

From Principles to Practice

Ethically Aligned Design Conceptual Framework

Ethically Aligned Design, First Edition (EAD1e) represents more than a comprehensive report, distilling the consensus of its vast community of creators into a set of high-level ethical principles, key issues, and practical recommendations. EAD1e is an in-depth seminal work, a one-of-a-kind treatise, intended not only to inform a broader public but also to inspire its audience and readership of academics, engineers, policy makers, and manufacturers of autonomous and intelligent systems¹ (A/IS) to take action.

This Chapter, “From Principles to Practice”, provides a mapping of the conceptual framework of *Ethically Aligned Design*. It outlines the logic behind “Three Pillars” that form the basis of EAD1e, and it connects the Pillars to high-level “General Principles” which guide all manner of ethical A/IS design. Following this, the content of the Chapters of EAD1e is mapped to the Principles. Finally, examples of EAD1e already in practice are described.

Sections in this Chapter:

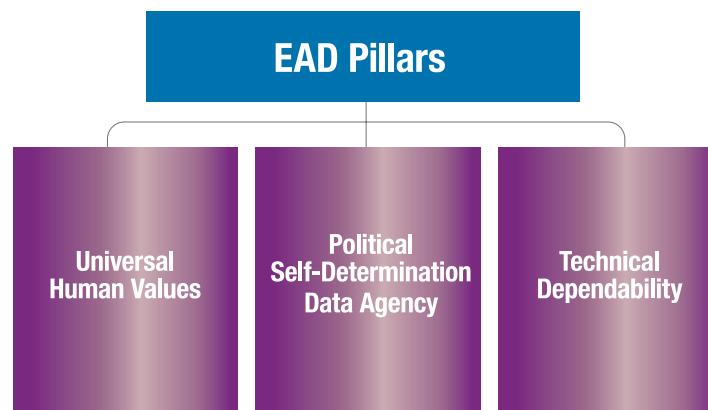
- The Three Pillars of the *Ethically Aligned Design* Conceptual Framework
- The General Principles of *Ethically Aligned Design*
- Mapping the Pillars to the Principles
- Mapping the Principles to the Content of the Chapters
- From Principles to Practice
- *Ethically Aligned Design* in Implementation

From Principles to Practice *Ethically Aligned Design* Conceptual Framework

The Three Pillars of the *Ethically Aligned Design* Conceptual Framework

The Pillars of the *Ethically Aligned Design* Conceptual Framework fall broadly into three areas, reflecting anthropological, political, and technical aspects:

- 1. Universal Human Values:** A/IS can be an enormous force for good in society provided they are designed to respect human rights, align with human values, and holistically increase well-being while empowering as many people as possible. They should also be designed to safeguard our environment and natural resources. These values should guide policy makers as well as engineers, designers, and developers. Advances in A/IS should be in the service of all people, rather than benefiting solely small groups, a single nation, or a corporation.
- 2. Political Self-Determination and Data Agency:** A/IS—if designed and implemented properly—have a great potential to nurture political freedom and democracy, in accordance with the cultural precepts of individual societies, when people have access to and control over the data constituting and representing their identity. These systems can improve government effectiveness and accountability, foster trust, and protect our private sphere, but only when people have agency over their digital identity and their data is provably protected.
- 3. Technical Dependability:** Ultimately, A/IS should deliver services that can be trusted.² This trust means that A/IS will reliably, safely, and actively accomplish the objectives for which they were designed while advancing the human-driven values they were intended to reflect. Technologies should be monitored to ensure that their operation meets predetermined ethical objectives aligning with human values and respecting codified rights. In addition, validation and verification processes, including aspects of explainability, should be developed that could lead to better auditability and to certification³ of A/IS.



The General Principles of *Ethically Aligned Design*

The General Principles of *Ethically Aligned Design* have emerged through the continuous work of dedicated, open communities in a multi-year, creative, consensus-building process. They articulate high-level principles that should apply to all types of autonomous and intelligent systems (A/IS). Created to guide behavior and inform standards and policy making, the General Principles define imperatives for the ethical design, development, deployment, adoption, and decommissioning of autonomous and intelligent systems. The Principles consider the role of A/IS creators, i.e., those who design and manufacture, of operators, i.e., those with expertise specific to use of A/IS, other users, and any other stakeholders or affected parties.

The General Principles⁴ of *Ethically Aligned Design*

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.



From Principles to Practice

Ethically Aligned Design Conceptual Framework

Mapping the Pillars to the Principles

Whereas the Pillars of the *Ethically Aligned Design* Conceptual Framework represent broad anthropological, political, and technical aspects relating to autonomous and intelligent systems, the General Principles provide contextual filters for deeper analysis and pragmatic implementation.

