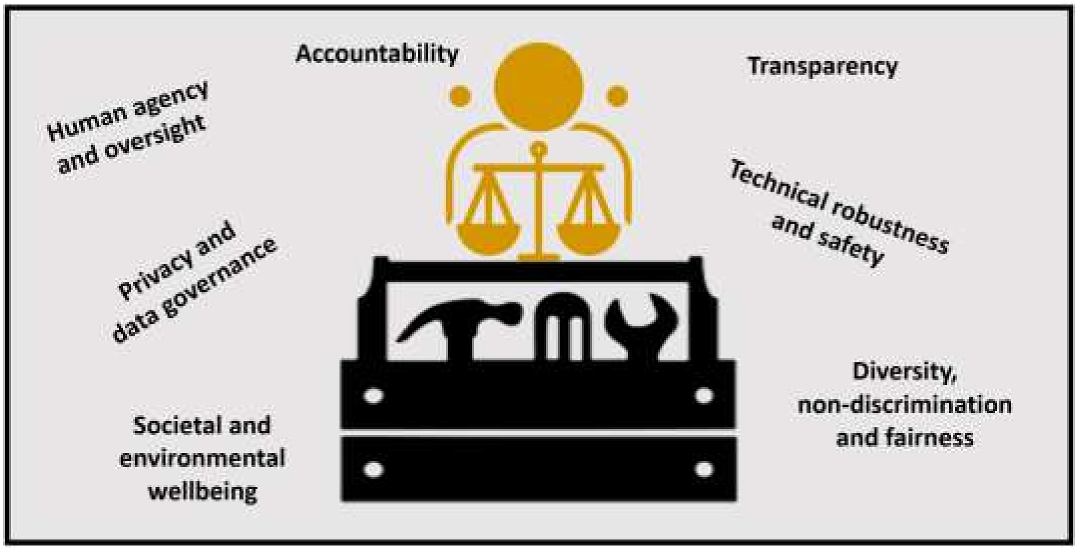
Created by Murat Durmus (CEO AISOMA)
LinkedIn: https://www.linkedin.com/in/ceosaisoma/

A brief overview of some

Ethical-AI Toolkits


Accountability
Transparency
Human agency and oversight
Technical robustness and safety
Privacy and data governance
Diversity, non-discrimination and fairness
Societal and environmental wellbeing

audit-AI (Bias Testing for Generalized Machine Learning Applications)

audit-AI is a tool to measure and mitigate the effects of discriminatory patterns in training data and the predictions made by machine learning algorithms trained for the purposes of socially sensitive decision processes.

The overall goal of this research is to come up with a reasonable way to think about how to make machine learning algorithms more fair. While identifying potential bias in training datasets and by consequence the machine learning algorithms trained on them is not sufficient to solve the problem of discrimination, in a world where more and more decisions are being automated by Artificial Intelligence, our ability to understand and identify the degree to which an algorithm is fair or biased is a step in the right direction.

Link: https://github.com/pymetrics/audit-ai

Playing with AI Fairness (Google's new machine learning diagnostic tool lets users try on five different types of fairness)

Researchers and designers at Google's PAIR (People and AI Research) initiative created the What-If visualization tool as a pragmatic resource for developers of machine learning systems. Using the What-If tool reveals, however, one of the hardest, most complex, and most utterly human, questions raised by artificial intelligence systems: What do users want to count as fair?

link: https://pair-code.github.io/what-if-tool/ai-fairness.html

Ethics & Algorithms Toolkit (A risk management framework for governments)

Government leaders and staff who leverage algorithms are facing increasing pressure from the public, the media, and academic institutions to be more transparent and accountable about their use. Every day, stories come out describing the unintended or undesirable consequences of algorithms. Governments have not had the tools they need to understand and manage this new class of risk.

link: https://ethicstoolkit.ai/

AI Explainability 360 (IBM)

