Ethical, Legal, Societal Aspects

- Artificial vs natural intelligence
- Ethics
- Future of AI
- Social impact of AI
- Practical aspects of today's life
- Ethically aligned design of AI systems

Humans vs Computers

- We talk about computers: Turing machine
- Many things humans do but computers cannot do by themselves
  - Prove or disprove that a grammar is unambiguous
- Yet, many of these things are added to the list of those done by computers
  - Driving vehicles
  - Playing soccer or ping pong, etc

- Many things computers can do better than us, or we can't do
  - Add two one-billion digit numbers
  - Even some tasks when driving a car!
  - Play Chess, Go, Jeopardy, etc

Human vs Robots

- Robots can do a lot of tasks humans do
- ... And better than us
  - Better means faster, more accurate, cheaper, etc.
- Examples:
  - Playing soccer, ping pong, etc.
  - Run Amazon's warehouse
    - https://spectrum.ieee.org/automaton/robotics/industrial-robots/amazon-introduces-two-new-warehouse-robots
  - Agility robots can beat Usain Bolt
    - https://spectrum.ieee.org/automaton/robotics/industrial-robots/agility-robotics-introduces-cassie-a-dynamic-and-talented-robot-delivery-ostrich
- But how can they beat the humans on everything?
- Societal impact:
  - Tons of people's jobs will be taken by AI/Robots

https://lnkd.in/e64Dan5

Can machines really think?

- Church-Turing thesis
  - A function on natural numbers can be computed by an algorithm iff it can be computed by a Turing machine
- Turing Test
  - A computer pretends to be a human
  - Ask questions to computer and reveal its identity (human/computer)

Mathematical view:

- Godel's incompleteness theorem
  - In any axiomatic system F
  - Powerful enough to do arithmetic
  - There are sentences, aka G(F) that cannot be proved within F
  - Put in other words:
    - G(F) can be proved to be true or false, unless supported by external evidence
  - This is the main issue of unsupervised learning!

Can machines really act as humans?

- Computers don’t feel emotions
  - love, lie intentionally, have religious beliefs
  - Well… so far
- The debate is still open

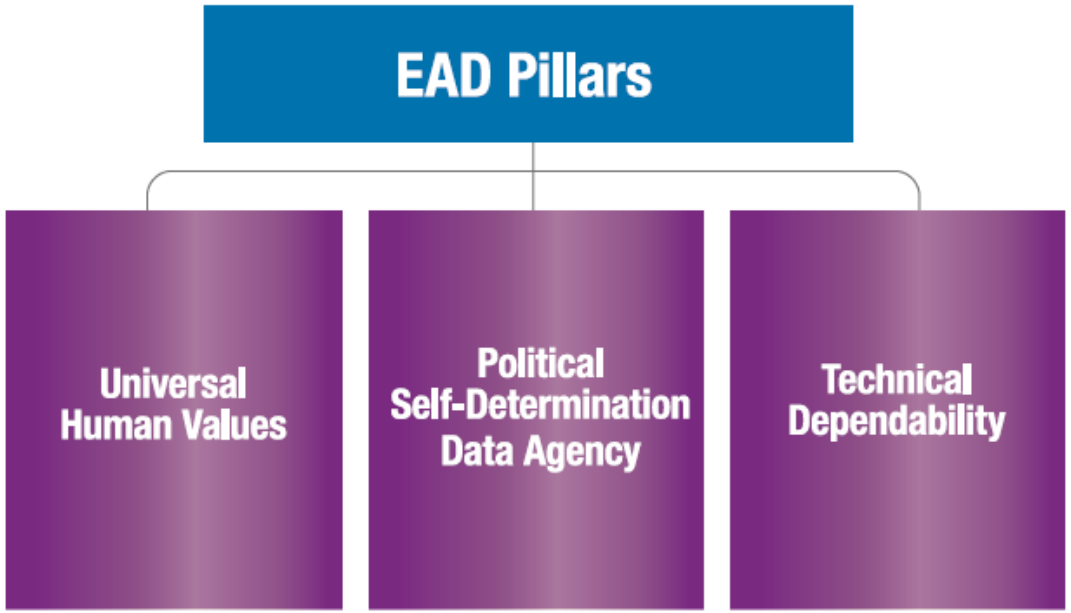
Ethics and Risks – Societal Impact

- Obvious implications of progress in AI
  - People might lose their jobs due to automation
    - Example: Amazon's warehouse
  - People might have too much time for leisure -> laziness
  - People might lose their sense of being unique
  - AI systems might be used toward undesirable ends
  - The use of AI systems might lose accountability
  - People might lose their ability to think
  - The success of AI might imply the end of the human race
    - Too drastic though!
  - Problems of liability
    - If a driverless car gets involved in an accident,
    - whom are we going to blame?
    - what will insurance cover?

