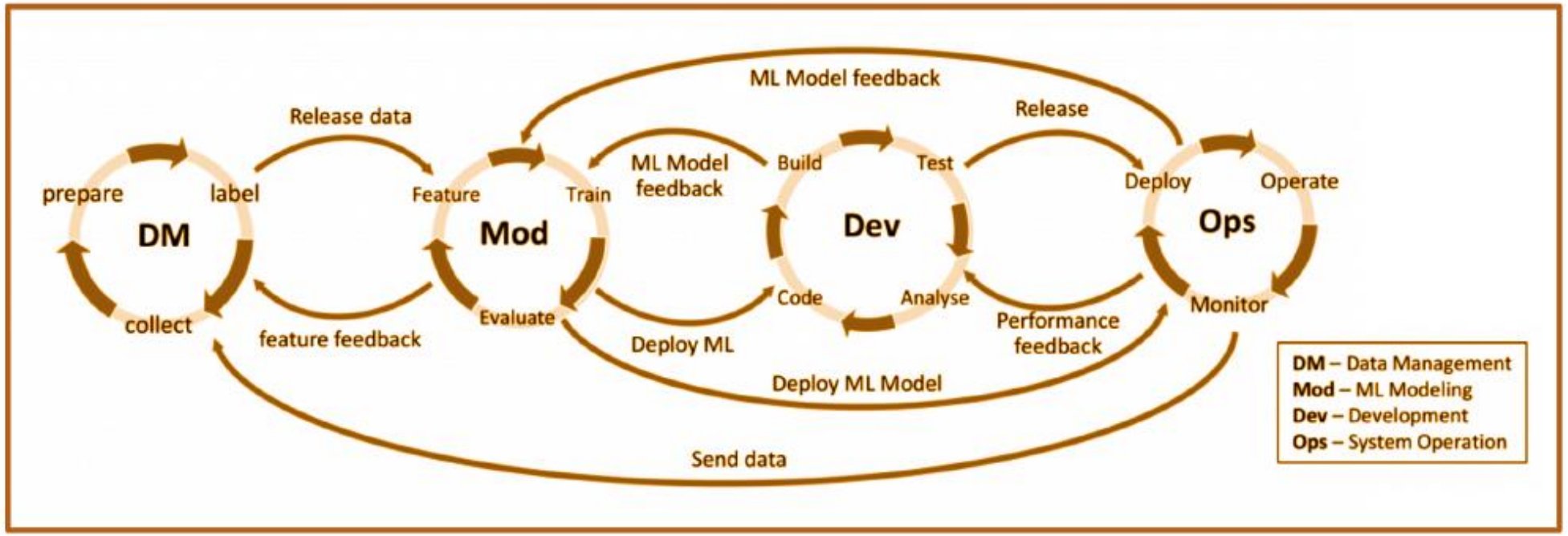


Fig. 1. Four phases of AI application development: Data Management, Model Selection, Development, and Operation.

Audit and Assurance of AI Algorithms: A framework to ensure ethical algorithmic practices in Artificial Intelligence, Ramya Akula, Ivan Garibay