This extensible open source toolkit can help you comprehend how machine learning models predict labels by various means throughout the AI application lifecycle. We invite you to use it and improve it.

link: http://aix360.mybluemix.net/

PwC's Responsible AI

PwC's Responsible AI Toolkit is a suite of customizable frameworks, tools and processes designed to help you harness the power of AI in an ethical and responsible manner - from strategy through execution.

link: https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html

InterpretML (Microsoft)

InterpretML is an open-source package that incorporates state-of-the-art machine learning interpretability techniques under one roof. With this package, you can train interpretable glassbox models and explain blackbox systems. InterpretML helps you understand your model's global behavior, or understand the reasons behind individual predictions.

Interpretability is essential for:

- Model debugging - Why did my model make this mistake?
- Detecting fairness issues - Does my model discriminate?
- Human-AI cooperation - How can I understand and trust the model's decisions?
- Regulatory compliance - Does my model satisfy legal requirements?
- High-risk applications - Healthcare, finance, judicial

link: https://github.com/interpretml/interpret

Diversity Toolkit: A Guide to Discussing Identity, Power and Privilege (USC)

This toolkit is meant for anyone who feels there is a lack of productive discourse around issues of diversity and the role of identity in social relationships, both on a micro (individual) and macro (communal) level. Perhaps you are a teacher, youth group facilitator, student affairs personnel or manage a team that works with an underserved population. Training of this kind can provide historical context about the politics of identity and the dynamics of power and privilege or help build greater self-awareness.

link: https://msw.usc.edu/mswusc-blog/diversity-workshop-guide-to-discussing-identity-power-and-privilege/#intro

WEF Empowering AI Leadership - An Oversight Toolkit for Boards of Directors (World Economic Forum)

This resource for boards of directors consists of: an introduction; 13 modules intended to align with traditional board committees, working groups and oversight concerns; and a glossary of artificial intelligence (AI) terms.

Eight modules focus on strategy oversight and the responsibilities connected with them. They cover: brand, competition, customers, operating model, people and culture, technology, cybersecurity and sustainable development. Five other modules cover additional board oversight topics: ethics, governance, risk, audit and board responsibilities.

Follow the links in the descriptions below to select a module. Continue scrolling down to read the introduction, which includes a description of AI.

Each module provides: a description of the topic, the board responsibilities specific to that module topic, the oversight tools, suggestions for setting an agenda for board discussions and resources for learning more about the topic.

link: https://adobe.ly/2WTBOmI

Responsible Innovation: A Best Practices Toolkit (Microsoft)

This toolkit provides developers with a set of practices in development, for anticipating and addressing the potential negative impacts of technology on people. We are sharing this as an early stage practice for feedback and learning.

Link: https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/

Bias and Fairness Audit Toolkit (Aequitas - Center for Data Science and Public Policy - University of Chicago)

The Bias Report is powered by Aequitas, an open-source bias audit toolkit for machine learning developers, analysts, and policymakers to audit machine learning models for discrimination and bias, and make informed and equitable decisions around developing and deploying predictive risk-assessment tools.

Link: http://aequitas.dssg.io/

An Ethical Toolkit for Engineering/Design Practice (Markkula Center for Applied Ethics - Santa Clara University)

The tools below represent concrete ways of implementing ethical reflection, deliberation, and judgment into tech industry engineering and design workflows.

Used correctly, they will help to develop ethical engineering/design practices that are:

- Well integrated into the professional tech setting, and seen as a natural part of the job of good engineering and design (not external to it or superfluous)
- Made explicit so that ethical practice is not an ‘unspoken’ norm that can be overlooked or forgotten
- Regularized so that with repetition and habit, engineers/designers/technologists can gradually strengthen their skills of ethical analysis and judgment
- Operationalized so that engineers/designers are given clear guidance on what ethical practice looks like in their work setting, rather than being forced to fall back on their own personal and divergent interpretations of ethics

Link: https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/

Fairlearn (Microsoft)

Fairlearn is a Python package that empowers developers of artificial intelligence (AI) systems to assess their system's fairness and mitigate any observed unfairness issues. Fairlearn contains mitigation algorithms as well as a Jupyter widget for model assessment. Besides the source code, this repository also contains Jupyter notebooks with examples of Fairlearn usage.

