



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



Context

- We are all dependent on Technology
- Our influence on these is declining/vanishing
- Decisions on functionality, pricing and customization is made by Multinationals 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



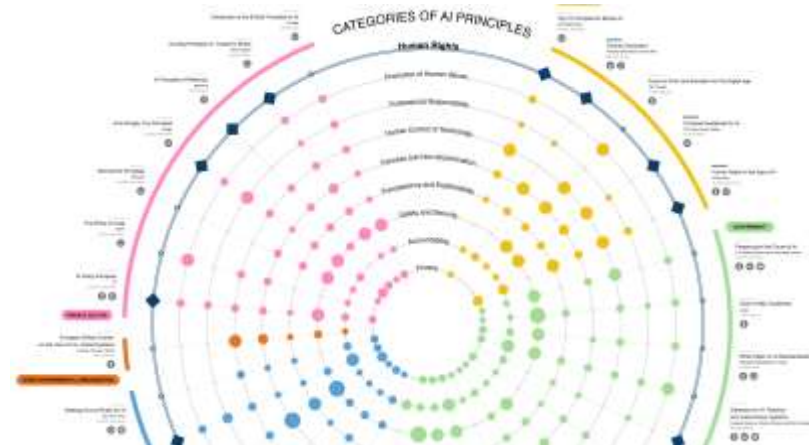
Context III

- 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		
		Universal Human Values	Political Self-Determination Data Agency	Technical Dependability
EAD General Principles	Human Rights	■	■	
	Well-being	■	■	
	Data Agency	■	■	■
	Effectiveness			■
	Transparency	■	■	■
	Accountability	■	■	■
	Awareness of Misuse			■
	Competence			■



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)

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

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

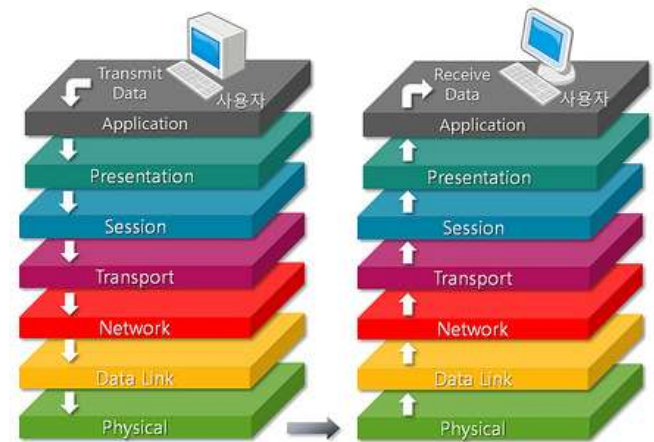
- Appointed a Chair, Secretary and a panel of experts
- Developed 6 months Work Plan
- Key Tasks were;
 - ✓ Develop Expert Competence Criteria and Invitation to Experts
 - ✓ Develop Terms of Reference (ToR)
 - ✓ Develop/Adopt Specific Reference Model from ECPAIS Generic Reference Model
- Develop Ethical 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



The Ethics Certification Program for Autonomous & Intelligent Systems-ECPAIS

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

Expert Focus Group ToR

ECPAIS Accountability Expert Focus Group

Issue 1 Draft 2, April 2019

Terms of Reference

Topics
1. Remit <ul style="list-style-type: none">Develop accountability requirements for autonomous intelligent systems (A/IS)Focus on accountability requirements during the design and verification, manufacture and deployment stages
2. Scope <ul style="list-style-type: none">Identify System Level Accountability Requirements (SLAR) for the design, testing and verification, manufacture/production and deployment stagesEnsure relevance to Products, Systems and Services in A/ISAdopt the P7000 Core Process Model or an equivalent Accountability in A/IS ImplementationsDevelop SAR structure:<ul style="list-style-type: none">ObjectivesFoundational Accountability RequirementsConsolidation of Accountability RequirementsApportionment of Accountability RequirementsDevelop Accountability Specific Reference Model (ASRM) ECPAIS Generic Reference ModelDevelop Accountability Ethical Foundational RequirementsDevelop Accountability Satisfaction Criteria (ASC)Trial/pilot SAR, ASRM, AEFR and ASC on a suite of diverse use cases to ensure coverage, completeness and integrity