AI Audit

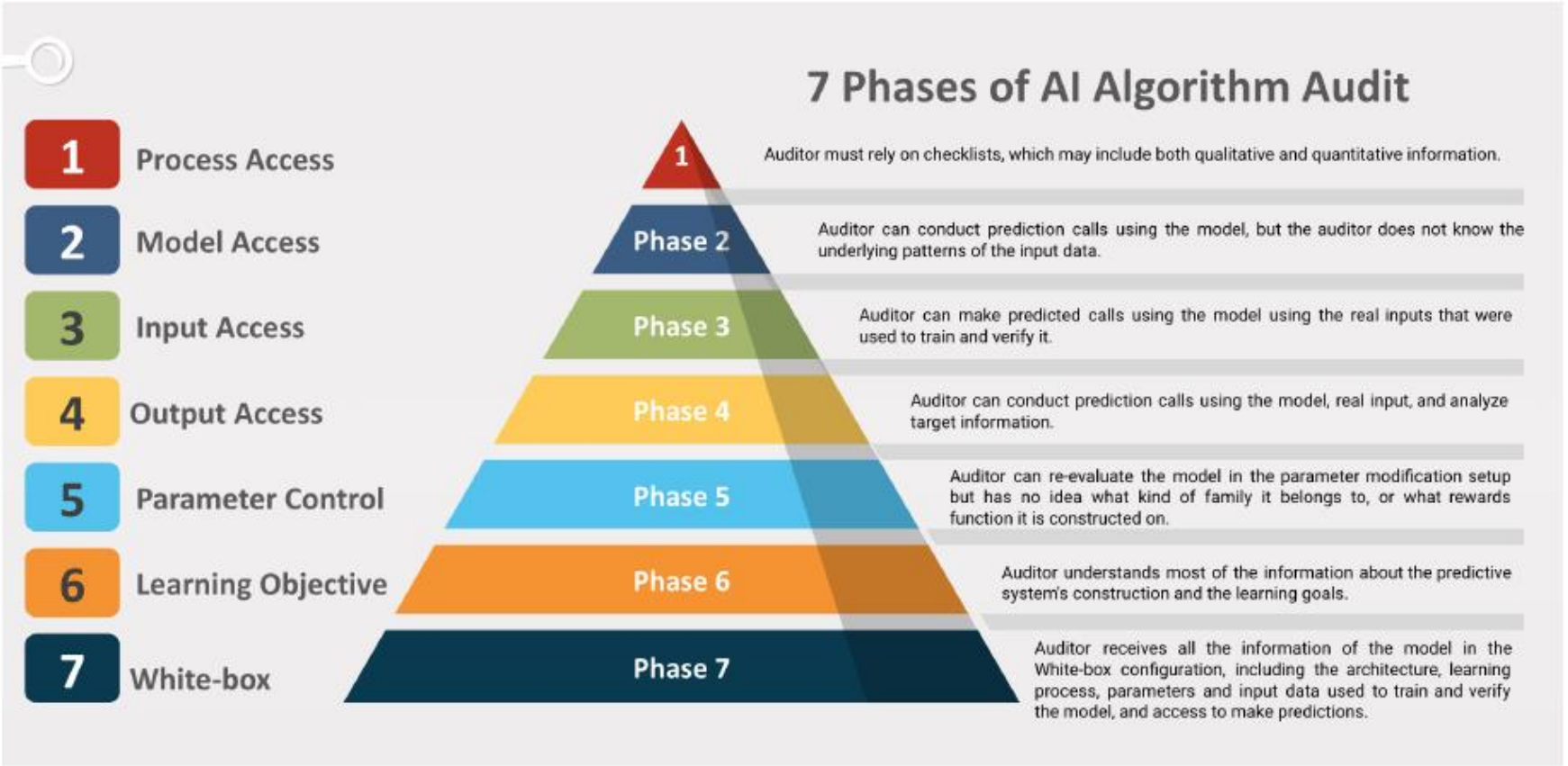


Fig. 2. Seven potential phases for the AI Algorithm Audit. In each phase, an auditor has various degrees of access to conduct legitimate check.

Audit and Assurance of AI Algorithms: A framework to ensure ethical algorithmic practices in Artificial Intelligence, Ramya Akula, Ivan Garibay

