

Applied Machine Learning Days 2022

Tuesday March 29, 2022 / 10:45 AM - 11:00 AM

"

AI Ethics Standards and Guidelines



Prof. Dr. Dagmar Monett



Prof. Dr. Dagmar Monett





@dmonett

Hello!



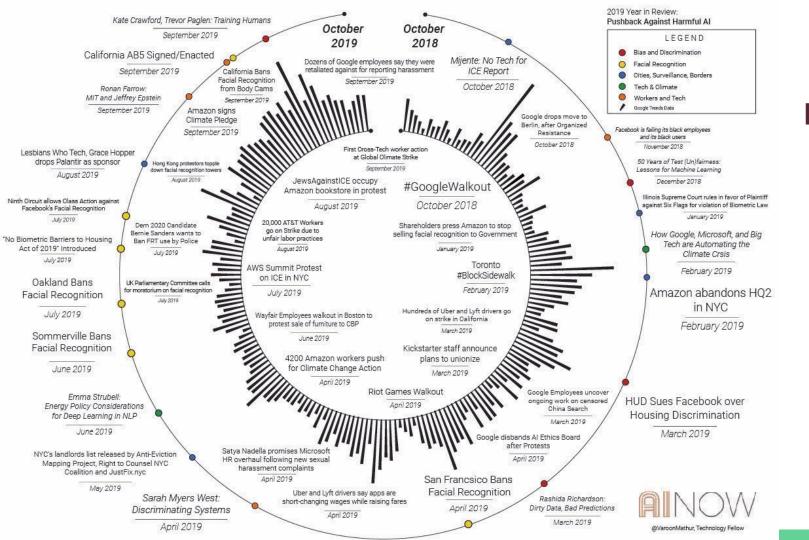
- Prof. Dr. Computer Science Artificial Intelligence, Software Engineering
- ► Head of the Computer Science Dept. (Apr. 2022–) Faculty of Cooperative Studies Berlin School of Economics and Law (HWR Berlin)
- Co-Director M.Sc. Digital Transformation (since Dec. 2020)
 Berlin Professional School
 Berlin School of Economics and Law (HWR Berlin)



http://monettdiaz.com



dagmar.monett-diaz@hwr-berlin.de



Pushback Against Harmful Al

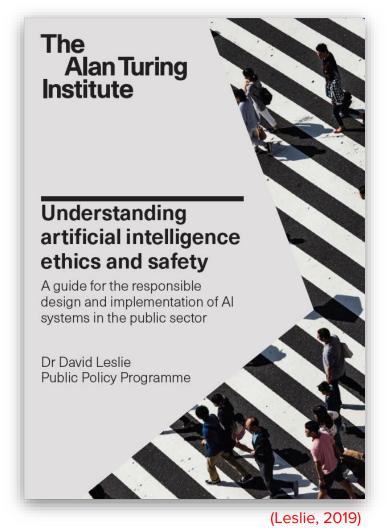
Courtesy of @VaroonMathur, Tech fellow @AlNowInstitute

"Ethical judgment corresponds with an ethical system."

If the system is not ethical, we cannot have ethical outcomes.

It's actually much more than only ethics...

It is also standards, policies, principles, law, values, guidelines, recommendations.



"Al ethics is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of Al technologies."

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of Al systems in the public sector. *The Alan Turing Institute*.

https://doi.org/10.5281/zenodo.3240529

Ethical Platform for the Responsible Delivery of an Al Project

SUM Values

that support, underwrite, and motivate a responsible innovation ecosystem



Respect, Connect, Care, Protect

Objectives: to provide an accessible framework for consideration of the moral scope of the social and ethical impacts of your project and to establish well-defined criteria to evaluate its ethical permissibility.

FAST Track Principles

that facilitate an actionable orientation to the ethical design and use of AI systems



Fairness, Accountability, Sustainability, Transparency

Objectives: to make sure that your project is biasmitigating, non-discriminatory, and fair, and to safeguard public trust in your project's capacity to deliver safe and reliable Al innovation.

