Session 602: *Ethical Considerations of Artificial Intelligence*

Colleen Dorsey
Director, Organizational Ethics & Compliance, University of St. Thomas School of Law
Eileen M. Lach
Executive Committee, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Tuesday, October 23, 2018, 1:00 pm

IEEE

What is Artificial Intelligence?

IEEE

Look to Shakespeare

"To be, or not to be, that is the question:

Whether 'tis nobler in the mind to suffer the slings and arrows of outrageous fortune,

Or to take arms against a sea of troubles

And by opposing end them."

Hamlet, Hamlet's Soliloquy, Act III, Scene I

2

Look to Developments in Technology

- The theory and development of computer systems able to perform tasks that normally require human intelligence
- Examples:
  - Visual Perception
  - Speech Recognition
  - Decision-Making
  - Translation Between Languages
- Beyond Machine Learning
  - Artificial General Intelligence (AGI) – Capable of sequential learning through the retention of past task
  - Artifical Superintelligence (ASI) – Capable of reprogramming and improving itself
  - Extended Intelligence – Machine assistance to ecosystem of society

3

Hypothetical

ABC Corporation is a mid-sized public manufacturing company in the United States. The CEO has watched the stock plummet for companies that have had their customer files hacked, made assumptions about consumer preferences, been sued for discriminatory hiring practices, and embedded chips in their employees wrists to make cafeteria payments easier. As a 21st century company, the CEO and the COO have been told they need to develop a compliance program that embeds the ethical use of Artificial Intelligence. The Board is supportive and encourages this direction. The CEO's goal is to enhance the image of the company as a good corporate citizen, avoid the negative publicity that has plagued other companies and prevent losses in shareholder value (not necessarily in that order). As the Chief Ethics & Compliance Officer, the CEO comes to you to help determine how ABC Corporation might begin the process of enhancing its ethics and compliance program to account for AI. Where do you begin?*

*The presenters are assuming for the purposes of this hypothetical that ABC Corp. has an existing E&C program and that the direction of including A/IS is one that is a strategic priority for ABC Corp based on direction and support of the board.

IEEE

4

A Starting Point

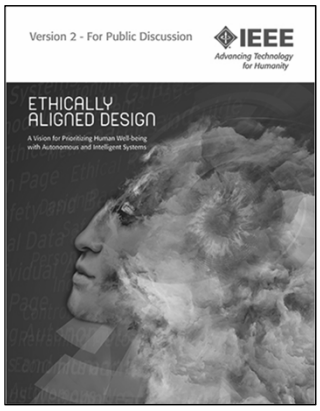
- Research:
  - The IEEE Ethically Aligned Design – v2. https://standards.ieee.org/industry-connections/ec/auto-sys-form.html.
  - Everyday Ethics for Artificial Intelligence – A practical guide for designers & developers, IBM, https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf
  - Self Study

IEEE

5

Ethically Aligned Design

A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems



Version 2

- Launched December 2017 as a Request for Input
- Created by over 250 Global A/IS & Ethics professionals, in a bottom up, transparent, open and increasingly globally inclusive process.
- Incorporates over 200 pages of feedback from public RFI and new Working Groups from China, Japan, Korea and more.
  ~ 300 pages very pertinent input received for EAD V2.
- Contains **over one hundred twenty** key Issues and Candidate Recommendations
- Version 3 to be published in Q1 2019, with over 1,000 contributors and reviewers

IEEE

Inspiration for IEEE P7000 Working Groups

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™ - Standard on Child and Student Data Governance

IEEE P7005™ - Standard on Employer Data Governance

IEEE P7006™ - Standard on Personal Data AI Agent Working Group

IEEE P7007™ - Ontological Standard for Ethically driven Robotics and Automation Systems

IEEE P7008™ - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems

IEEE P7009™ - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems

IEEE P7010™ - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems

IEEE P7011™ - Standard for the Process of Identifying and Rating the Trustworthiness of News Sources

IEEE P7012™ - Standard for Machine Readable Personal Privacy Terms

IEEE P7013™ - Inclusion and Application Standards for Automated Facial Analysis Technology.

IEEE

A Starting Point continued

- Appointment of A/IS Master Work Group
  - Include diverse representatives from:
    - ✓ Ethics & Compliance
    - ✓ HR
    - ✓ Marketing
    - ✓ Sales
    - ✓ Finance
    - ✓ Operations
    - ✓ Audit
    - ✓ Legal
    - ✓ Environmental, Health & Safety
    - ✓ Product/Research Development (from each major product unit or division)

IEEE

8

A Starting Point continued

- Broad issues for A/IS Master Work Group to consider:
  - Time Lines – for completion of goals; for reporting
  - Resources – do they have the right talent?; enough people to do the work? do they have the budget?
  - Authority:
    - ✓ who has the ultimate say in how to move forward?
    - ✓ how and when will progress be reported to the board?
    - ✓ are further governance actions necessary to allow for the A/IS Master Work Group to move forward: policies? Code of Conduct updates? board subcommittees?
  - Scope – what are the leader groups responsible for and how broad or narrow is that responsibility?
  - Deliverables – what does the board and leadership want as deliverables from the A/IS Working Group?
  - Education/training requirements – are there courses or programs engineers, technologists, lawyers, etc. can take to get them grounded in this subject?
  - Incentive development for each group within ABC Corp to ensure ethical concerns are raised as a matter of course