AI Audit

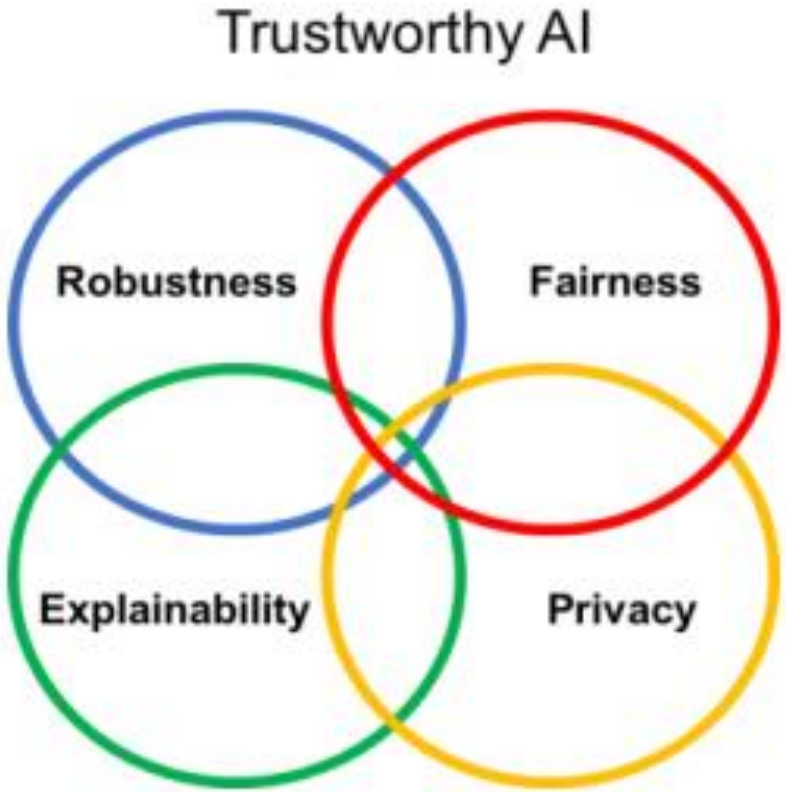


Figure 5: The overlaps between Algorithm Robustness, Fairness, Explainability and Privacy.

Towards Algorithm Auditing -A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms*
Adriano Koshiyama et.al.

