Technology Ethics Standardization

IEEE Summit on Communications Futures
A G Hessami, ECPAIS VC & Process Architect
18 Jan 2020, Honolulu
Overview of the Talk

➢ Definitions and Concepts
➢ Current Technology Challenges
➢ IEEE Global Initiative
➢ The Ethics Certification Programme
➢ Way Forward
Definitions

➢ **Ethics**
   - A branch of knowledge that deals with moral principles (that govern a person’s behaviour)

➢ **Morals**
   - Considerations of right & wrong behaviour (acceptable in a particular society/culture)
Principal Ethics Theories

- **Consequentialism/Utilitarianism (JS Mill)**
  - Deals with Happiness & Well being
  - Everyone ought to act to bring greatest happiness for greatest No. of People

- **Deontological/Duty (E Kant-Categorical Imperative)**
  - put yourself as a universal law maker and whether the object will influence people, destroy, threaten or create values?
  - Also WD Ross on prima facie duties covering Fidelity, Reparation, Gratitude, Promoting Max Good & non-Maleficence
  - Top Management’s Personal Ethics impacting on decisions to align with stakeholders expectations

- **Virtue Ethics (Aristotle)**
  - What stakeholders are affected? Virtues are character qualities borne by persons
  - How is virtuous behavior impacted?
  - Vice is opposite to virtue but core principle is to “be good” hence Virtue
We are all dependent on Technology
Our influence on these is declining/vanishing
Decisions on functionality, pricing and customization is made by Multinational Enterprises (MNEs)
Technology & Apps pose unknown Risks
There’s insufficient regulation and control
Context II

- Our interactions are our Personal Profile
- Our activities are monitored by Apps
- Our Transactional Data is of commercial value/interest
- Not all services are honest and transparent to the users
There’s a need for Transparency of Services
There’s a need for Accountability of Service Providers
There’s a need for Explainability of Automation
Not every immoral behavior is Illegal!
There’s a need for a new responsive paradigm
Technology Ethics

I. Algorithmic Bias

II. Transparency

III. Accountability

IV. Privacy

V. Safety & Security
## IEEE Initiative - Ethically Aligned Design (EAD)

### EAD Pillars

<table>
<thead>
<tr>
<th>EAD General Principles</th>
<th>Universal Human Values</th>
<th>Political Self-Determination Data Agency</th>
<th>Technical Dependability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Rights</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well-being</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Agency</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparency</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accountability</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness of Misuse</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Ethically Aligned Design-Principles

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 P7000- Standard
Model Process for Addressing Ethical Concerns During System Design

➢ Purpose

- help to create a shared mission around values, value priorities and value harms to avoid and
- help to assure value based system engineering, by building a bridge between the value mission and the actual development of a system.
Draft IEEE P7000 - Process

➢ **Ethical Mission Analysis**
  - Define System of Interest and Context of Use
  - Collect unstructured Ideas as Harms & Benefits and weight these
  - Use Life Cycle as per ISO/IEC TR24748-1:2010

➢ **Value Needs Refinement**
  - Identify Harms and Benefits’ Underlying Values
  - Conceptually Investigate Value Qualities

➢ **Risk Based Value Sensitive Design**
  - Treat Values as System Qualities at Risk
  - Analyse, Evaluate and Treat Values at Risks
  - Identify and select System Controls for all System Qualities at Risk
IEEE P7000™ Standardization Projects

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 P7014™ - Standard for Ethical considerations in Emulated Empathy in Autonomous and Intelligent Systems
The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS) has the goal to create specifications for certification and marking processes that advance transparency, accountability, and reduction in algorithmic bias in autonomous and intelligent systems. ECPAIS intends to offer a process and define a series of marks by which organizations can seek certifications for their processes around the A/IS products, systems, and services they provide.

For More Information:
standards.ieee.org/industry-connections/ecpais.html
The Ethics Certification Program for Autonomous and Intelligent Systems

**Expert Focus Groups - Formation**

1) Experts to self-declare interests and competencies

2) Formed Three Expert Focus Groups for workstreams:
   1) Accountability (AEFG)
   2) Algorithmic Bias (BEFG)
   3) Transparency (TEFG)

3) Assigned Members, leads, support etc.

