

USING TECH/AI ETHICS: *WHY* ARE WE DOING THIS? (WHY DO TECHNOLOGY COMPANIES WANT ETHICAL INPUT?)

1. Provide (human language) metrics to determine AI tool's success
+ Enhance technical design

Determine what *counts* as success

2. AI Ethics Evaluations (Audits): How do you do them?
→ Risk mitigation and legal compliance

USING TECH/AI ETHICS: *Why* ARE WE DOING THIS?



Provide metrics to determine success

- What does privacy mean?
- How do you design for it?
- How far will you go against other values (Social wellbeing)?

USING TECH/AI ETHICS: *WHY* ARE WE DOING THIS?

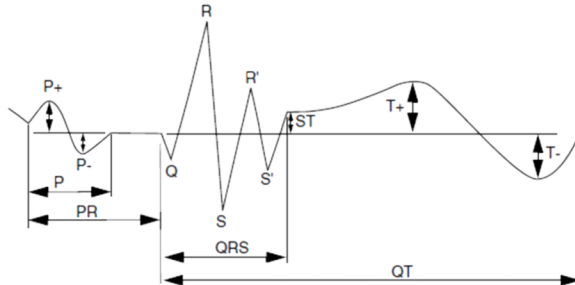
Provide metrics to determine success

What *counts* as accuracy? (How balance Sensitivity / Specificity?)

How does balance change:

Autonomy/Dignity/Privacy/Fairness/Equity/Social Wellbeing...?

Human values versus profit versus medical goals...



USING TECH/AI ETHICS: *WHY* ARE WE DOING THIS?

(WHY DO TECHNOLOGY COMPANIES WANT ETHICAL INPUT?)

1. Provide (human language) metrics to determine AI tool's success
+ Enhance technical design
Determine what *counts* as success
2. AI Ethics Evaluations (Audits): How do you do them?
→ Risk mitigation and legal compliance

USING TECH/AI ETHICS: AI Ethics Evaluations (Audits)

Because innovation is outstripping legal regulation, incorporating of ethics provides justification for decisions about the design and use of AI models, while also clarifying accountability.

Evaluations (Audits) increasingly required by law

Risk mitigation and legal compliance

USING ETHICS: EVALUATIONS (AUDITS)

What are they?

- Cast technology in ethical/humanist terms

 - As opposed to scientific, economic, legal (iPhone example)

- Evaluate (ensure) how well the technology serves fundamental human rights

USING ETHICS: EVALUATIONS (AUDITS)

How do you do them?

1. Social/Technical Scenario
2. Evaluate
3. Report/Recommend

USING ETHICS: EVALUATIONS (AUDITS): SOCIAL/TECHNICAL SCENARIO

What do you need to know to know what we are talking about?

- What does the AI/technology do?
- Who does it do it for?
- What counts as success?
- ...?

Gather stakeholders

Translation

Meeting/Zoom

Draft and commentary (Google doc)

USING ETHICS: EVALUATIONS (AUDITS): EVALUATE

Working group(s) with Principles

- EC Ethics Guidelines/Other...

Top down (Dignity: status of traditional music repurposed: At what point is the original author of the music no longer the author...? (*Whose music ethically versus legally?*))

Bottom up (Autonomy: new capacities to interact with music for everyone, for hearing impaired...)

Meeting/Zoom

Consolidate in terms of principles

Draft and commentary (Google doc)

USING ETHICS: EVALUATIONS (AUDITS): EVALUATE

EXAMPLE GOOGLE DOC...

USING ETHICS: EVALUATIONS (AUDITS): REPORT/RECOMMEND

Convert Google doc into (academic) narrative

Major areas:

- Autonomy (Cardisio case)

- Privacy (Covid Case)

- Fairness + Explainability (exAID case)

USING ETHICS: EVALUATIONS (AUDITS): REPORT/RECOMMEND

Convert Google doc into academic / public policy narrative

272

IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY, VOL. 3, NO. 4, DECEMBER 2022

Assessing Trustworthy AI in Times of COVID-19: Deep Learning for Predicting a Multiregional Score Conveying the Degree of Lung Compromise in COVID-19 Patients

