ISSN: 2756-2638
Transactions on Engineering and Computing Sciences; Vol. 14 No. 04 (2026) (133-148)
https://doi.org/10.14738/tmlai.1404.12061

Value Sensitive Design

Oleg Namicheishvili[1], Guram Javakhadze[2], & Vera Prangishvili[3]

1. Institute of Quantum Physics Engineering Technologies, Georgian Technical University, Tbilisi, Georgia
2. Joint Stock Company "Medinserv", Tbilisi, Georgia
3. Communications Commission, Tbilisi, Georgia

Abstract: This paper presents the concept of Value Sensitive Design (VSD) as a framework for embedding and preserving the values that society seeks to uphold in innovative technologies. Value Sensitive Design can be formally represented through a Values Hierarchy Matrix, which may be applied using either a top-down or a bottom-up approach. The visual representation of values enables stakeholders to discuss, evaluate, and reconcile these values in a constructive manner. Furthermore, it facilitates the critical deconstruction and examination of operational design criteria by raising fundamental questions: Does a given design genuinely reflect the values that we consider important, or, conversely, do the established design criteria enable the realization of the desired values?

Keywords: Information Systems, Value Sensitive Design, Interactionism, Substantivism, Sustainability, Values Hierarchy Matrix.

INTRODUCTION

This paper focuses on the significance of Value Sensitive Design (VSD) for the development of responsible innovation. VSD seeks to provide a practical framework for translating moral and ethical values into concrete technical parameters that can be incorporated into technological systems and innovative products.

Historically, Value Sensitive Design emerged as a distinct discipline within computer science. However, the underlying concept extends far beyond computing and has broad applicability across technological development. When computers first appeared in the mid-twentieth century, scholarly attention was directed primarily toward the technology itself. Computers were rightly regarded as general-purpose technologies capable of addressing a wide range of problems across numerous disciplines. Nevertheless, relatively little attention was devoted to the social context of computing or to the needs and experiences of its users.

During the second stage of computing development, spanning the 1970s and 1980s, the social and behavioral sciences assumed increasing importance in information technology research. This shift was reflected in the emergence of human–computer interaction (HCI), participatory design, and social informatics.

At that time, the emphasis on social context and patterns of user behavior was motivated primarily by practical concerns. Researchers sought to identify potential obstacles to the successful implementation of information systems, reduce system failures and accidents, and prevent unsuccessful investments. Ultimately, however, this line of research evolved into the systematic study of usability, practicality, and user acceptance.

A third stage of development emerged around the turn of the twenty-first century, when it became increasingly evident that the success of information technologies depends not merely on usability but also on their ability to accommodate a broad spectrum of human values. Individuals—whether acting as employers, consumers, citizens, or patients—possess moral values, ethical preferences, and normative ideals. Every society continuously debates issues such as equality, property rights, privacy, sustainability, autonomy, responsibility, and many other fundamental values. Consequently, computer networks and information systems should, whenever feasible and appropriate, be designed to reflect and support these values.

Over the past decade, values have ceased to be regarded merely as constraints on technological implementation. Instead, they have become fundamental design objectives and proactive drivers of technological innovation. An illustrative example is the establishment of the Center for Information Technology Research in the Interest of Society (CITRIS) in California in 2001. This development suggests that information technology has entered a fourth logical stage, in which users' needs and values are recognized as central design considerations. Information technology is increasingly viewed as serving human beings as moral agents, supporting their individual ethical commitments and broader social aspirations.

Accordingly, perspective has shifted from viewing technology solely through its technical characteristics to understanding it within its broader social context. Likewise, moral values are no longer regarded merely as constraints that technologies must satisfy but as guiding principles for creating technologies that serve human needs and societal goals. This intellectual development is clearly reflected in the widespread adoption of the concept of Value Sensitive Design (VSD) during the past decade.

The central premise of VSD is that moral values can be expressed explicitly in engineering terms. Ethical principles can therefore be translated into concrete design requirements concerning, among other aspects, the sustainability, safety, and privacy of the technologies and artifacts that we design and develop.

Related Intellectual Traditions

The VSD framework synthesizes several influential strands of research and scholarship.

Do Artifacts Have Political Significance?

One of the foundational contributions to this line of inquiry was Langdon Winner's influential 1980 essay, *Do Artifacts Have Politics?* [1]. Winner argued that values can become embedded not only in social institutions but also in physical artifacts and technological systems, thereby profoundly influencing human behavior. His compelling examples demonstrated how values, political assumptions, and power relations incorporated into technological designs can shape and constrain human action, profoundly influencing subsequent discussions on the ethics of technology design. One of Winner's most widely cited examples concerns the low overpasses constructed on parkways leading to Jones Beach in New York by the urban planner and architect Robert Moses. Moses intentionally designed these overpasses to be too low for buses while allowing passenger cars to pass freely. As a consequence, middle-class citizens who owned private automobiles could easily reach Jones Beach, whereas residents of poorer neighborhoods, who relied primarily on public transportation, were effectively excluded. In this interpretation, the overpasses functioned as a form of materialized social discrimination, separating wealthier citizens from poorer communities.

Although historians continue to debate the factual accuracy of this example, it vividly illustrates how values can become embedded in material artifacts and subtly shape people's lives.

Science and Technology Studies (STS)

Throughout the 1980s, numerous studies in the philosophy and sociology of technology contributed to the emergence of the interdisciplinary field known as Science and Technology Studies (STS). This body of research demonstrated that sociopolitical biases and prejudices—particularly those related to race, gender, and socioeconomic status—can become embedded within technical artifacts, infrastructures, and technological systems.

Among the leading contributors to this field are Geoffrey C. Bowker, Susan Leigh Star, and Lucy A. Suchman, whose work has significantly advanced our understanding of the social dimensions of technology.