AI Audit

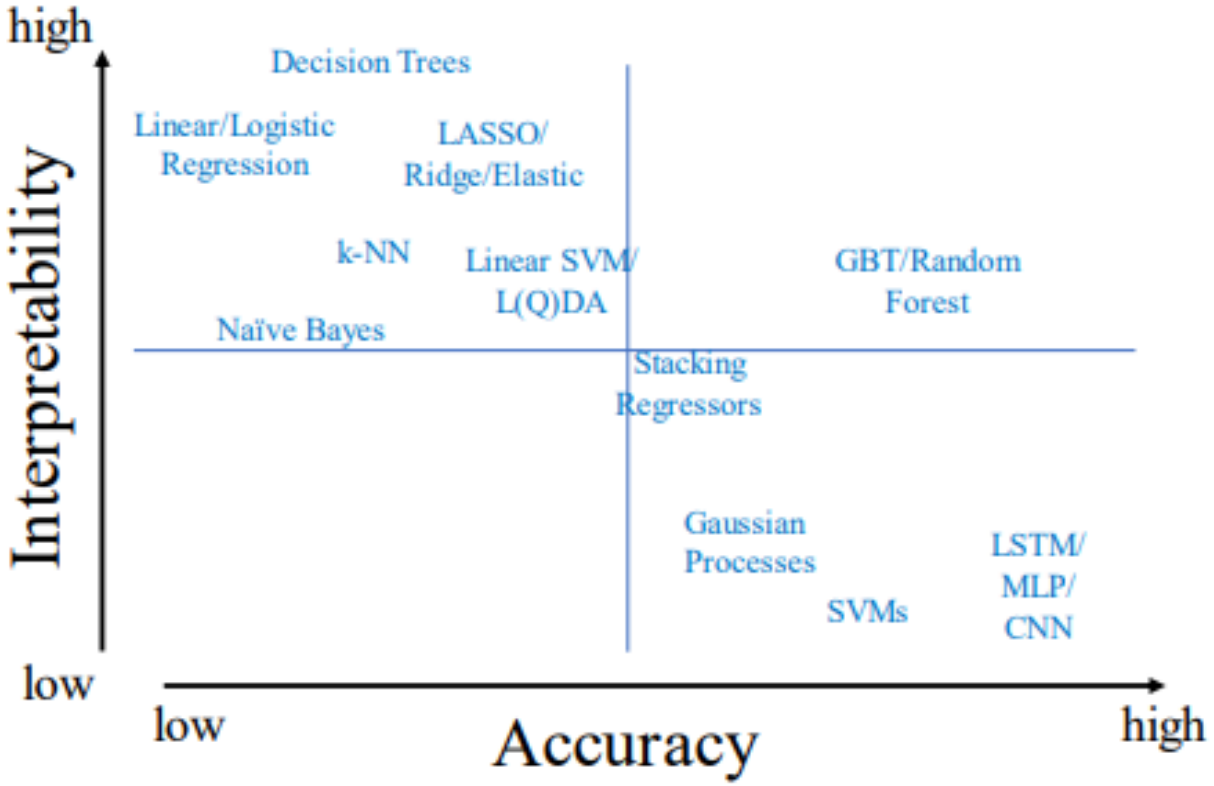


Figure 6: Algorithm selection trade-offs: model-specific Interpretability vs Accuracy.