Link: https://github.com/fairlearn/fairlearn

TOOLBOX: Dynamics of AI Principles (AI ETHICS LAB)

You can use it to think through the ethical implications of the technologies

that you are evaluating or creating.

The Box is a simplified tool that

- lists important ethical principles and concerns,
- puts instrumental ethical principles in relation to core principles,
- helps visualize ethical strengths & weaknesses of technologies, and
- enables visual comparison of technologies.

Link: https://aiethicslab.com/big-picture/

Algorithmic Accountability Policy Toolkit (AI Now Institute)

The following toolkit is intended to provide legal and policy advocates with a basic understanding of government use of algorithms including, a breakdown of key concepts and questions that may come up when engaging with this issue, an overview of existing research, and summaries of algorithmic systems currently used in government. This toolkit also includes resources for advocates interested in or currently engaged in work to uncover where algorithms are being used and to create transparency and accountability mechanisms

Link: https://ainowinstitute.org/aap-toolkit.pdf

From Principles to Practice – An interdisciplinary framework to operationalize AI ethics (AIEI Group)

The AI Ethics Impact Group is an interdisciplinary consortium led by VDE Association for Electrical, Electronic & Information Technologies and Bertelsmann Stiftung. We have come together in 2019 to bring AI ethics from principles to practice.

With our labelling and specification frameworks we aim to support the enforcement of European values and the protection of citizens in Europe, to create quality transparency and comparability in the market, and to prevent unnecessary red tape for companies with a straightforward implementation only where necessary.

We also want to make sure that AI ethics becomes easy to communicate and understand for organizations and citizens all over Europe and beyond.

We present our findings here and are looking forward to a broad implementation-oriented debate.

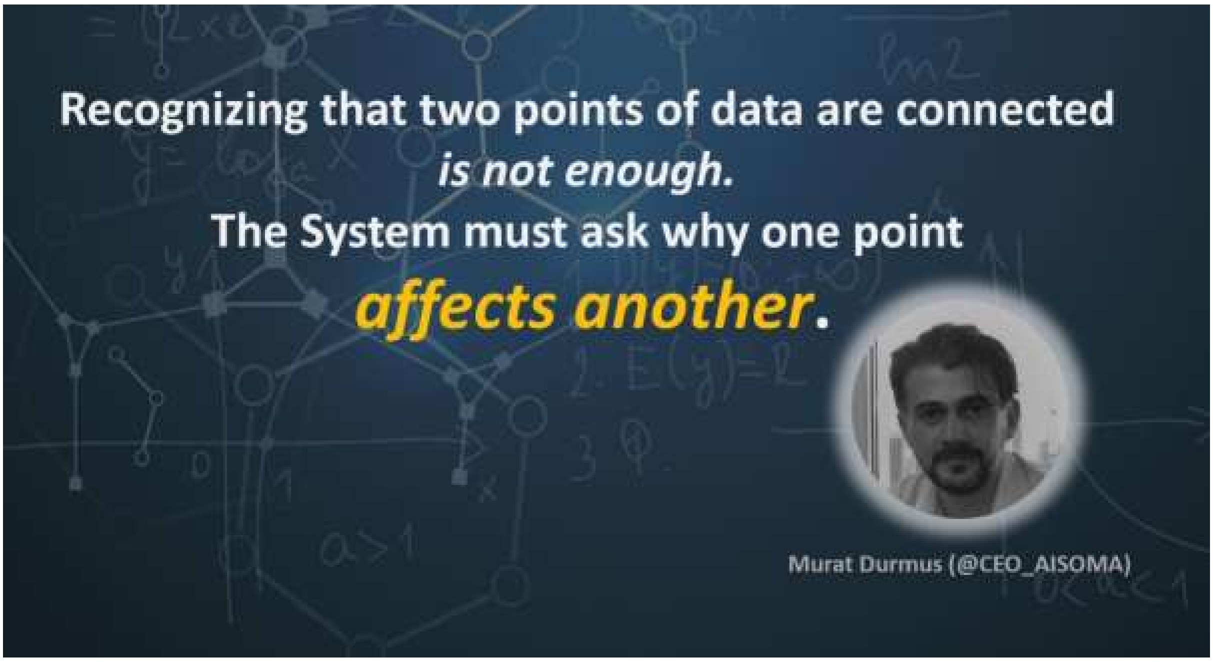
Link: https://www.ai-ethics-impact.org/en