PBG Framework

that operationalises the values and principles in an end-to-end workflow governance model

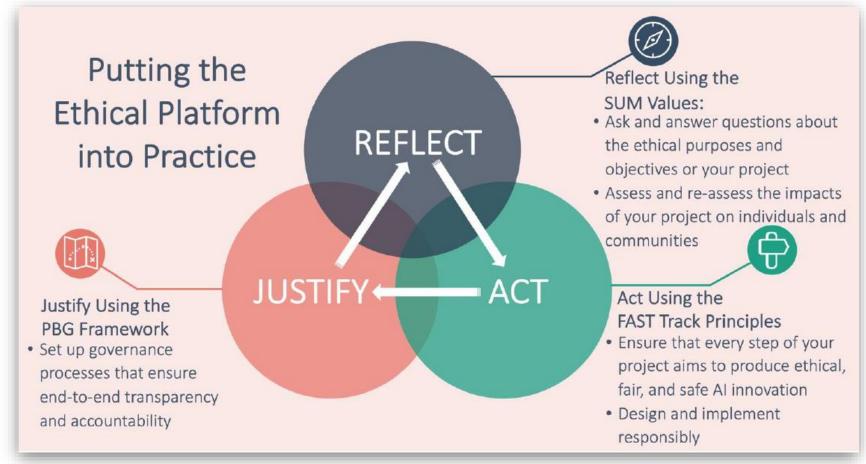


Process-Based
Governance Framework

Objective: to set up transparent processes of design and implementation that safeguard the justifiability of both your AI project and its product as well as enable end-to-end accountability.

"A framework of ethical values that **S**upport, **U**nderwrite, and **M**otivate a responsible data design and use ecosystem."

(Leslie, 2022)





of produce, printer and gene in the seasons.

Second of the second of the seasons of the seasons

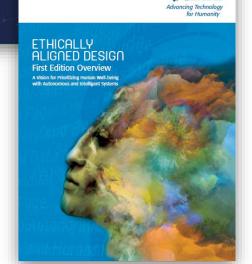
we do take species of the contemporary of the

INDEPENDENT HIGH-LEVEL EXPERT GROUP ON **ARTIFICIAL INTELLIGENCE** SET UP BY THE EUROPEAN COMMISSION **POLICY AND INVESTMENT RECOMMENDATIONS**





ETHICALLY ALIGNED DESIGN First Edition A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems



IEEE



UNIT

INDEPENDENT

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

ETHICS GUIDELINES FOR TRUSTWORTHY AI



THE UNIVERSAL DECLARATION of Human Rights

extend recognition of the informat signify and of the equal and melianship eights of all members of the human family is the juminion of freedom, justice and power in the model.

process disregard and commut for homes rights have treated in me acts which here introped the constraint of modified, and the solvest of a world in which haven beings shall super freedom of querk and fullef and freedom from four and west has been proclaimed so the highest represents of the removes people,

process it is exceeded, if most in not to be compelled to have recommesee at first rewart, to rebuillion regarded concern and approaches, that haven sights should be protected by the rain of law.

American is in concentral to pressure the development of Disable settings. comp miles

process the peoples of the Entirel Nations have in the Charter regionsel. the beams person and in the equal rights of new and source and how . Homeotics and councy the peoples of merimons ander their people in

Member Notes here pledged themselve to unknow in cooperation with the Exited Nations, the promotion of anterest request for and observance of houses rights and Josephoreand Fundame.

a common anticomorphic of their stakes and freedom is of the greatest impositions for the full evaluation of this pholon.

THE RESIDENCE INC. LABOUR. CO., LABOUR.

received this Patients Derivative of Banner Rights on a common streeterd of achievement for all peoples and all nations to the end that every individual and every segon of series, keeping this Beckennian constant's in soled, shall stell to breeking and education in parameter request for those rights and freedom and hypeographic measures. noticed and intractional, is some their national and effective their falls in parelement homes sights in the dignity and most of recognition and abstracts, but many the peoples of Manter State

marked board on authorite and distributed

water of boson a self-directly registerables artist or the femous artists deposite at any local code or the order day temposite, respect position or who approxirefrest e sone pligo propora, bell or also retu-E furbacion, se distante stad la male et les boss el fu-palitud production à secondaries sons el fils comme e moltre la cital a programme, alcular de section la sebarrier, but a feeled despring print, a pair of after Debation of Sciences,

and the story trade about the published is not time factor.