Expert Focus Group ToR

ECPAIS Algorithmic Bias Expert Focus Group

Issue 1 Draft 1, April 2019

Terms of Reference

Topics
1. Remit <ul style="list-style-type: none">Requirements for identifying illegal/undesirable bias in algorithmic systemsRequirements for a system that <u>minimises</u> unintended and unjustified biasApplication of Value Sensitive Design & EAD principles
2. Scope <ul style="list-style-type: none">Cover system design & development processesCover system deployment processesCover system monitoring and recall processesAddress Human (e.g. team diversity), Data, and Computational optimization target definitions levelsAddress products, systems and processes

Expert Focus Group ToR

ECPAIS Transparency Expert Focus Group

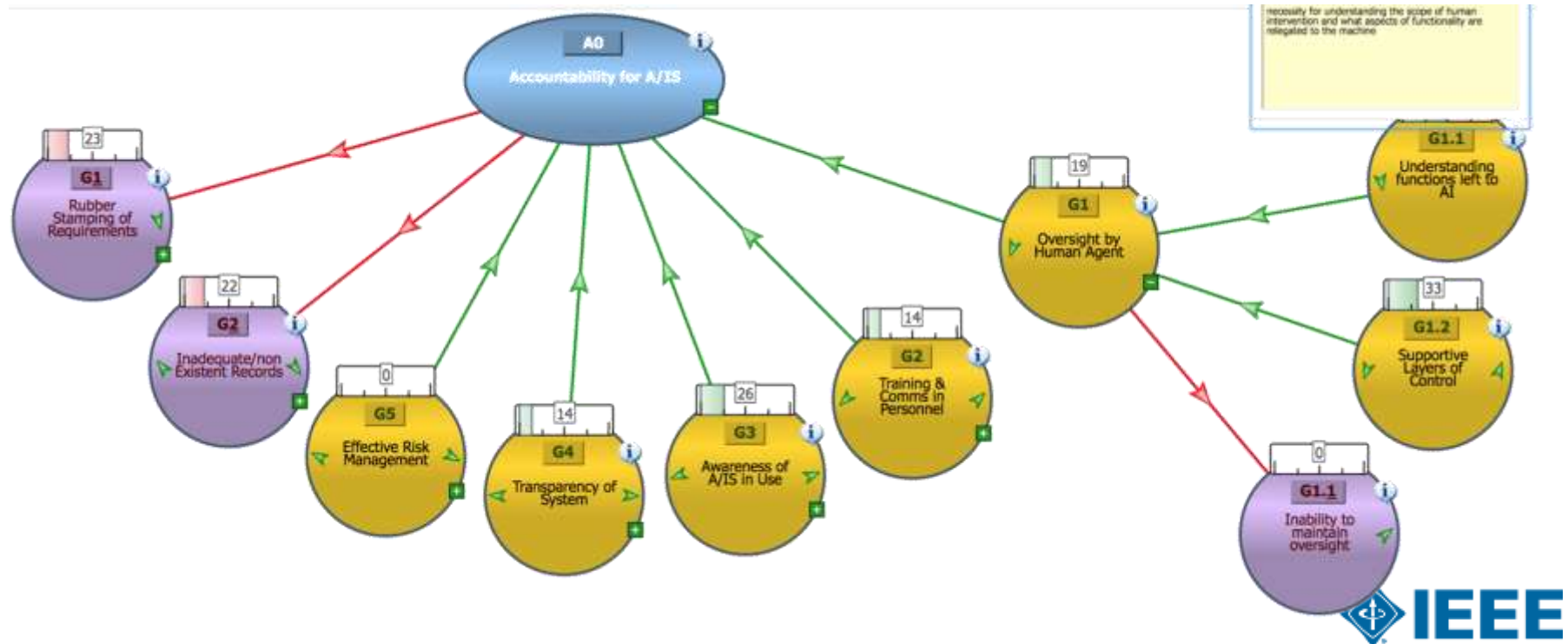
Issue 1 Draft 2, April 2019

Terms of Reference

Topics	Members & Status
1. Remit & Aim <ul style="list-style-type: none">Develop transparency (inclusive of explainability) requirements for autonomous and intelligent systems (A/IS) certificationFocus on transparency requirements during the design, testing, verification, manufacture and deployment stages of A/IS	
2. Scope <ul style="list-style-type: none">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 synthetic working list that can succinctly covers the main points.Referring to this research list, then:<ul style="list-style-type: none">Identify System-Level Transparency Requirements (STR) covering the design, testing and verification, manufacture/production and deployment stages.Ensure relevance to Products, Systems and Services incorporating A/IS.Adopt the P7000 Core Process Model or an equivalent and adapt for Transparency/Explainability in A/IS Implementations.Develop STR structure:<ul style="list-style-type: none">Objectives;Foundational Transparency Requirements;Consolidation of Transparency/Explainability Requirements;Apportionment of Transparency Requirements.Develop Transparency Specific Reference Model (TSRM) from ECPAIS Generic Reference Model.	