It is also important to recognize that the General Principles do not live in isolation of EAD’s Pillars and vice versa. While the General Principle of “Transparency” may inform the design of a specific autonomous or intelligent system, the A/IS must also account for universal human values, political self-determination, and data agency. Moreover, Transparency goes beyond technical features. It is an important requirement also for the processes of policy and lawmaking. In this way, EAD1e’s Pillars form the holistic ethical grounding upon which the Principles can build, and the latter may apply in various spheres of human activity.

EAD1e Pillars Mapped to General 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			■

■ Indicates General Principle mapped to Pillar.

Mapping the Principles to the Content of the Chapters

The Chapters of *Ethically Aligned Design* provide in-depth subject matter expertise that allows readers to move from the General Principles to more deeply analyze ethical A/IS issues within the context of their specific work.

The mapping or indexing provided in the table below serve as directional starting points since elements of a Principle like “Competence” may resonate in several EAD1e Chapters. In addition, where core subjects are primarily covered by specific Chapters, we have done our best to indicate this via our mapping below.

EAD1e General Principles Mapped to Chapters

		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
EAD General Principles	Human Rights	■	■	■	■	■	■	■	■	■	■
	Well-being	■	■	■ ■ ■	■	■		■	■	■	■
	Data Agency	■		■	■	■ ■ ■	■	■	■	■	
	Effectiveness	■			■		■		■	■	■
	Transparency	■			■		■		■	■	■
	Accountability	■			■		■	■	■	■	■
	Awareness of Misuse	■	■		■		■		■	■	■
	Competence	■			■		■		■	■	■

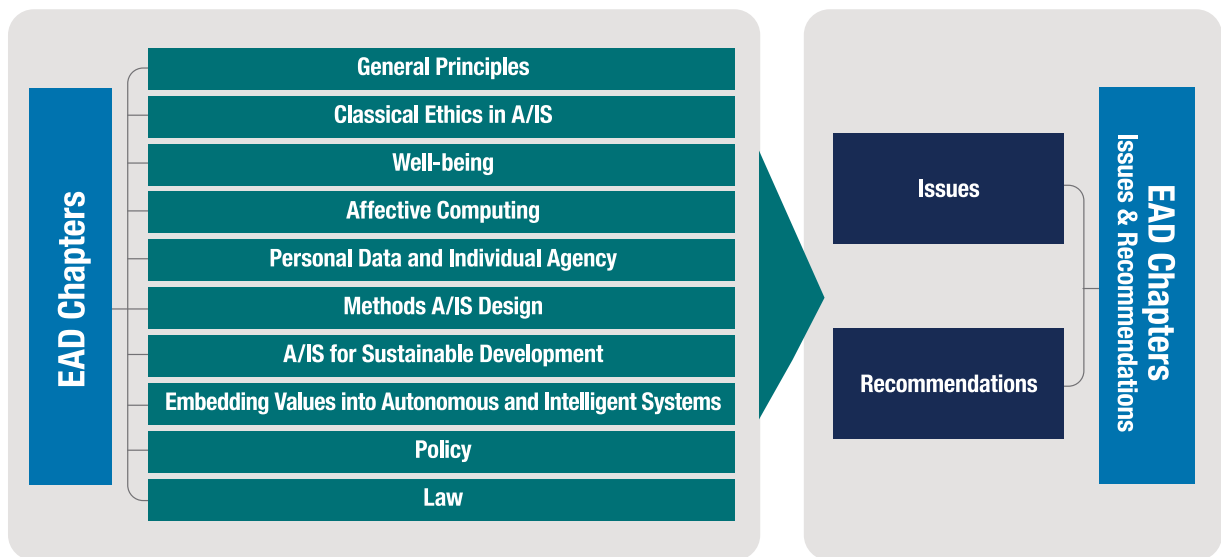
- Indicates General Principle mapped to Chapter.
- Indicates primary EAD Chapter providing elaboration on a General Principle.

From Principles to Practice

Ethically Aligned Design Conceptual Framework

From Principles to Practice

It is at this step of the *Ethically Aligned Design* Conceptual Framework that readers will be able to identify the Principles and Chapters of key relevance to their work. Content provided in EAD1e Chapters is organized by “Issues” identified as the most pressing ethical matters surrounding A/IS design to address today and “Recommendations” on how it should be done. By reviewing these Issues and Recommendations in light of a specific A/IS product, service, or system being designed, readers are provided with a simple form of impact assessment and due diligence process to help put their “Principles into Practice” for themselves. Of course, more fine-tuned customization and adaptation of the content of EAD1e to fit specific sectors or applications are possible and will be pursued in the near future. See below for some implementation examples already happening.