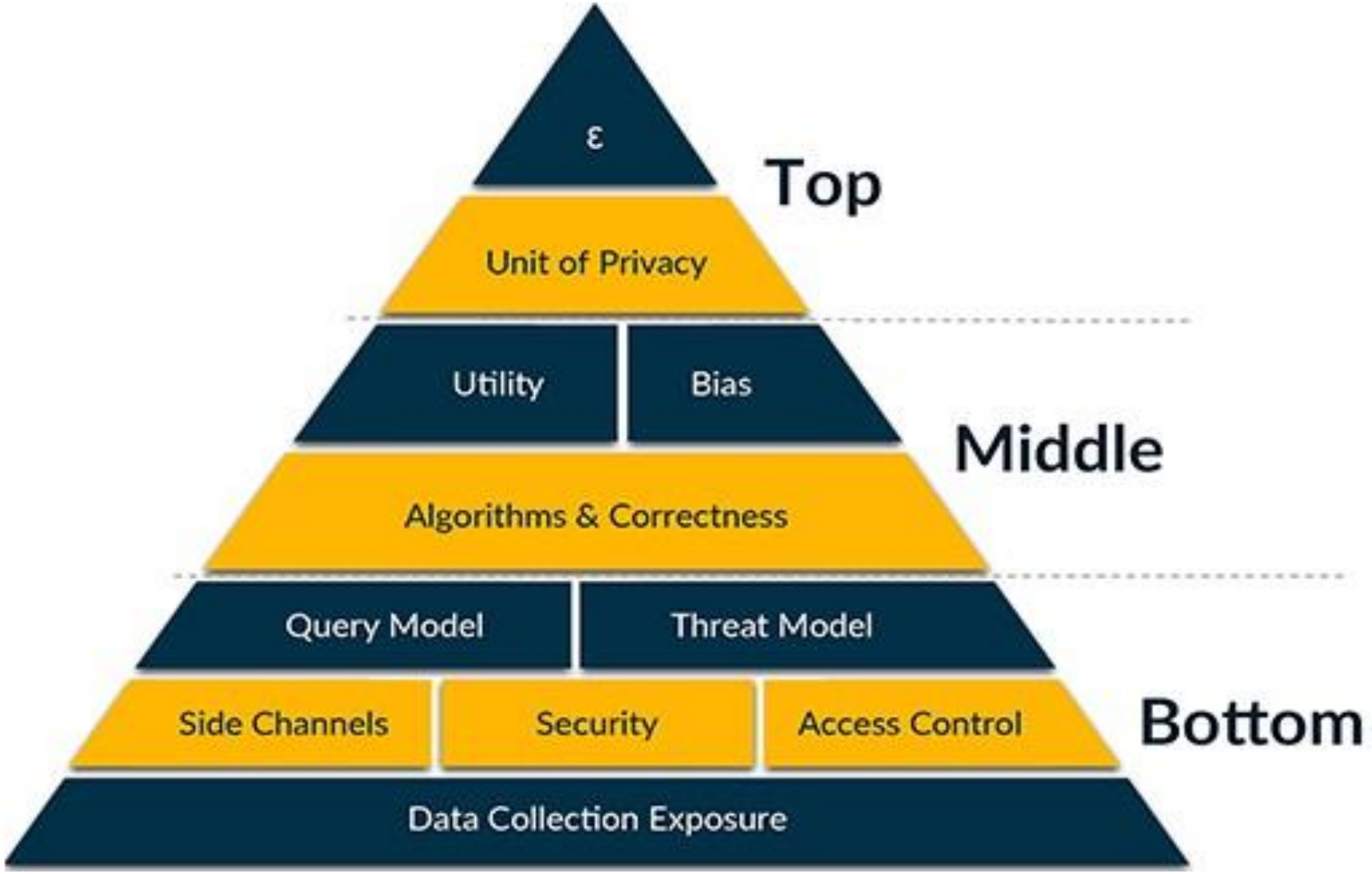
Towards Algorithm Auditing -A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms*
Adriano Koshiyama et.al.

