Responsible Innovation approaches in Industry

DR. ANSGAR KOENE,
GLOBAL AI ETHICS AND REGULATORY LEADER, EY
SENIOR RESEARCH FELLOW, HORIZON DER, UNIVERSITY OF NOTTINGHAM
CHAIR OF IEEE P7003 STANDARD FOR ALGORITHMIC BIAS CONSIDERATIONS
30TH JANUARY 2020

Drivers for Responsible Innovation

Regulatory Compliance

Reputation

Ideological conviction

Governance frameworks

Market solutions (reputation)

DEMAND SIDE SOLUTIONS

Marmite maker Unilever threatens to pull ads from Facebook and Google
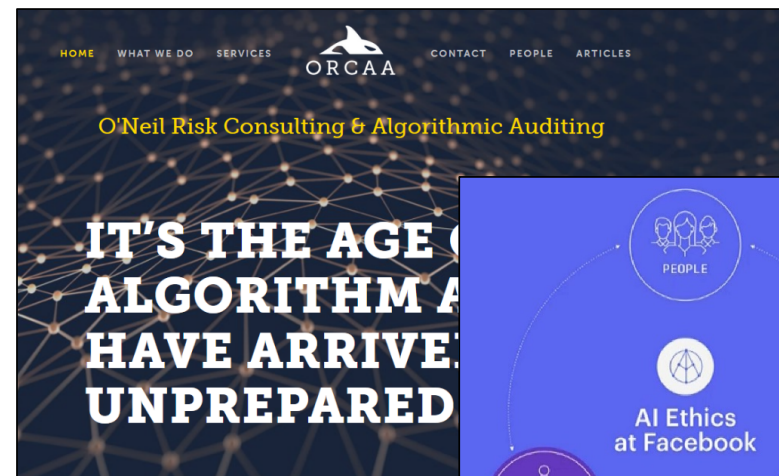
One of world's biggest advertisers says it will avoid platforms that 'create division'

▲ Unilever brands include Marmite and Pot Noodle. Photograph: Chris Radburn/PA

Unilever has threatened to withdraw its advertising from online platforms such as Facebook and Google if they fail to eradicate content which "create division in society and promote anger and hate" .

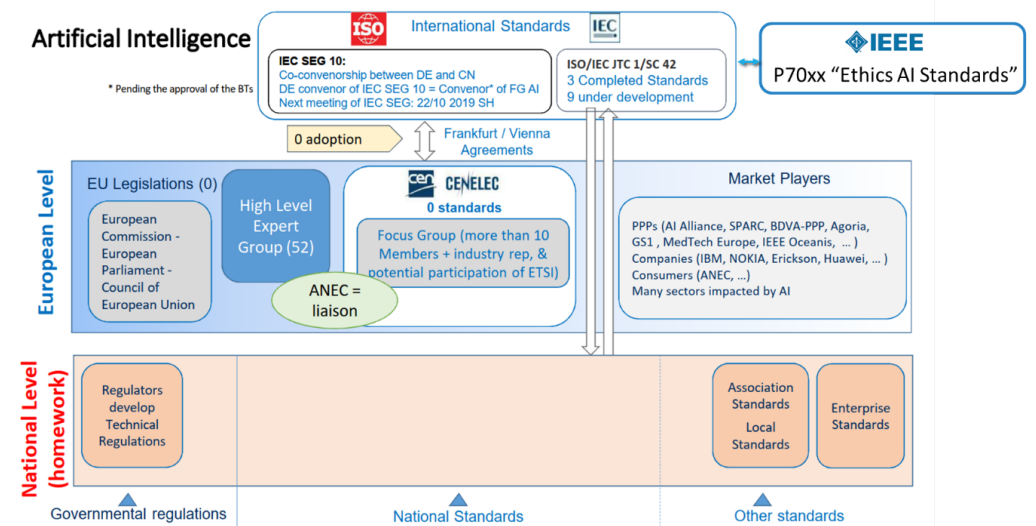## SUPPLY SIDE SOLUTIONS

HOME   WHAT WE DO   SERVICES   ORCAA   CONTACT   PEOPLE   ARTICLES

O'Neil Risk Consulting & Algorithmic Auditing

IT'S THE AGE
ALGORITHM A
HAVE ARRIVE
UNPREPARED

PEOPLE

AI Ethics at Facebook

ALGORITHMS   DATA

Self-regulation (reputation / compliance)