The state of the state of the same of the same of the state of the same of the

ATTECH AND ADMINISTRATION OF THE PARTY AND ADMINISTRATION OF T est me distributed in regal proteins of the little SE are in Marie most protein against my distributed in relation of the following and against any included in such distributed.

company tomoral influence for such conting the formation of the second flow in the contrastor or to him.

COLUMN TO THE PARTY OF THE PART public fracting by an independent and impacted relation, in the description of the rights and self-public and of any criment

The contract of the second section in the the second secon

I No are stell to half galls of any good officer on mount of the set in advance which did not make in a good officer. And the stellar of the stellar is good officer. period the staff of feature parents for impact from the case that was applicable to the feature of distance and contribute

PATE TO THE RESIDENCE OF STREET OF STREET, AND STREET,

and it is not become to the safe or bracket of mornior It foregoes too the light to beautiful country training the local

3. Not right you are to received to the read of proposition yes

more in the human his tis spirit is a second If the constability interests from the constability of the con-

If the property of the party of desire being not a to decimal No high shall be contract because with the housest bull-conserof the incoding species.

and it profited to profession to court, and the basis. committee of the part of the part is not present them. so and in a comment with other. I be an old to consideration of the papers.

Wigner of Select and Parallel and the Color of the Color

And the same of th Storage and results and requestions of Storage

or the state of the last of th of second description and resolution If the sea was in complete to being to an assessment

personal for more than a bangli test man E. Remoted from the right of easier locate to public spicing to fin

program the left had be represed to served; and places are seen when that he has provided and seem sublings and shall be Marrie and the trib particular for company to the

and many or a wild to relate the large recording to the state of the s affected rights indigenouslike for this Highly and the Free Heatings

I finance offers an distribution for the sale

Education of any or operation of the same

area fourth for financial and the best of manager and to all

A Souper has being the best and a get being chara-No the promoted of the interest.

which it is because the the right temperaphilipped behaling

gapte for the results out and large, differently and all his feature, minding freel contings, by only of resulted past and separate which recome seen the right in another to the result of sampling tools asking disability with which all legs as notice but all hadred to distribute broad to write

more, but designer, inhalter have an an last of markets death with

COLD, with Resident for the right to place the Manusco that is the at the last to consider and believed any and strong diff to been proved events and up-I district and to discoul a factor description of technical attention and in the completing of manual for before option and features fraction, it shall produce patentioning also one yet brottle sering of paties, such a copper grow and and looks the private of the privat Spotter by the state

where t = 0. Assume that the sign bady is particular, to the other the ϕ fits constantly, in some the second in which is observed whether the properties and its bandle. If you have been presented to be provided and the party of the party o If the same has the man or the distriction of the stand in

THE RESIDENCE OF STREET mile it still by upts and frames on high to the fact

their trades would envirance of the personal to people. I with morning of the rights and Tonation, company also because ATTACK TO A RESIDENCE OF AN ADMINISTRATION OF THE PARTY O

A New India and Sendant was to be replaced in the Sendant Sendant.

marries, for our filter, group or present any right to original to your merring or to contract and not contract of the of the right and fundament, on facility beauty.

rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status." (Article 2)