Ethical OS Toolkit

What's in the Toolkit:

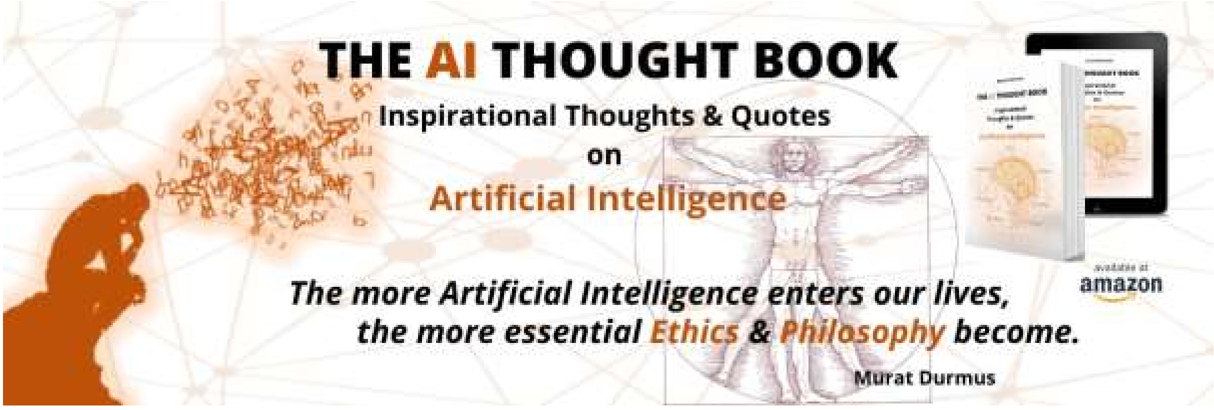
- A checklist of 8 risk zones to help you identify the emerging areas of risk and social harm most critical for your team to start considering now.
- 14 scenarios to spark conversation and stretch your imagination about the long-term impacts of tech you're building today.
- 7 future-proofing strategies to help you take ethical action today.

Link: https://ethicalos.org/


Recognizing that two points of data are connected
is not enough.
The System must ask why one point
affects another.
Murat Durmus (@CEO_AISOMA)

Created by Murat Durmus (CEO AISOMA)
LinkedIn: https://www.linkedin.com/in/ceosaisoma/

This might be also of interest (my new Book on Mindful AI):


THE AI THOUGHT BOOK
Inspirational Thoughts & Quotes
on
Artificial Intelligence
The more Artificial Intelligence enters our lives,
the more essential Ethics & Philosophy become.
Murat Durmus
available at
amazon

(Amazon) THE AI THOUGHT BOOK: Inspirational Thoughts & Quotes on Artificial Intelligence (including 13 colored illustrations & 3 essays for the fundamental understanding of AI)

Contact

Murat Durmus
CEO & Founder @ AISOMA AG
Frankfurt am Main, Hessen, Deutschland

https://www.linkedin.com/in/ceosaisoma/

murat.durmus@aisoma.de

https://www.aisoma.de