COMPANY LEVEL

PRIVACY AND SECURITY

Google Plans Not to Renew Its Contract for Project Maven, a Controversial Pentagon Drone AI Imaging Program

Kate Conger

Apple Is Trying to Make Your iPhone Less Addictive

Harrison Weber
6/04/18 2:18pm · Filed to: APPLE

Former Google and Facebook Staff Launch Anti-Addiction Lobby
February 5, 2018

BUSINESS INSIDER UK — TECH

Over 100 Amazon employees, including senior software engineers, signed a letter asking Jeff Bezos to stop selling facial-recognition software to police

BRANCH/SECTOR LEVEL - STANDARDS

State Intervention (compliance)

CO-REGULATION

The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)

Developing metrics and processes towards the implementation of a certification methodology addressing transparency, accountability and algorithmic bias

< Industry Connections

The Ethics Certification Program for Autonomous and Intelligent Systems

About

The goal of The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) is to create specifications for certification and marking processes that advance transparency, accountability and reduction in algorithmic bias in Autonomous and Intelligent Systems (A/IS).

Privacy Shield Framework

LEGISLATIVE ACTION

Could New York City's AI Transparency Bill Be a Model for the Country?

The recently-passed legislation is aimed at ensuring government is not "black boxed."

BY ELIZABETH ZIMA / JANUARY 4, 2018

problem sol

langu

algo

step

lem

26 January 2016

The Digital Bill

—

The National Assembly has just passed the Bill for a Digital Republic. The bill was jointly prepared in consultation with citizens over a period of 3 weeks. An unprecedented method!

IMPRIMER

Market & Reputation driven RI services

Lack of trust is impacting investments in AI

User mistrust and aversion is slowing down the growth of AI

It can take only one mistake -- or a perception of a mistake – for a user to stop trusting AI

Lack of trust in AI is #1 barrier for deployment of AI in organizations

Sustaining trust in AI will be a long-term struggle and current governance practices are not keeping up