4) Identified Gaps, Overlaps, need for extra resources, ....
The Ethics Certification Program for Autonomous and Intelligent Systems

Expert Focus Groups - Working Process

- Appointed a Chair, Secretary and a panel of experts
- Developed 6 months Work Plan
- Key Tasks were:
  - Develop Expert Competence Criteria and Invitation to Experts
  - Develop Terms of Reference (ToR)
  - Develop/Adopt Specific Reference Model from ECPAIS Generic Reference Model
  - Develop Ethical Foundational Requirements (EFRs) & Validate
  - Develop Satisfaction Criteria for EFRs
  - Plan/initiate pilot trials and refine the EFRs
  - Develop & release final Certification guidelines by 2019 Q4
Core Process/Reference Model & Workstreams

➢ There’s a need for Generic Reference Model (GRM) to provide a common thread in the whole ECPAIS’ work
➢ The ECPAIS contexts will be tailored derivatives as Specific Reference Models (SRMs) for each Workstream
➢ There’s also a Terms of Reference for each workstream to ensure clarity, transparency & provide a schedule
**The Ethics Certification Program for Autonomous and Intelligent Systems**

**Expert Focus Groups Terms of Reference - ToR**

---

### ECPAIS Accountability Expert Focus Group

**Issue 1 Draft 2, April 2019**

**Terms of Reference**

<table>
<thead>
<tr>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Remit</td>
</tr>
<tr>
<td>- Develop accountability requirements for autonomous intelligent systems (A/IS)</td>
</tr>
<tr>
<td>- Focus on accountability requirements during the design and verification, manufacture and deployment stages</td>
</tr>
<tr>
<td>2. Scope</td>
</tr>
<tr>
<td>- Identify System Level Accountability Requirements (SLAR) design, testing and verification, manufacture/production and deployment stages</td>
</tr>
<tr>
<td>- Ensure relevance to Products, Systems and Services in A/IS</td>
</tr>
<tr>
<td>- Adopt the P7000 Core Process Model or an equivalent Accountability in A/IS implementations</td>
</tr>
<tr>
<td>- Develop SAR structure:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

---

### ECPAIS Algorithmic Bias Expert Focus Group

**Issue 1 Draft 1, April 2019**

**Terms of Reference**

<table>
<thead>
<tr>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Remit</td>
</tr>
<tr>
<td>- Requirements for identifying illegal/undesirable bias in algorithmic systems</td>
</tr>
<tr>
<td>- Requirements for a system that minimises unintended or unjustified bias</td>
</tr>
<tr>
<td>- Application of Value Sensitive Design &amp; EAD principles</td>
</tr>
<tr>
<td>2. Scope</td>
</tr>
<tr>
<td>- Cover system design &amp; development processes</td>
</tr>
<tr>
<td>- Cover system deployment processes</td>
</tr>
<tr>
<td>- Cover system monitoring and recall processes</td>
</tr>
<tr>
<td>- Address Human (e.g. team diversity, Data, and Compute optimization target definitions) levels</td>
</tr>
<tr>
<td>- Address products, systems and processes</td>
</tr>
</tbody>
</table>

---

### ECPAIS Transparency Expert Focus Group

**Issue 1 Draft 2, April 2019**

**Terms of Reference**

<table>
<thead>
<tr>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Remit &amp; Aim</td>
</tr>
<tr>
<td>- Develop transparency (inclusive of explainability) requirements for autonomous and intelligent systems (A/IS) certification</td>
</tr>
<tr>
<td>- Focus on transparency requirements during the design, testing, verification, manufacture and deployment stages of A/IS</td>
</tr>
<tr>
<td>2. Scope</td>
</tr>
<tr>
<td>- Examine extant draft AI Ethics Principles from various sources (National Governments, Department of Defense, European Unions, Standards Bodies, etc.), select articles that relate to transparency, and distill them into a succinct working list that can succinctly cover the main points</td>
</tr>
<tr>
<td>- Referring to this research list, then:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>- Develop STR structure:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>- Develop Transparency Specific Reference Model (TSRM) from ECPAIS Model</td>
</tr>
</tbody>
</table>
The Ethics Certification Program for Autonomous and Intelligent Systems

Expert Focus Groups – Accountability Ethical Foundational Requirements
The Ethics Certification Program for Autonomous and Intelligent Systems

Expert Focus Groups – Transparency Ethical Foundational Requirements
The Ethics Certification Program for Autonomous and Intelligent Systems