AI Audit

NIST – AI Risk Management Framework https://www.nist.gov/itl/ai-risk-management-framework

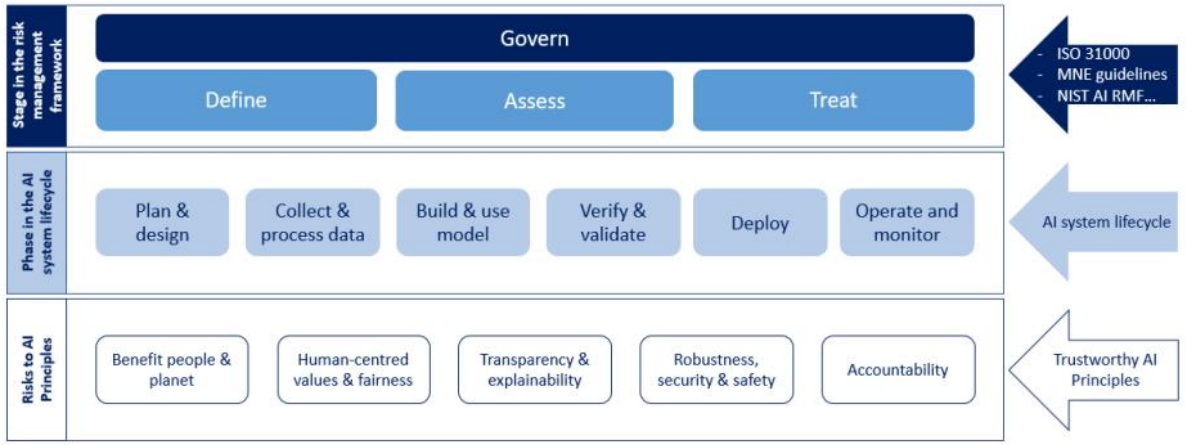
AI Audit

Evaluating a Privacy Protection Technique for the AI Era

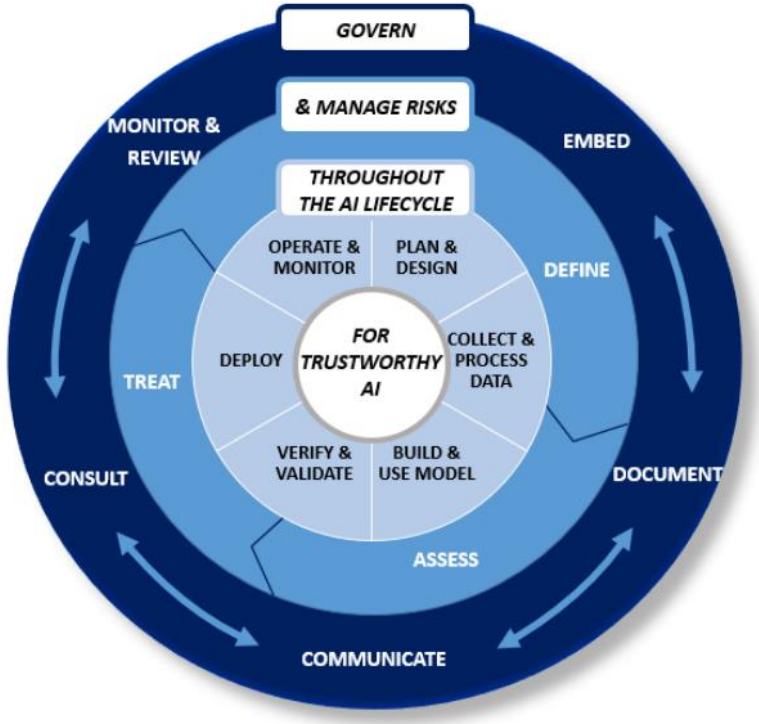


NIST – AI Risk Management Framework https://www.nist.gov/news-events/news/2023/12/nist-offers-draft-guidance-evaluating-privacy-protection-technique-ai-era