IEEE

9

A Starting Point continued

- Assign specific Leader Groups from within the A/IS Master Work Group to more fully explore *general and ethically focused principles* which apply to all types of autonomous and intelligent systems (A/IS):
  - Human Rights Leader Group– consider reps from HR, Product/Research Development, Legal, E&C
  - Well-Being Leader Group– consider reps from all - leads could be HR and E&C, EHS
  - Accountability Leader Group – consider reps from Product/Research Development, Legal, Operations, E&C, EHS
  - Transparency Leader Group – consider reps from Legal, Product/Research Development, Operations, E&C
  - Misuse and Awareness Leader Group– consider reps from Legal, Product Development, Marketing, Sales, Operations, E&C, EHS

IEEE

Human Rights Leader Group

- Example areas for this leader group to explore:
  - Which human values and legal norms should be promoted in the design of A/IS systems?
  - What are the international human rights standards which may impact our A/IS products?
  - Who should be included in the team to further develop the human rights aspect? (should be diverse teams of different cultures, ages, ethnicities, genders and educational backgrounds)
  - Should there be a feedback mechanism installed to keep dialogue open with users to raise awareness of user-identified bias?
  - How should we train and test for bias?
  - How do we audit for unintentional bias in the design and development of A/IS?
  - How might cultural (and other) bias unintentionally inform information flows, information systems, algorithmic decision-making and value by design> Examples:
    - ✓ Algorithms in developing customer preference data
    - ✓ Stereotyping
    - ✓ Sunk cost bias
    - ✓ Confirmation bias
    - ✓ Hiring practices:
      - o Algorithmic bias in resume selection
      - o Sensitive personal data

IEEE

Well-Being Leader Group

- Example areas for this leader group to explore:
  - Employee Well-Being:
    - ✓ What security systems do we want/need on ABC Corp. computers?
    - ✓ Consider models of consent for data collection and agency
    - ✓ Consider how and if to gather subjective well-being measurements based on company values such as:
      - oMental health
      - oEmotions
      - oSense of self
      - oAutonomy
      - oAbility to achieve their goals
      - oOther dimensions of well being
      - oPersonal, environmental and social factors
  - Consider use of "nudging" with consent
  - Consider how to establish norms/embedding values of the relevant user group into A/IS
    - ✓ Which norms will govern? Consider the culture you are designing within
    - ✓ How will you update continually changing norms?
    - ✓ How will you resolve norm conflicts?

12

Accountability Leader Group

- Example areas for this leader group to explore:
  - ABC Corp. should have clear policies about who is accountable for product and design of A/IS
    - ✓ Where does responsibility begin and end?
    - ✓ Ensure code of conduct or other policies cover A/IS development processes
  - How will there be access to remedies for consumers and employees following any violation of rights due to use of A/IS?
  - Does accountability change according to user level and their influence over the A/IS system?
  - Privacy – opt in requirements beyond GDPR
- How can oversight of proprietary manufacturing processes keep the information confidential and yet be transparent to safeguard the public?
  - Look to biomedical, civil and aerospace engineering for ideas on how to balance these competing issues
- Determine if ABC Corp. should consider if multi stakeholder ecosystems will be necessary to include reps from law enforcement, insurers, manufacturers, engineers, lawyers

13

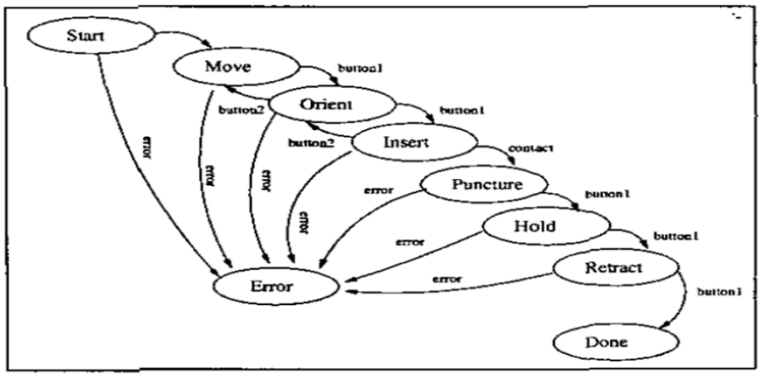
Transparency Leader Group

- Example areas for this leader group to explore:
  - Self assessing transparency standards
    - ✓ Explicitly present empirical evidence of the consideration of ethical values
    - ✓ Present methodology used such as data used to train the system, algorithms and components used and results of behavior monitoring
    - ✓ This documentation should be subject to auditability
    - ✓ Documentation should also be required to be accessible, meaningful and readable
  - Record keeping necessary:
    - ✓ intended use documented
    - ✓ training data and environments (if applicable)
    - ✓ data sources
    - ✓ algorithms
    - ✓ process graphs
    - ✓ model features
    - ✓ user interfaces
    - ✓ outputs
    - ✓ reward/optimization function
    - ✓ Post-launch documentation regarding ethical design choices and considerations
  - Systems design:
    - Provide transparent signals such as explanations or inspection capabilities
    - Look to algorithmic traceability in black-box software systems but with caution and with another independent means of validating results and detecting harms

IEEE

Engineering Process Task Graph Example

IEEE



IEEE

Misuse and Awareness Leader Group

- Example areas for this leader group to explore:
  - Depending upon the level of the A/IS (develop internal levels to trigger certain processes), product developers may need to convene user/consumer focus groups to develop education and security awareness and appropriate warnings (groups would include potential users, government and enforcement agency reps)

16

Thank you!