Ethically Aligned Design (EAD) by IEEE

Three main pillars of the Ethically Aligned Design Conceptual Framework [2]

- Universal Human Values
  - AI systems should be designed to protect human rights, human values and well being
  - Should safeguard environment and natural sources
  - Should be in the service of people
    - Not benefiting solely smaller groups
- Political Self-Determination and Data Agency
  - Encourage and align to political freedom and democracy
  - Accordance with cultural precepts
  - Grant people have access to and control over data
- Technical Dependability
  - AI should deliver services that can be trusted
  - Trust means reliable, safe and accomplish objectives for which they were designed



General Principles of Ethically Aligned Design

- Human Rights
- Well-being
- Data Agency
- Effectiveness
- Transparency
- Accountability
- Awareness of Misuse
- Competence

Mapping the Pillars to the Principles

		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			■

Chapters in EAD, First Edition

- From Principles to Practice
- General Principles
- Classical Ethics in AI
- Well-being
- Affective Computing
- Personal Data and Individual Agency
- Methods to Guide Ethical Research and Design
- AI fro Sustainable Development
- Embedding Values into Autonomous and Intelligent Systems
- Policy
- Law

Full chapter content available in [2]

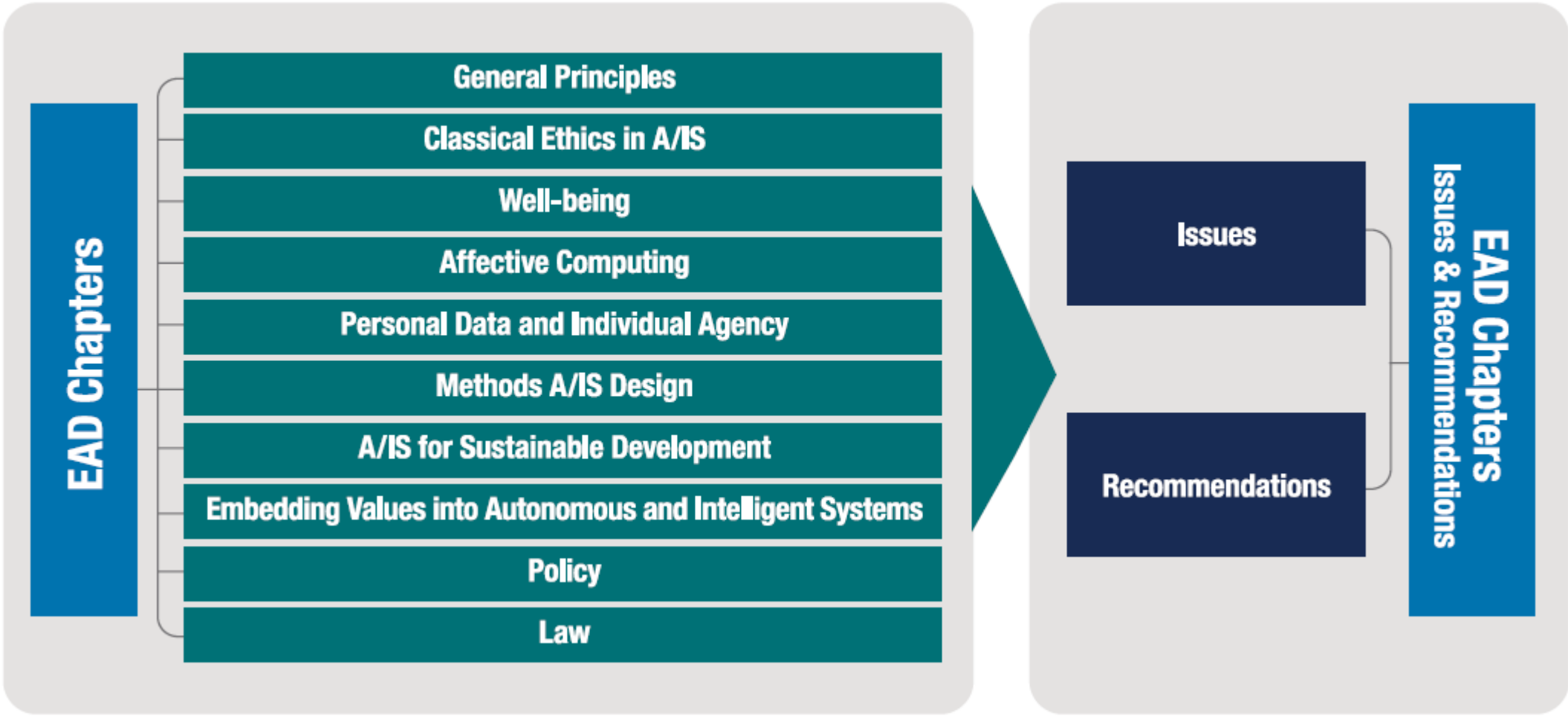
Mapping the Principles to Contents of Chapters