"Everyone is entitled to all the

"No one shall be subjected to arbitrary arrest, detention or exile." (Article 9)

https://www.un.org/en/universal-declaration-human-rights/



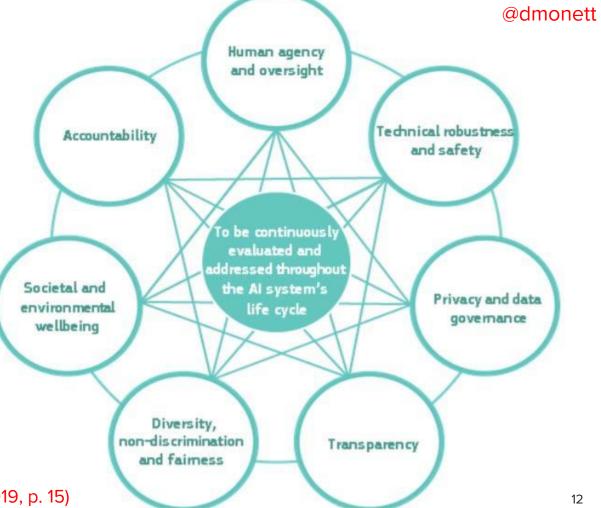
"Ensure that the development, deployment and use of Al systems meets the seven key requirements for Trustworthy Al: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

societal well-being and (7)

accountability."

requirements for responsible Al



"Trustworthy AI has three components, which should be met throughout the system's entire life cycle:

- 1. it should be lawful, complying with all applicable laws and regulations;
- **2.** it should be **ethical**, ensuring adherence to ethical principles and values; and
- **3.** it should be **robust**, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm."

The realization of Trustworthy Al

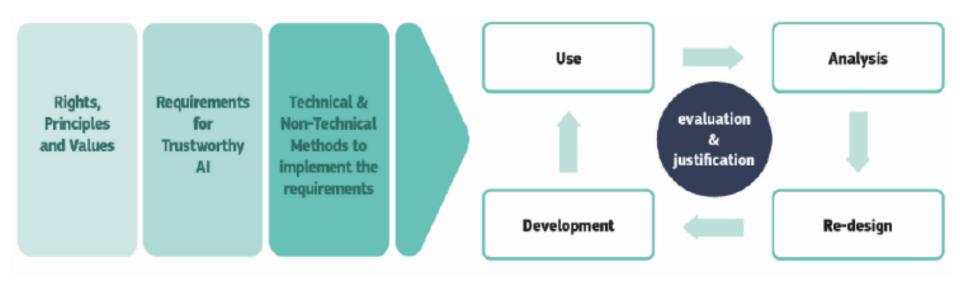


Figure 3: Realising Trustworthy AI throughout the system's entire life cycle

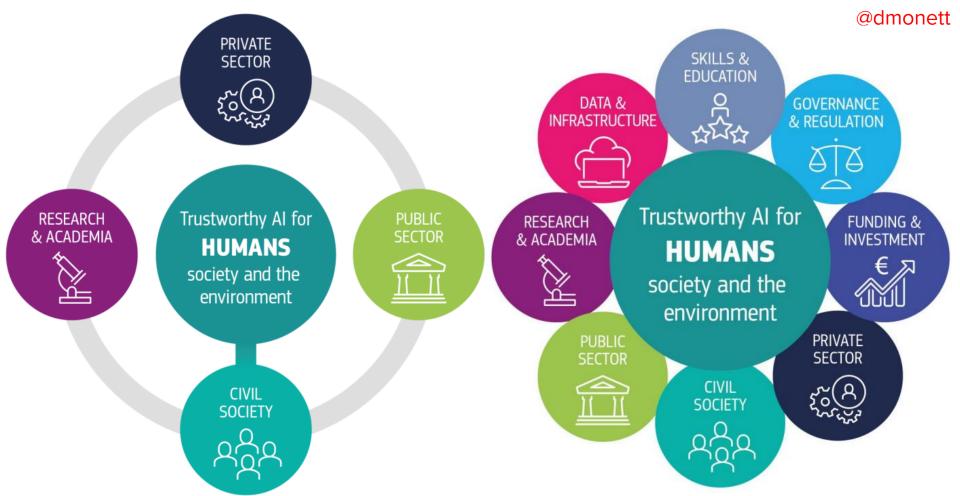
"An [Al system] cannot genuinely be deemed 'ethical' without accounting for the business it is involved in."

"There can be no ethical Al without ethical businesses to build it."