only **9%** of consumers would allow AI to run their **finances**, and only 4% would trust it with **HR-related work**
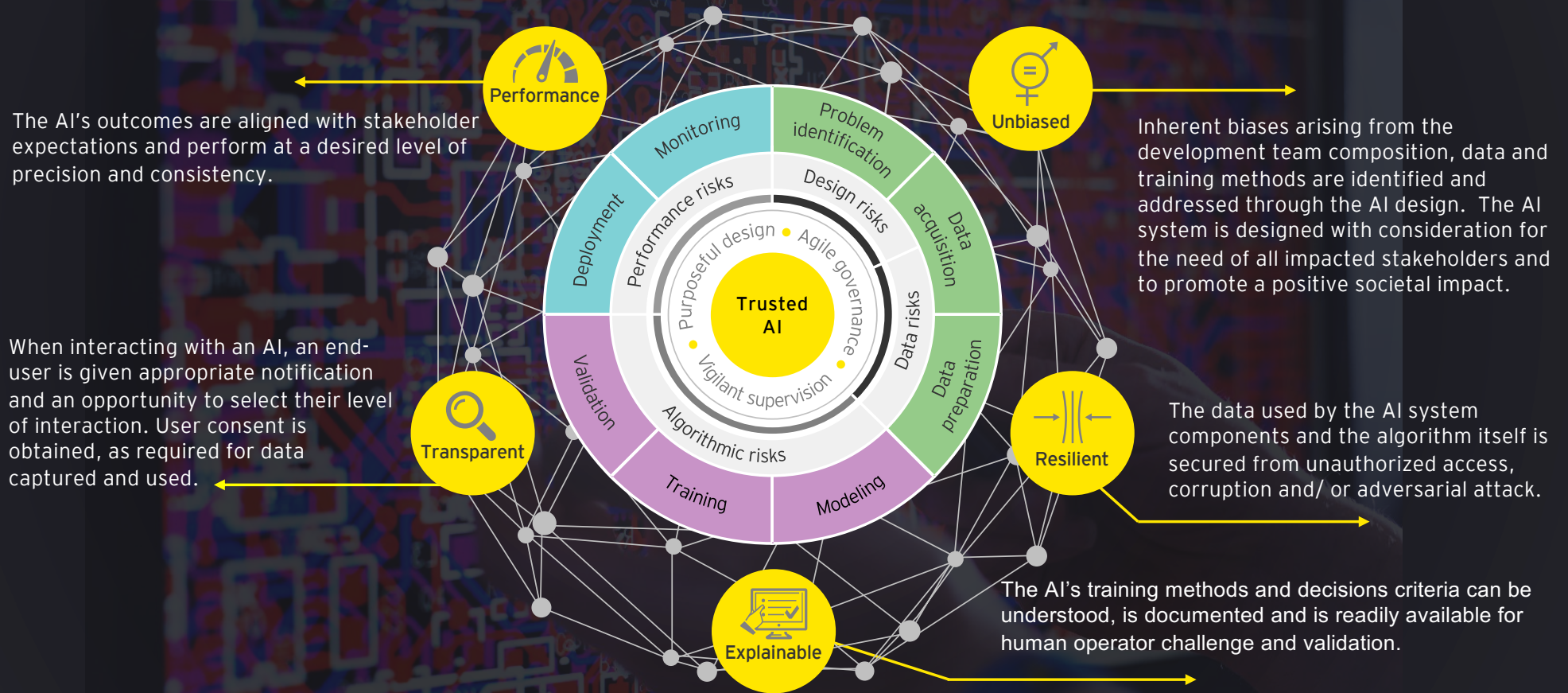
Source: InsideSales.com 2017 AI user survey