EAD General Principles	EAD Chapters: General Principles	Classical Ethics in A/IS	Well-being	Affective Computing	Data & Individual Agency	Methods A/IS Design	A/IS for Sustainable Dev.	Embedding Values into A/IS	Policy	Law
Human Rights	■	■	■	■	■	■	■	■	■	■
Well-being	■	■	■ ■	■	■		■	■	■	■
Data Agency	■		■	■	■ ■	■	■	■	■	
Effectiveness	■			■		■		■	■	■
Transparency	■			■		■		■	■	■
Accountability	■			■		■	■	■	■	■
Awareness of Misuse	■	■		■		■		■	■	■
Competence	■			■		■		■	■	■

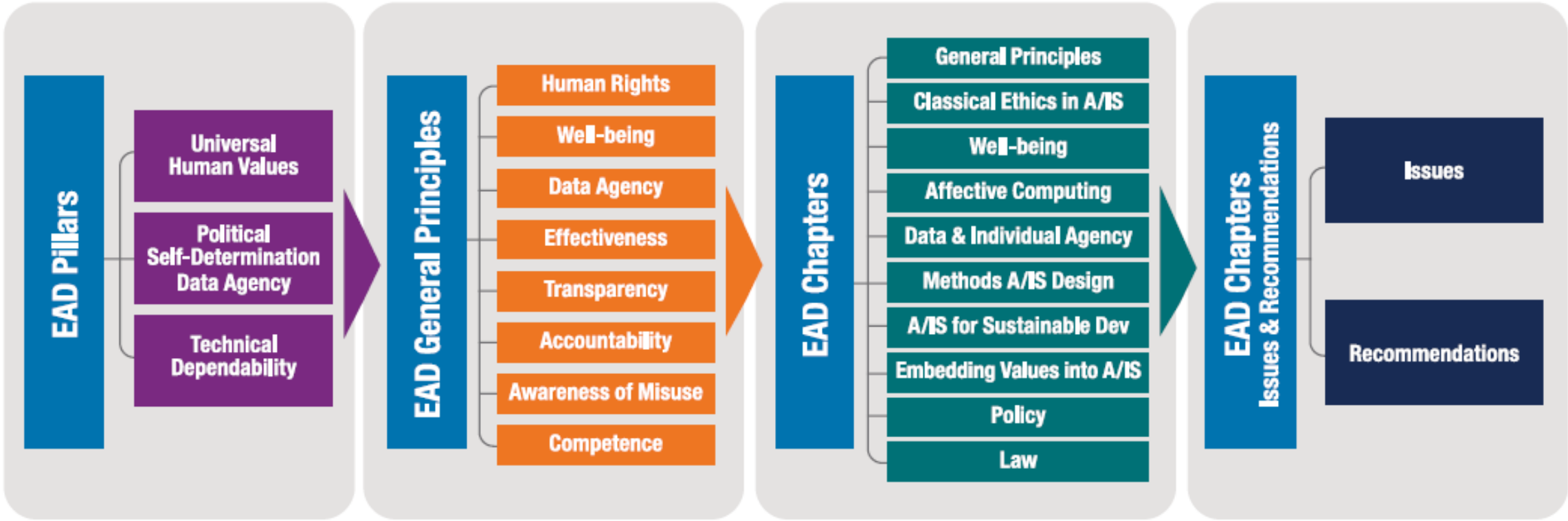
■ Indicates General Principle mapped to Chapter.

■ Indicates primary EAD Chapter providing elaboration on a General Principle.