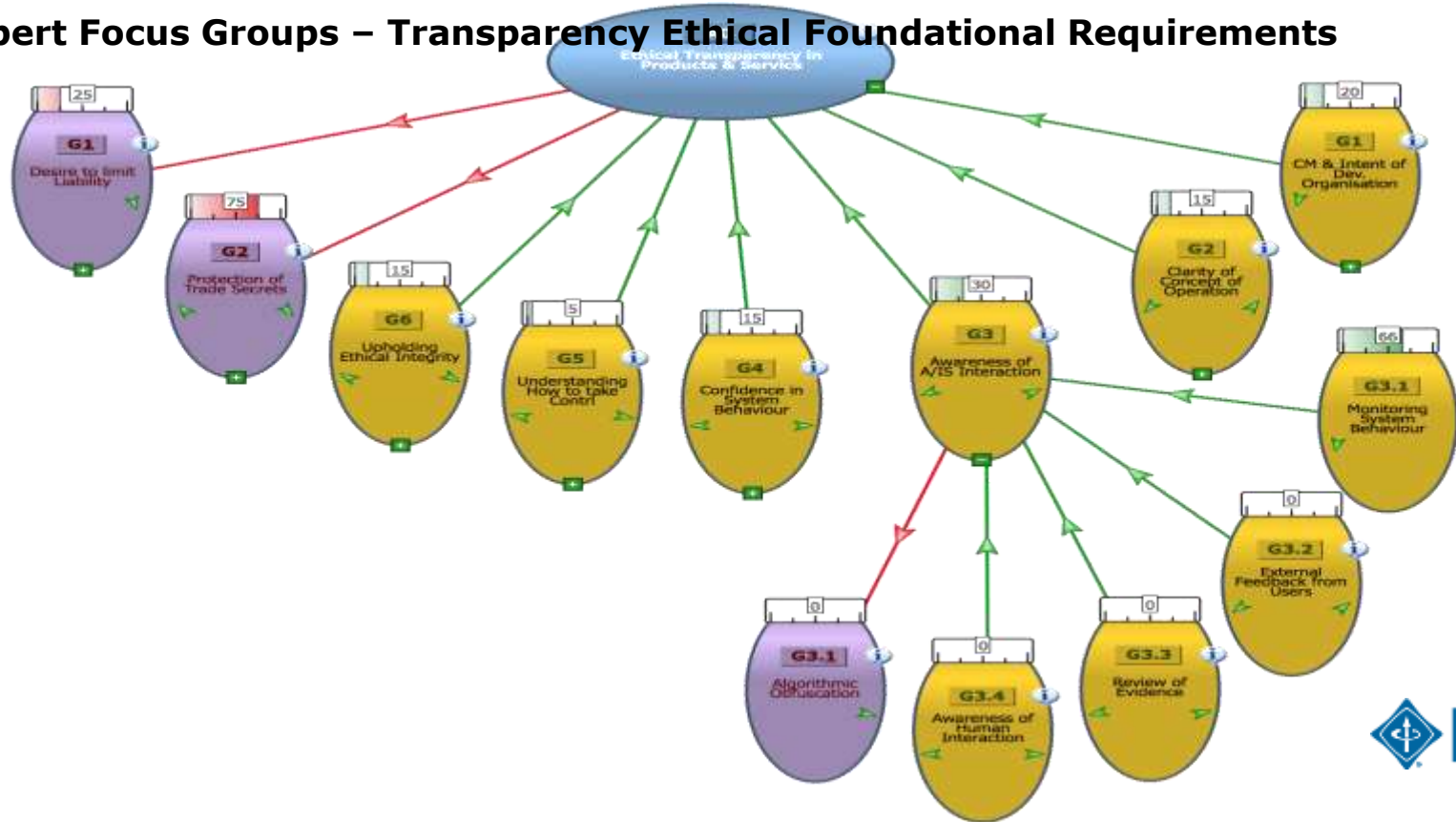
The Ethics Certification Program for Autonomous and Intelligent Systems