85% of AI projects through 2022 will deliver **erroneous outcomes** due to bias in data, algorithms or development teams
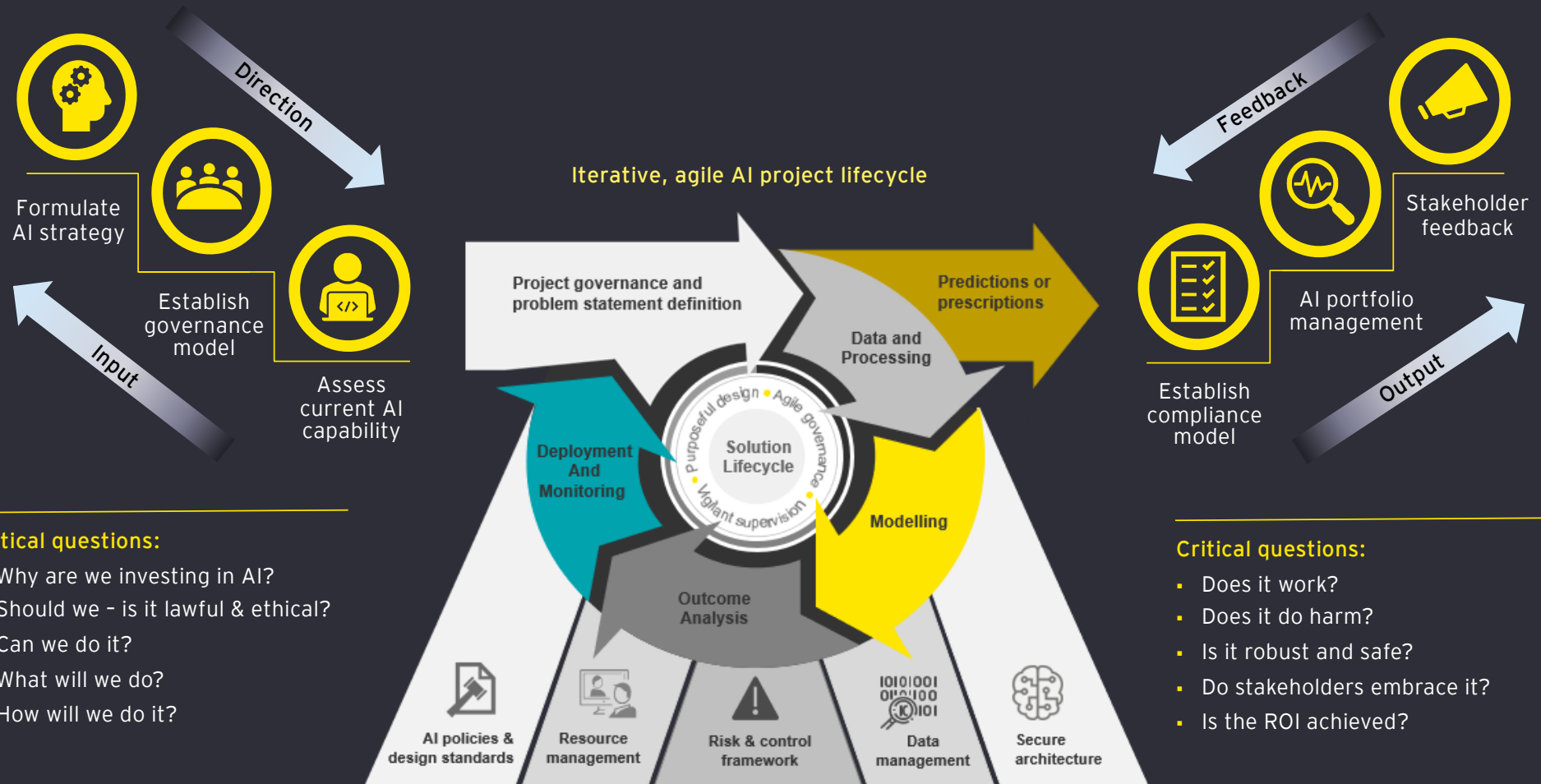
Source: Gartner's 2018 CIO Agenda Survey