"[R]ecommendations that can guide Trustworthy AI towards sustainability, growth and competitiveness, as well as inclusion – while empowering, benefiting and protecting human beings."

https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence





"[A] non-exhaustive
Trustworthy AI assessment
list [with 147 questions] to
operationalise Trustworthy AI.
It particularly applies to AI
systems that directly interact
with users, and is primarily

https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

addressed to developers and

deployers of Al systems."





What the company says about its Al product



How robust the product is

"Did you assess how your system behaves in unexpected situations and environments?"

(Trustworthy AI assessment list, 2019)





What the end user trully needs



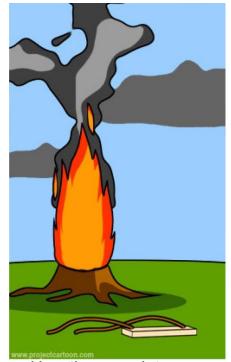
How human-centered the product is

"In case the AI system features a chat bot or conversational system, are the human end users made aware of the fact that they are interacting with a non-human agent?"

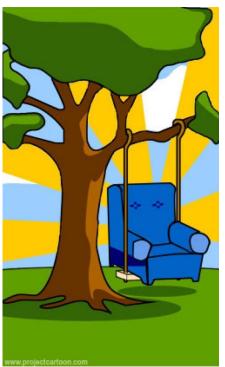
(Trustworthy Al assessment list, 2019)

Cartoon created at www.projectcartoon.com





How the user data was processed

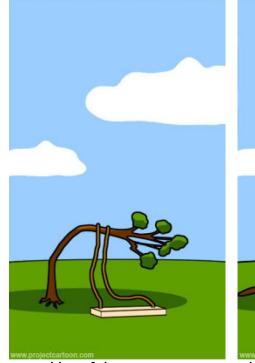


The product and the bias

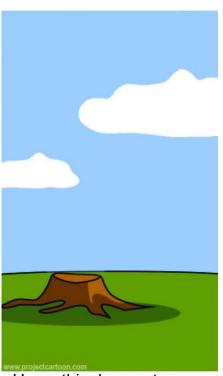
"Did you consider ways to develop the AI system or train the model without or with minimal use of potentially sensitive or personal data?"

(Trustworthy Al assessment list, 2019)





How fairness was considered



How ethical aspects were considered

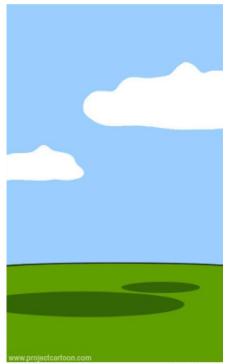
"Did you consider diversity and representativeness of users in the data? Did you test for specific populations or problematic use cases?"

(Trustworthy AI assessment list, 2019)









What the company thinks about regulation

"Did you put in place mechanisms that facilitate the system's auditability by internal and/or independent actors?"

(Trustworthy AI assessment list, 2019)

What makes us human makes us also Al (ir)Responsible





"[IEEE's] commitment to engage global communities in creating practical guidelines that will help ensure the development of ethically aligned autonomous and intelligent products, services, and systems."

https://ethicsinaction.ieee.org/#series

Human Rights EAD General Principles Well-being **Data Agency** Effectiveness Transparency Accountability **Awareness of Misuse** Competence

- **1. Human Rights**—A/IS shall be created and operated to respect, promote, and protect internationally recognized human rights.
- **2. Well-being**—A/IS creators shall adopt increased human well-being as a primary success criterion for development.
- **3. Data Agency**—A/IS creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity.
- **4. Effectiveness**—A/IS creators and operators shall provide evidence of the effectiveness and fitness for purpose of A/IS.
- **5. Transparency**—The basis of a particular A/IS decision should always be discoverable.
- **6. Accountability**—A/IS shall be created and operated to provide an unambiguous rationale for all decisions made.
- **7. Awareness of Misuse**—A/IS creators shall guard against all potential misuses and risks of A/IS in operation.
- **8. Competence**—A/IS creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

(IEEE, 2019, p. 11)

Be responsible.

Be responsible.