Expert Focus Groups – Accountability Ethical Foundational Requirements



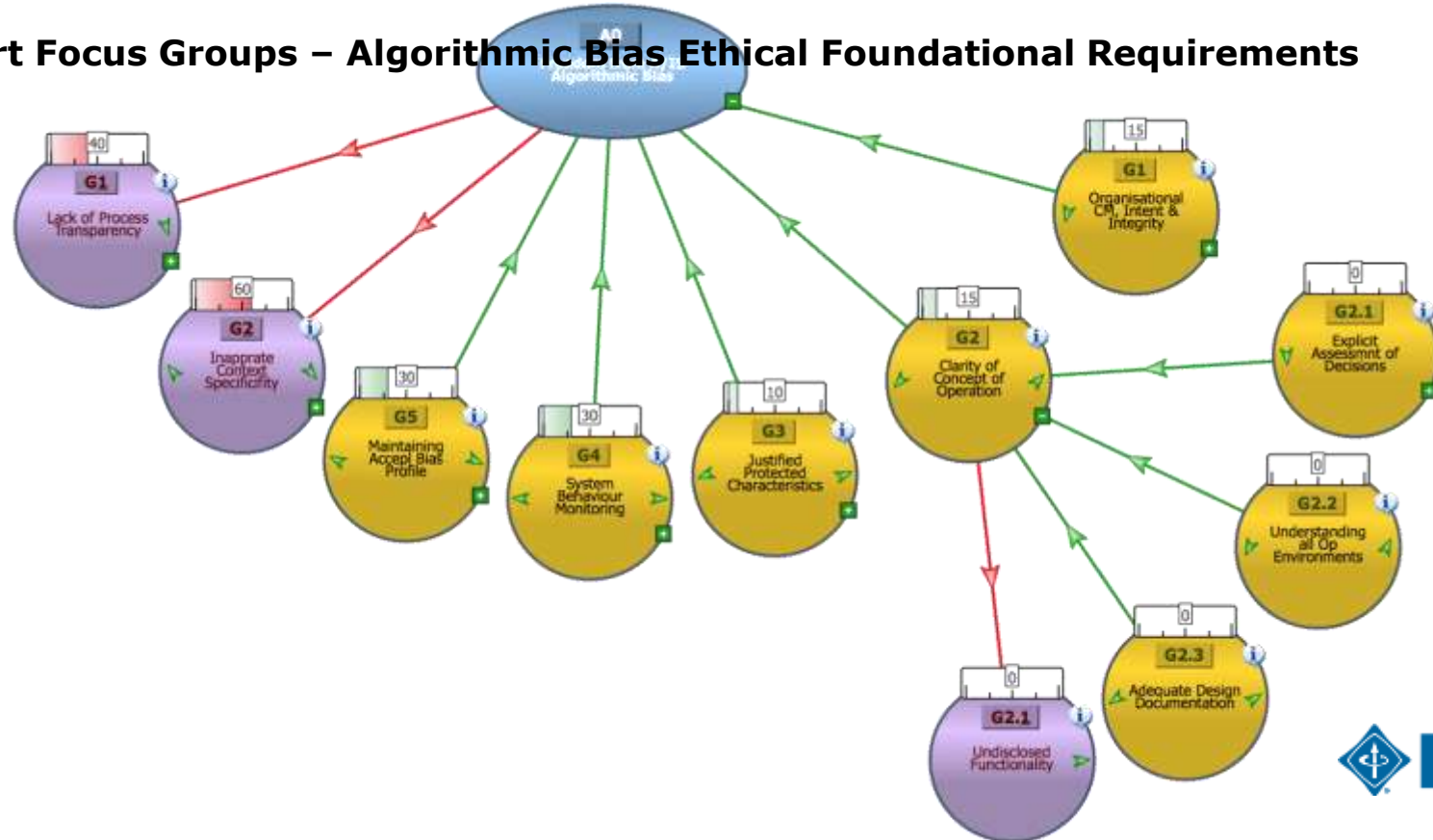
The Ethics Certification Program for Autonomous and Intelligent Systems

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



The Ethics Certification Program for Autonomous and Intelligent Systems

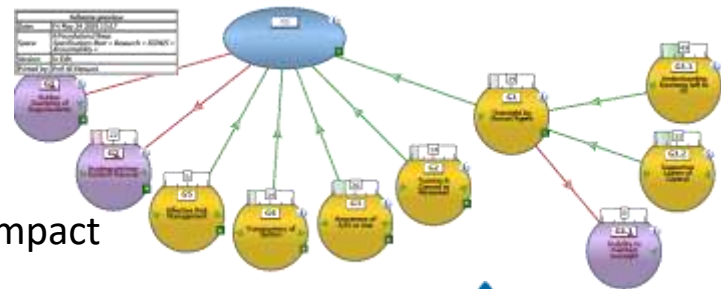
Ethical Foundational Requirements – Transparency Example