Himanshi Allahabadi, Julia Amann, Isabelle Balot[✉], Andrea Beretta, Charles Binkley[✉], Jonas Bozenhard[✉],
Frédéric Bruneault, James Brusseau[✉], Sema Candemir, Luca Alessandro Cappellini[✉],
Subrata Chakraborty[✉], *Senior Member, IEEE*, Nicoleta Cherciu, Christina Cociancig, Megan Coffee[✉], Irene Ek,
Leonardo Espinosa-Leal, Davide Farina, Geneviève Fieux-Castagnet, Thomas Frauenfelder[✉], Alessio Gallucci,
Guya Giuliani, Adam Golda[✉], Irmhild van Halem, Elisabeth Hildt[✉], Sune Holm, Georgios Kararigas[✉],
Sebastien A. Krier, Ulrich Kühne, Francesca Lizzi[✉], Vince I. Madai, Aniek F. Markus[✉], Serg Masis[✉],
Emilie Wiinblad Mathez, Francesco Mureddu, Emanuele Neri, Walter Osika, Matiss Ozols[✉], Cecilia Panigutti,
Brendan Parent, Francesca Pratesi[✉], Pedro A. Moreno-Sánchez, Giovanni Sartor, Mattia Savardi[✉],
Alberto Signoroni[✉], Hanna-Maria Sormunen[✉], Andy Spezzatti, Adarsh Srivastava[✉], Annette F. Stephansen,
Lau Bee Theng[✉], *Senior Member, IEEE*, Jesmin Jahan Tithi, Jarno Tuominen[✉], Steven Umbrello[✉],
Filippo Vaccher, Dennis Vetter[✉], Magnus Westerlund, Renee Wurth, and Roberto V. Zicari[✉]

Abstract—This article's main contributions are twofold: 1) to demonstrate how to apply the general European Union's High-Level Expert Group's (EU HLEG) guidelines for trustworthy AI in practice for the domain of healthcare and 2) to investigate the research question of what does "trustworthy AI" mean at the time of the COVID-19 pandemic. To this end, we present the results of a *post-hoc* self-assessment to evaluate the trustworthiness of an AI system for predicting a multiregional score

pandemic. The AI system aims to help radiologists to estimate and communicate the severity of damage in a patient's lung from Chest X-rays. It has been experimentally deployed in the radiology department of the *ASST Spedali Civili clinic* in Brescia, Italy, since December 2020 during pandemic time. The methodology we have applied for our *post-hoc* assessment, called Z-Inspection[®], uses sociotechnical scenarios to identify ethical, technical, and domain-specific issues in the use of the AI system in the context

USING ETHICS: EVALUATIONS (AUDITS)

QUICK EXAMPLE CASE (AI TUTOR, LOW STAKES?)

1. Social/Technical Scenario
2. Evaluate
3. Report/Recommend



USING ETHICS: EVALUATIONS (AUDITS)

QUICK EXAMPLE CASE (AI TUTOR, LOW STAKES)

Optimize For:

Autonomy/Dignity/Privacy/
Fairness/Equity/Social wellbeing
Explainability/Safety/Performance

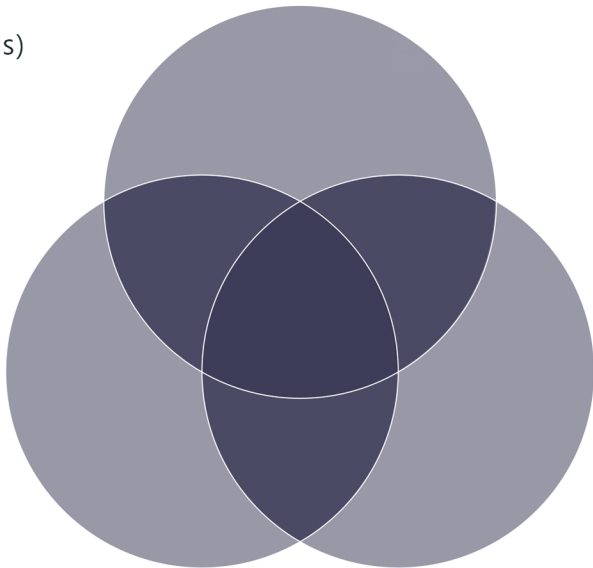
What does this mean?

For whom?

How can you do it?

What if there is no "sweet spot"?

*(Does maximizing user autonomy
subtract from privacy, dignity?)*



OVERVIEW

USING ETHICS: WHY ARE WE DOING THIS?

1. Provide human language/metrics to measure AI success
+ Enhance technical design
2. Risk mitigation and legal compliance

USING AI ETHICS: EVALUATIONS (AUDITS)

What are they?

How do you do them?

1. Social/Technical Scenario
2. Evaluate
3. Report/Recommend

Quick case