Trusting AI will require expanding the risk and control attributes

EY's Trusted AI Framework emphasizes the following five attributes necessary to sustain trust.

Performance
The AI's outcomes are aligned with stakeholder expectations and perform at a desired level of precision and consistency.

Unbiased
Inherent biases arising from the development team composition, data and training methods are identified and addressed through the AI design. The AI system is designed with consideration for the need of all impacted stakeholders and to promote a positive societal impact.

Transparent
When interacting with an AI, an end-user is given appropriate notification and an opportunity to select their level of interaction. User consent is obtained, as required for data captured and used.

Resilient
The data used by the AI system components and the algorithm itself is secured from unauthorized access, corruption and/ or adversarial attack.

Explainable
The AI's training methods and decisions criteria can be understood, is documented and is readily available for human operator challenge and validation.

Trusted AI

Purposeful design • Agile governance
Vigilant supervision

Performance risks
Design risks
Data risks
Algorithmic risks

Monitoring
Problem identification
Deployment
Data acquisition
Validation
Data preparation
Training
Modeling

EY

An agile governance model is needed to address the high-pace dynamics of AI

Direction

Feedback

Formulate AI strategy

Establish governance model

Assess current AI capability

Input

Iterative, agile AI project lifecycle

Project governance and problem statement definition

Data and Processing

Predictions or prescriptions

Deployment And Monitoring

Solution Lifecycle

- Purposeful design • Agile governance • Vigilant supervision

Modelling

Outcome Analysis

AI policies & design standards

Resource management

Risk & control framework

Data management

Secure architecture

Stakeholder feedback

AI portfolio management

Establish compliance model

Output

Critical questions:
- Why are we investing in AI?
- Should we - is it lawful & ethical?
- Can we do it?
- What will we do?
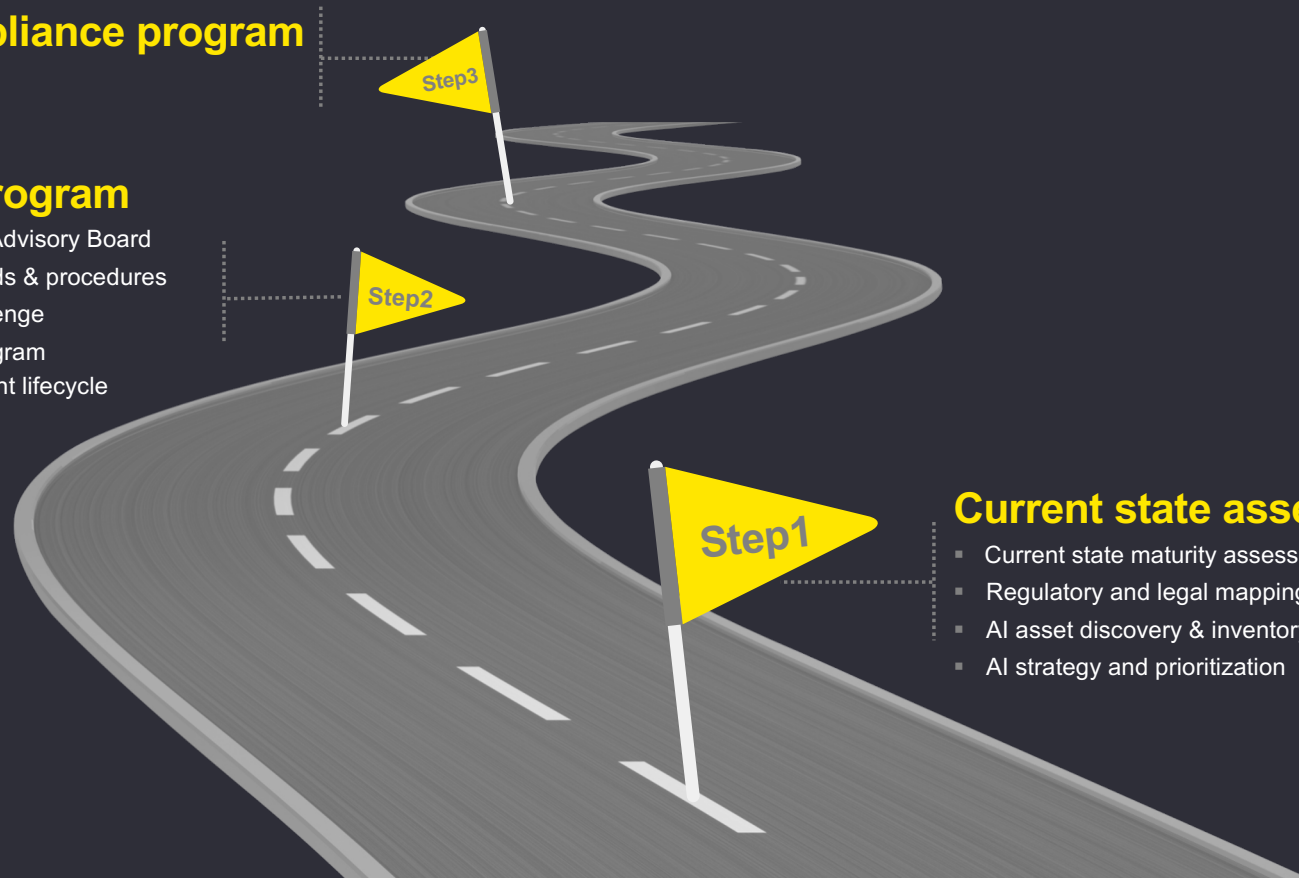- How will we do it?

Critical questions:
- Does it work?
- Does it do harm?
- Is it robust and safe?
- Do stakeholders embrace it?
- Is the ROI achieved?