Ethically Aligned Design in Implementation

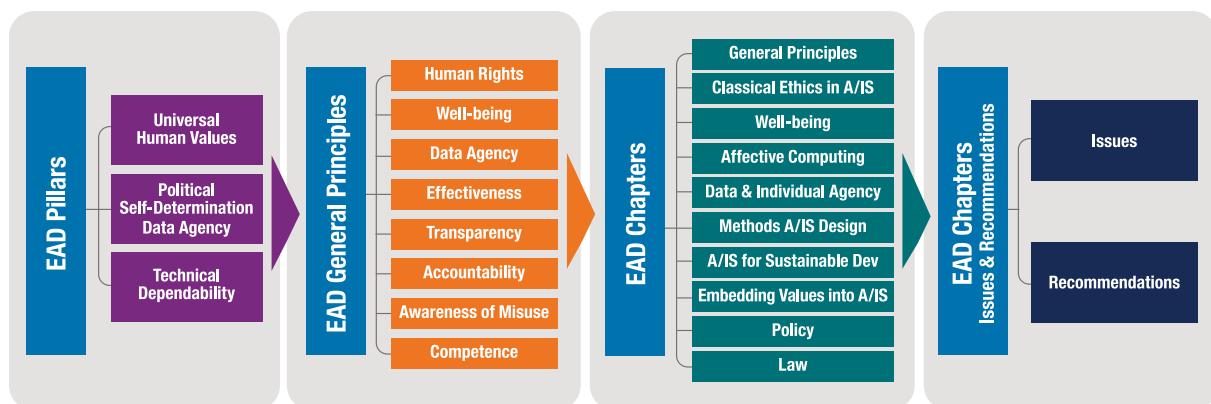
Ethically Aligned Design, First Edition represents the culmination of a three-year process guided bottom-up since 2015 by the rigor and standards of the engineering profession and by a globally open and iterative process involving hundreds of global experts. The analysis of the Principles, Issues, and Recommendations generated as part of an iterative process have already inspired the creation of fourteen IEEE Standardization Projects, a Certification Program, A/IS Ethics Courses, and multiple other action-oriented programs currently in development.

In its earlier manifestations, *Ethically Aligned Design* informed collaborations on A/IS governance with a broad range of governmental and civil society organizations, including the United Nations, the European Commission, the Organization for Economic Cooperation and Development and many national and municipal governments and institutions.⁵ Moreover, the engagement in all of these arenas and with such partners has put the collective knowledge and creativity of The IEEE Global Initiative in the service of global policy-making with tangible and visible results. Beyond inspiring the policy arena, EAD1e and this growing body of work has also been influencing the development of industry-related resources.⁶

It is time to move “From Principles to Practice” in society regarding the governance of emerging autonomous and intelligent systems. The implementation of ethical principles must be validated by dependable applications of A/IS in practice while honoring our desire for political self-determination and data agency. To achieve societal progress, the autonomous and intelligent systems we create must be trustworthy, provable, and accountable and must align to our explicitly formulated human values.

It is our hope that *Ethically Aligned Design* and this conceptual framework will provide action-oriented inspiration for your work as well.

Ethically Aligned Design Conceptual Framework—From Principles to Practice



For information on disclaimers associated with EAD1e, see [How the Document Was Prepared](#).

From Principles to Practice

Ethically Aligned Design Conceptual Framework

Endnotes

¹ We prefer not to use—as far as possible—the vague term “AI” and use instead the term autonomous and intelligent systems (A/IS). This terminology is applied throughout *Ethically Aligned Design, First Edition* to ensure the broadest possible application of ethical considerations in the design of the addressed technologies and systems.

² See also [Draft Ethics Guidelines for Trustworthy AI](#) of The European Commission’s High Level Expert Group on AI.

³ A/IS should be subject to specific certification procedures by competent and qualified agencies with participation or control of public authorities in the same way other technical systems require certification before deployment. The IEEE has launched one of the world’s first programs dedicated to creating A/IS certification processes. [The Ethics Certification Program for Autonomous and Intelligent Systems](#) (ECPAIS) offers processes by which organizations can seek certified A/IS products, systems, and services. It is being developed through an extensive and open public-private collaboration.

⁴ For their overall framing, see the “General Principles” Chapter.

⁵ As an example, the recently published report [Draft Ethics Guidelines for Trustworthy AI](#) of The European Commission’s High Level Expert Group on AI explicitly mentions EAD as a major source of their inspiration. EAD has also been guiding policy creation for efforts of the United Nations and the Organization for Economic Cooperation and Development.

⁶ [Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers and Developers](#)