From Principles to Practice



EAD Conceptual Framework: From Principles to Practice



Practical Aspects of AI – when robots're not ready

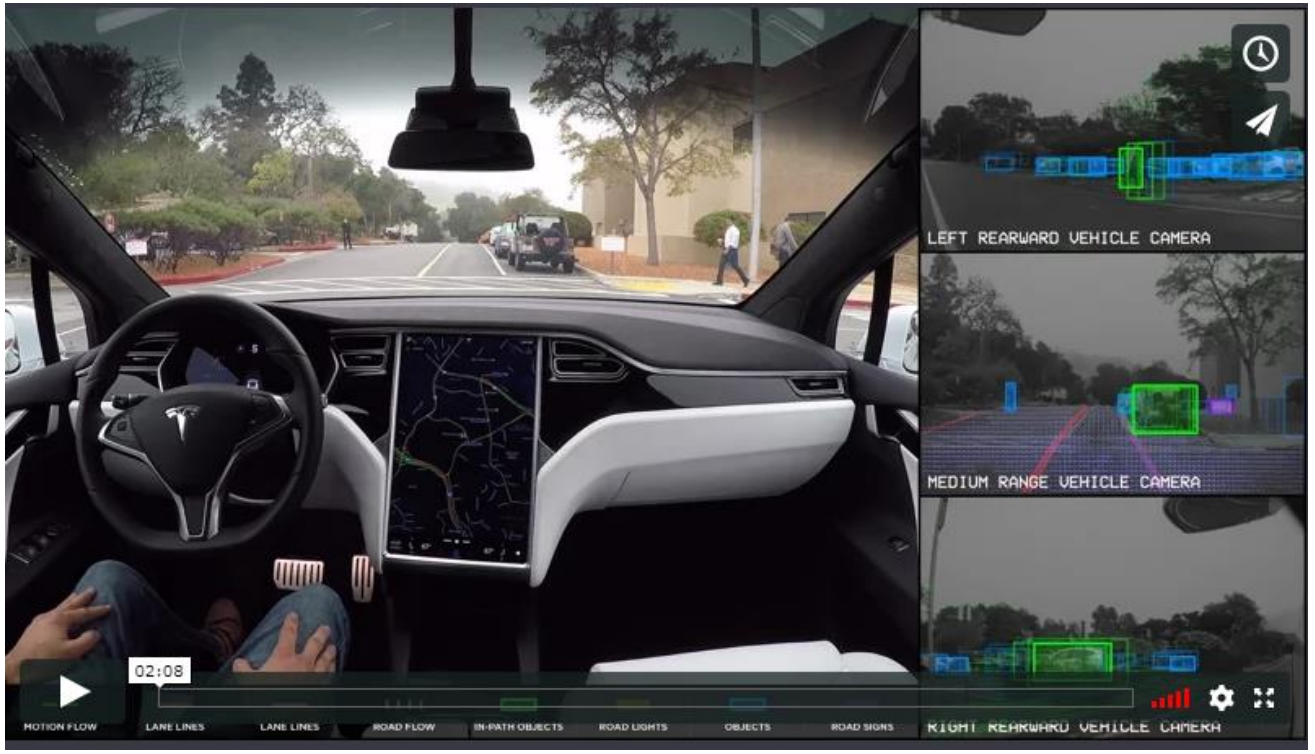
- Practical case: personal view
  - Robot pool cleaner vs manual vacuum
  - Manual vacuum cleaner:
    - Need 50 minutes of work per week (usually done on Sundays):
    - Setting up vacuum system
    - Raking leaves, cleaning bottom, edges, etc.
    - Filter draining, backwashing, etc.
    - The "actual cleaning" takes 10-15 minutes.
  - Robot cleaner:
    - Cleaning bottom (robot) takes 10 minutes.
    - But I need to maintain the robot (charge battery, cleaning, etc.). This add up more time.
    - When I manually clean the pool, I do other things (included in the 50 minutes)
    - Costs of robot $500+ to $1,000s
  - Utility-based decision: not profitable

vs

State of the Art in Industry

Tesla's autopilot

- Advanced sensor coverage
- 360 degree visibility
- Up to 250 meters of range
- 12 ultrasonic sensors for vision
- Forward facing radar
- Enhanced and redundant processing
- Can see through heavy rain, fog, dust and even the car ahead
- Can drive from point to point with no assistance
- Not perfect though
  - Full self-driving in almost all circumstances



The Future of AI

- Improved sensors
  - Cameras, infrared, gyroscopes, GPS, smell, taste, etc.
- Modeling
  - Learning schemes being improved
  - Agent architectures
- Computing
  - Improved hardware, but with limitations
  - Biocomputing, quantum computing
  - The abstract computer (Turing machine) stays the same!
  - NP completeness: Is P = NP?
- Definition of rationality
  - Perfect, calculative, bounded, probabilistic

References

1. Artificial Intelligence: A Modern Approach, 3rd Edition, by R. Norvig. Pearson Hall.
2. Ethically Aligned Design: The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, First Edition, 2019. https://ethicsinaction.ieee.org/