AI Audit



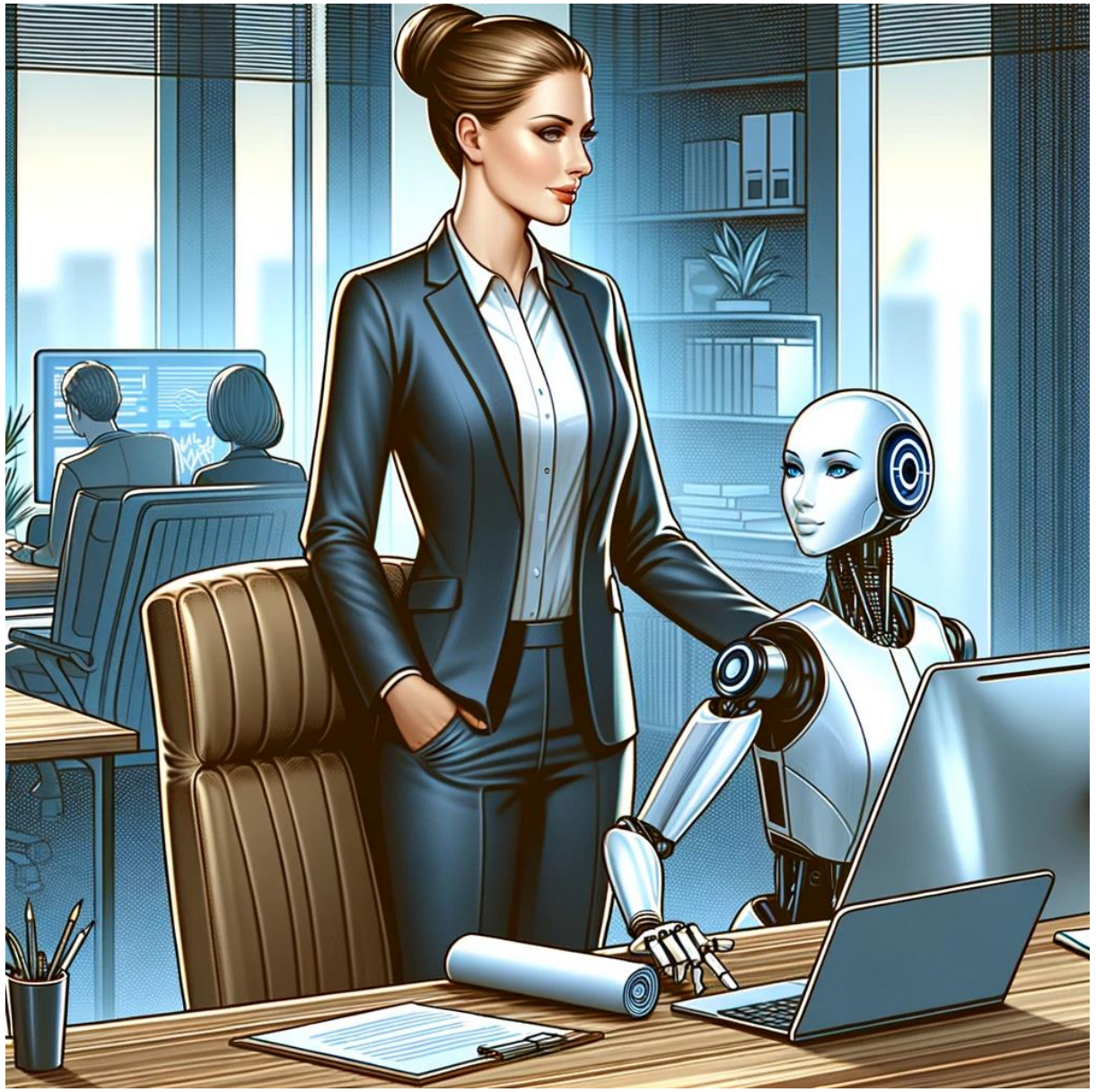
b) Functional view



OECD – Advancing Accountability in AI

Die Welt von morgen

Vielen Dank für Ihre Aufmerksamkeit!



MMag. Dr. Stefan Eder

Stefan Eder ist Rechtsanwalt und Partner der Wirtschaftskanzlei Benn-Ibler Rechtsanwälte GmbH (https://www.benn-ibler.com). Ein Schwerpunkt seiner juristischen Tätigkeit liegt im Bereich IT-Recht, Cybersicherheit, Datenschutz und Compliance, in dem er seit 30 Jahren Unternehmer, Unternehmen aber auch die öffentliche Hand berät. In seiner Tätigkeit beschäftigt er sich auch mit Legal Engineering im Rahmen der Auswertung größerer Datenbestände zum Zwecke der Informationsgewinnung für rechtliche Zwecke. Dabei greift er auf das Fachwissen und technische Verständnis zurück, das er sich in seinem Studium, der Betriebsinformatik, angeeignet hat. So gelingt es ihm, die Brücke zwischen den beiden Welten zu schlagen.

Stefan Eder ist Herausgeber des Usancen Rechtsblogs und eines Blogs für Cybersicherheit mit rund 8.000 Lesern. Er trägt regelmäßig zu Themen der Digitalisierung vor und ist im Rahmen einer Lehrveranstaltung zum Thema Rechtsinformatik an der Universität Wien als Lektor tätig. Weiters ist er seit mehr als 20 Jahren Lektor an der TU Wien im Masterlehrgang Immobilenmanagement und betreut in dieser Funktion auch Masterarbeiten.

Stefan Eder ist auch Gesellschafter der Cybly GmbH (https://cybly.tech), die die LawThek betreibt. Die LawThek ist eine frei zugängliche, mehrsprachige, internationale Rechtsdatenbank (https://lawthek.eu) in Form eines Knowledge Graphen. Ein weiterer Schwerpunkt in der Cybly GmbHsind internationale Forschungsprojekte und Anwendungen zu Aspekten des Legal Dataminings und der Einsatz von KI Tools in diesem Zusammenhang. Stefan Eder beschäftigt sich dabei auch besonders mit Aspekten der Automatisierung juristischer Workflows.

Stefan Eder ist Co-Chair der IRIS Rechtsinformatikkonferenz (https://iris-conferences.eu/) und Initiator und Chair der Reserach Meets Practise Plattform (https://remep.net), die seit 2018 jährlich führenden Forschern aus dem Bereich der (Rechts-)Informatik eine Plattform zur Präsentation und zum Diskurs mit Wissenschaftern und Praktikern bietet.

Kontakt

Kontakt

Dr. Stefan Eder
Benn-Ibler Rechtsanwälte GmbH
bzw.
Cybly GmbH

Tuchlauben 8, 1. Stock
1010 Wien

☎ +43 1 531 550-0
✉ stefan.eder@benn-ibler.com
✉ stefan.eder@cybly.tech