Perspective Published: 02 September 2019

The global landscape of AI ethics guidelines

<u>Anna Jobin</u>, <u>Marcello Ienca</u> & <u>Effy Vayena</u> □

Nature Machine Intelligence 1, 389–399 (2019) Cite this article

39k Accesses 522 Citations 740 Altmetric Metrics

"84 documents containing ethical principles or guidelines for Al."

"No single ethical principle appeared to be common to the entire corpus of documents, although there is an emerging convergence around the following principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy."

The Ethics of AI Business Practices: A Review of 47 AI Ethics Guidelines

Al and Ethics (Forthcoming)

55 Pages • Posted:

Blair Attard-Frost

University of Toronto, Faculty of Information

Andrés De los Ríos

Doblin, a Deloitte Business

Deneille R. Walters

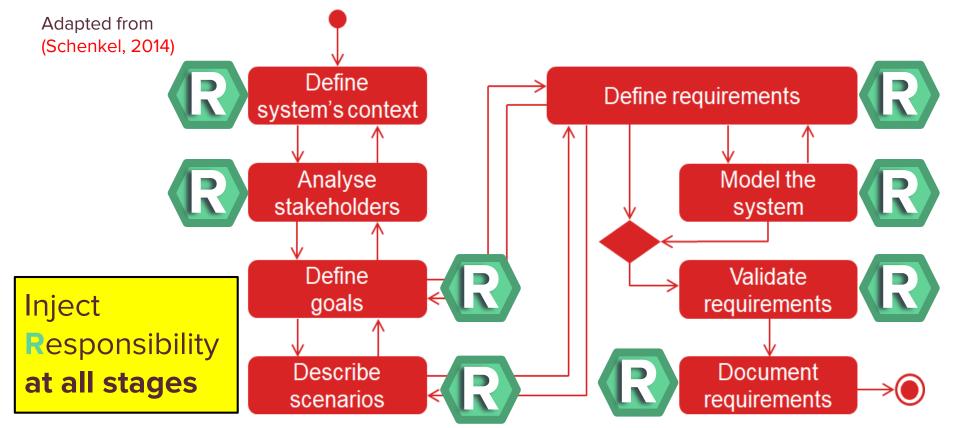
Independent Education Design & Technology Consultant

"The fairness and accountability of AI business practices are the two most widely observed principles."

Develop responsible Al.



Requirements engineering as a process

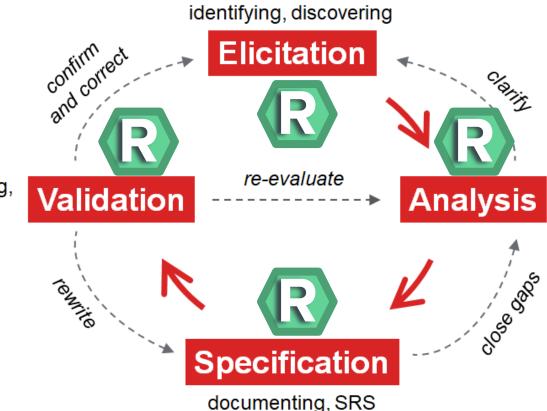


Requirements development in detail

Adapted from (Wiegers & Beatty, 2013)

evaluating, verifying

Inject
Responsibility
at all stages



classifying, representing, deriving, negotiating

High quality requirements

Adapted from (Rupp et al., 2014)

2: Characterise the activity of the system **5**: Phrase 1: Determine the process, conditions identify the functionality verb shall cess> provide <whom> object-When?/Under additional details the shouldwhat conditions? system with the ability to about the object be able to will cess>

Inject

Responsibility into steps 4 & 5. Make it a **MUST**



Determine legal obligation



4: Fine tune the requirement



Put it on your backlog!

That's how (R) gets done.

Responsible Al needs responsible humans.

Be responsible.



Applied Machine Learning Days 2022

Tuesday March 29, 2022 / 10:45 AM - 11:00 AM

66

AI Ethics Standards and Guidelines



Prof. Dr. Dagmar Monett