Expert Focus Groups – Algorithmic Bias Ethical Foundational Requirements
The Ethics Certification Program for Autonomous and Intelligent Systems

**Schedule & Targets**

- All three Expert Focus Groups progressed on Target
- Final Outcomes by early December 2019
- Process Trials:
  - Members Volunteered real-world Use Cases/Case Studies
  - ECPAIS supported Use Case Specification with bespoke Template
  - Interim Trials run by any of the Expert Focus Groups’ Processes
  - Outcome shared with Members as Value from Support & Participation
  - Any Learning taken Onboard with Processes Enhancements
  - Can revisit Trials at the end of Expert Focus Group Work as Appropriate
**Ethical Foundational Requirements – Transparency Example**

<table>
<thead>
<tr>
<th>Goal</th>
<th>Requirement</th>
<th>Measurement</th>
<th>Acceptable Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1- Suitable &amp; Sufficient Management</td>
<td>The duty holder shall fulfill the following Transparency Requirement(s);</td>
<td>Two-tier approach to encourage adoption:</td>
<td>The following item(s) shall be presented as evidence for conformity against the Transparency Requirement(s):</td>
</tr>
<tr>
<td></td>
<td>- Demonstrate that a suitable and sufficient organizational governance framework is in place reflecting capability, maturity, and processes to ensure legal responsibility and ethical accountability.</td>
<td>1. Top-level finding: “no critical findings in the detailed requirements” / “areas requiring attention for improvement”</td>
<td>• Organigramme showing lines of responsibility and accountability including the supply chain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Organizational readiness finding: on 1-5 scale (based on aggregate of satisfying sub level goals) such as:</td>
<td>• Designated positions for risk management, data protection compliance, legal compliance, stakeholder management, and ethical profile management and coordination across all roles</td>
</tr>
<tr>
<td></td>
<td>5- excellent</td>
<td></td>
<td>Minimum assessment requirements comprising a) sector risks, including web-based global operation risks; b) potential harms/adverse impacts from AI/IS; c) end-user needs (e.g. privacy); and d) supply chain awareness and compliance with minimum assessment requirements</td>
</tr>
<tr>
<td></td>
<td>4- good</td>
<td></td>
<td>• Implementation of local laws and requirements relevant above minimum assessment requirements</td>
</tr>
<tr>
<td></td>
<td>3- average</td>
<td></td>
<td>Overall legal compliance (dependent on cross-jurisdictional reach</td>
</tr>
<tr>
<td></td>
<td>2- poor</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Ethics Certification Program for Autonomous and Intelligent Systems

**ECPAIS Approach**

- Three suites of A/IS Ethics Certification Criteria based on Expert Schema
- Associated Satisfaction Criteria per Criterion
  - Ethical Foundational Requirements - EFRs
  - Claim-Evidence for Conformity
  - Evaluation scale, units and Benchmarks
- Guidelines for key Stakeholders
  - Developers, Integrators, Operators & Regulators
- Three Levels for Conformity & Certification according to Impact
  - *Baseline (L1), Compliant (M1), Critical (H1)*
Affinity to Emerging Regulatory Framework
Deliverables for the Programme & all Streams AEFG, BEFG and TEFG

- Bespoke Reference Model and Terms of Reference
- Generic Schema to third level of Taxonomy with Goal Descriptors
- Satisfaction Measures for all Criteria
- Two Comprehensive Reports for each Work Stream
  - *Development Process Reports for Transparency, Accountability & Algorithmic Bias*
  - *Criteria and Satisfaction Measures, Constraints and Way Forward*
- Overall ECPAIS Report and Strategy for next Steps
The Ethics Certification Program for Autonomous and Intelligent Systems

**Next Steps**

➢ The three Certification suites finalized and all Reports Issued

➢ Consultation with ECPAIS Members Continue with Pilots and New Developments

➢ A Press Release will be issued next week

➢ A number of Generic PoC projects are planned for launch in Q1-2020

➢ A number of Sector Tailoring Expert Panels are formed for development of Criteria derivatives
Thank you!

Meeri Haataja
Chair, ECPAIS
meeri@saidot.ai

Ali Hessami
VC & Process Architect, ECPAIS
A.G.Hessami@IEEE.Org
Hessami@VegaGlobalSystems.com