EY

High-level roadmap to a robust AI compliance program

Operate AI compliance program

Step3

Design AI compliance program
- Develop AI organization, incl. Ethics / AI Advisory Board
- Create ethical AI policies, design standards & procedures
- Develop AI compliance checklists + challenge
- Create independent model validation program
- Map compliance to AI system development lifecycle
- Training & awareness

Step2

Step1

Current state assessment
- Current state maturity assessment
- Regulatory and legal mapping
- AI asset discovery & inventory management
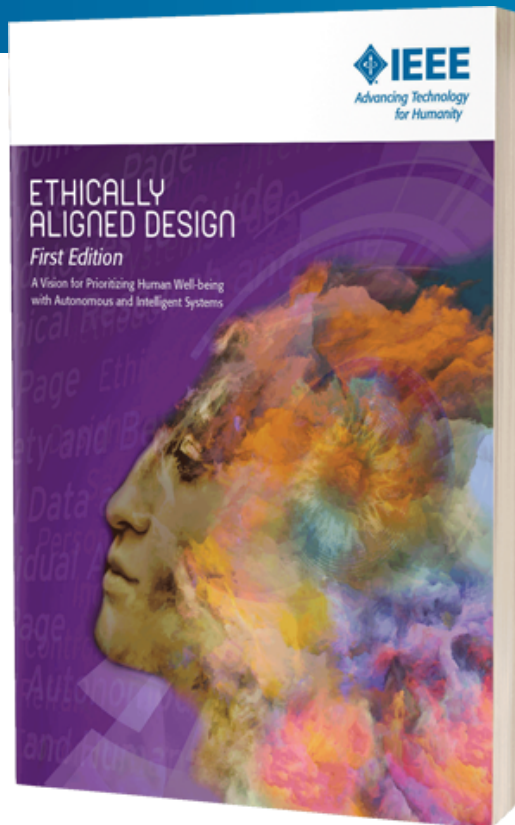- AI strategy and prioritization

EY

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

An incubation space for new standards and solutions, certifications and codes of conduct, and consensus building for ethical implementation of intelligent technologies

IEEE P7000: Model Process for Addressing Ethical Concerns During System Design

IEEE P7001: Transparency of Autonomous Systems

IEEE P7002: Data Privacy Process

IEEE P7003: Algorithmic Bias Considerations

IEEE P7004: Child and Student Data Governance

…

IEEE P7008: Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems

IEEE P7009: Fail-Safe Design of Autonomous and Semi-Autonomous Systems

IEEE P7010: Wellbeing Metrics Standard for Ethical AI and Autonomous Systems
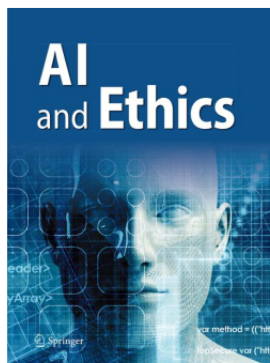
IEEE P7011: Process of Identifying and Rating the Trustworthiness of News Sources

IEEE P7013: Benchmarking of Automated Facial Analysis Technology

Algorithmic systems are socio-technical

- Algorithmic systems do not exist in a vacuum
- They are built, deployed and used:
  - by people,
  - within organizations,
  - within a social, political, legal and cultural context.

- The outcomes of algorithmic decisions can have significant impacts on real, and possibly vulnerable, people.

Thank you

Springer

AI and Ethics

Editors-in-Chief:

John MacIntyre
University of Sunderland, UK

Larry Medsker
The George Washington University, USA

ISSN: 2730-5953
eISSN: 2730-5961
2021: Volume 1, 4 issues
www.springer.com/43681

Don't forget to sign up for free Article Alerts to be informed whenever new articles are published!

AI and Ethics seeks to promote informed debate and discussion of the ethical, regulatory, and policy implications that arise from the development of AI. It will focus on how AI techniques, tools, and technologies are developing, including consideration of where these developments may lead in the future. The journal will provide opportunities for academics, scientists, practitioners, policy makers, and the public to consider how AI might affect our lives in the future, and what implications, benefits, and risks might emerge.

**IEEE P7003™,
Standard for Algorithmic
Bias Considerations
Working Group**

IEEE Computer Society/Software & Systems Engineering Standards Committee (C/S2ESC)

https://site.ieee.org/sagroups-7003/

EY

https://www.ey.com/en_gl/ai