ECPAIS Transparency Experts Focus Group: G1 Criteria				Name:	Ali Hessami	Date:	14Nov19
Goal	Requirement	Measurement	Acceptable Evidence				
G1- Suitable & Sufficient Management (This comprises the capability, maturity, and intent of the development organization in having the right motivation and resources, processes, etc, to achieve transparency.	<p>The duty holder shall fulfill the following Transparency Requirement(s):</p> <ul style="list-style-type: none">-Demonstrate that a suitable and sufficient organizational governance framework is in place reflecting capability, maturity, and processes to ensure legal responsibility and ethical accountability.	<p>Two-tier approach to encourage adoption:</p> <ol style="list-style-type: none">1. Top-level finding: "no critical findings in the detailed requirements" / "areas requiring attention for improvement"2. Organizational readiness finding: on 1-5 scale (based on aggregate of satisfying sub level goals) such as: 5- excellent 4- good 3- average 2- poor 1- unacceptable	<p>The following item(s) shall be presented as evidence for conformity against the Transparency Requirement(s):</p> <ul style="list-style-type: none">• Organigramme showing lines of responsibility and accountability including the supply chain• Designated positions for risk management, data protection compliance, legal compliance, stakeholder management, and ethical profile management and coordination across all roles• Minimum assessment requirements comprising a) sector risks, including web-based global operation risks; b) potential harms/adverse impacts from A/IS; c) end-user needs (e.g. privacy); and d) supply chain awareness and compliance with minimum assessment requirements• Implementation of local laws and requirements relevant above minimum assessment requirements• Overall legal compliance (dependent on cross-jurisdictional reach				

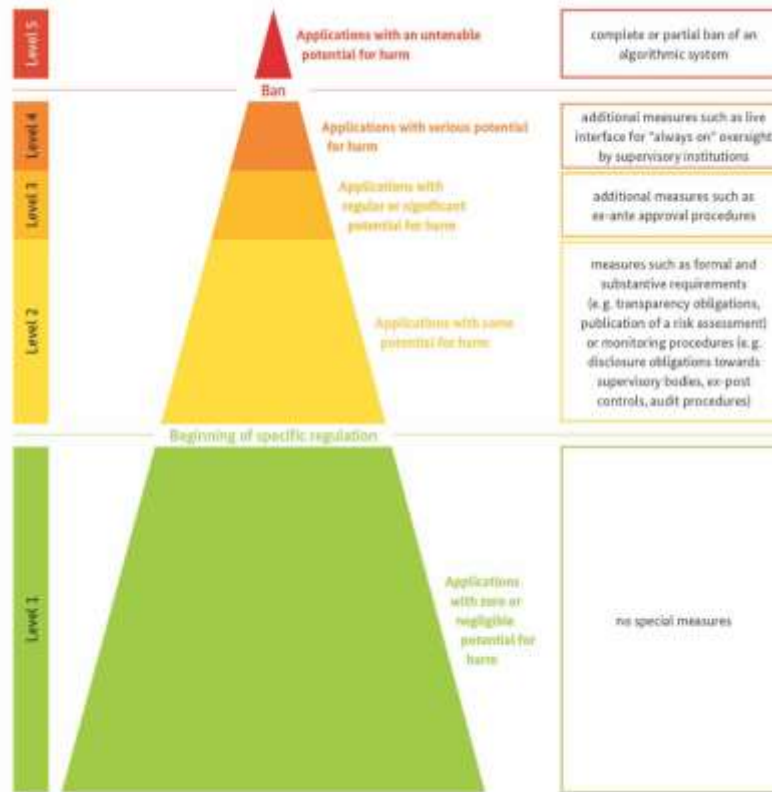
The Ethics Certification Program for Autonomous and Intelligent Systems

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 (LI), Compliant (MI), Critical (HI)*



Affinity to Emerging Regulatory Framework



The Ethics Certification Program for Autonomous and Intelligent Systems

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

IEEE ECPAIS
Certification Criteria for
Autonomous & Intelligent Systems
A/IS

Transparency Certification Requirements

Date: 16th December 2020, 2:00

Author: Dr. Eleanor Nelson, Transparency Expert Group Lead

Reviewer: A.G. Housley, ECPAIS-RC and Process Architect

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