Engineering Perspectives

At approximately the same time, several specialized fields within design and engineering also began to adopt the Value Sensitive Design (VSD) approach. This development can be illustrated by the following two examples.

The first concerns privacy-enhancing technologies. During the 1980s, researchers investigating privacy began exploring ways to design information systems and software applications that would increase the likelihood of users complying with privacy norms and regulatory requirements. Rather than relying solely on users' voluntary adherence to privacy policies, the systems themselves were designed so that compliance with privacy requirements would become the natural and default mode of operation.

A second example comes from architecture and urban planning during the 1980s. In these fields, proactive design strategies—aimed at anticipating and preventing potential problems—were increasingly incorporated into architectural design and urban planning to enhance public safety and reduce opportunities for crime. Research demonstrated that factors such as lighting, architectural diversity, spacing between buildings, and unobstructed sightlines, among others, significantly influence crime rates. Consequently, these factors were systematically incorporated into the design parameters of buildings and urban neighborhoods.

DEFINING THE VALUE SENSITIVE DESIGN METHOD

The clearest and most systematic formulation of the Value Sensitive Design (VSD) concept emerged during the 1970s and 1980s within the field of computer science at Stanford University, where its development was strongly influenced by the work of Terry Allen Winograd (b. 1946).

Value Sensitive Design is an approach to the development of systems and software that was established during the final decade of the twentieth century by Batya Friedman and David G. Hendry of the University of Washington Information School [2]. Drawing upon ideas developed within the Human–Computer Interaction (HCI) research community, they sought to direct scholarly and public attention toward the social and moral dimensions of technological design. The VSD approach places primary emphasis on incorporating a broad spectrum of human and moral values into the design of information technologies.

Although Value Sensitive Design is not tied to any single normative or ethical framework, Batya Friedman emphasizes that it is fundamentally concerned with values related to human well-being, human dignity, justice, social welfare, and human rights. VSD seeks to bridge the gap between those who design technological systems and interfaces and those who understand the values and interests of the stakeholders affected by these systems. As Friedman argues, Value Sensitive Design ultimately requires an expansion of the criteria and objectives used to evaluate the quality of technological systems so that they explicitly include requirements that promote and protect fundamental human values.

For example, at Delft University of Technology, Value Sensitive Design is regarded as a practical means of integrating ethical considerations and moral values into the processes of technological design, research, and innovation.

The principal methodological framework employed in VSD initiatives is an integrative and iterative three-part methodology consisting of conceptual, empirical, and technical investigations. These three forms of inquiry continuously inform one another through an iterative process, with insights from each component contributing to the refinement and advancement of the others. An early and influential presentation of this methodology can be found in the work of Batya Friedman, Peter H. Kahn Jr., and Alan Borning [3].

Value Sensitive Design possesses several characteristics that closely align with the principles of Responsible Innovation. The values and ethical concerns of all relevant stakeholders should be identified and articulated while they can still influence the design process. These values must be formulated in a manner that enables their effective incorporation into the project, and both technological artifacts and design outcomes should be evaluated according to the values they support and the ethical issues they raise.

It is important to recognize that although Value Sensitive Design originated within the fields of information technology and computer science, its applicability extends far beyond these disciplines. Today, VSD provides a valuable framework for virtually all forms of technological innovation, including the design, development, deployment, and implementation of emerging technologies and technological artifacts.

APPLYING VALUE SENSITIVE DESIGN IN PRACTICE

We have seen why the concept of Value Sensitive Design (VSD) is of fundamental importance. Contemporary socio-technical systems present a number of significant challenges. On the one hand, all individuals, as well as society as a whole, uphold values that they consider important. These include such fundamental values as safety, sustainability, justice, privacy, human well-being, and many others. Traditionally, these values were protected and promoted through human behavior, legal frameworks, and public policy.

Today, however, we live in a technological world in which technology increasingly shapes human behavior. One need only consider how profoundly the Internet, computers, and smartphones have transformed everyday life compared with previous generations. Consequently, humanity faces a critical challenge: we must ensure that these technologies embody and promote the values that we consider essential.

This requires translating the world of values and ethical ideals into the world of technology and material artifacts—a task that is far from straightforward, since these two domains have historically been treated as largely separate. This naturally raises a fundamental question: How can values be embodied in technological design?

Do Technologies Embody Values?

Three principal philosophical perspectives offer different answers: instrumentalism, substantivism, and interactionism.

Instrumentalism

Instrumentalism maintains that technology is value-neutral. Technology is viewed merely as a tool whose moral significance depends entirely on the way it is used. Whether a particular technology serves beneficial or harmful purposes is therefore determined not by the technology itself but by human action.

This instrumentalist perspective is succinctly captured in the well-known slogan of the American National Rifle Association (NRA):

"Guns don't kill people; people kill people."

According to this view, moral responsibility rests with human agents rather than with the technologies they employ.

Substantivism

Substantivism adopts the opposite position. It argues that technologies themselves embody particular values and exert influences that are largely independent of human intentions. From this perspective, technology is not merely an instrument but an autonomous carrier of value-laden assumptions.

For example, some substantivist thinkers argue that modern technology inherently promotes efficiency—understood as maximizing outputs relative to invested resources—and that it inevitably contributes to environmental degradation, diminishes human authenticity, or reduces meaningful forms of human interaction.

The principal limitation of this perspective is that it underestimates the capacity of individuals and societies to shape both the development and the use of technology. More broadly, substantivism may be understood as a philosophical doctrine that attributes to entities an essential and immutable set of defining properties. It assumes that objects possess an underlying reality or intrinsic nature that cannot be directly observed but nevertheless determines their fundamental character. Consequently, what is regarded as most significant is this hidden essence rather than the observable characteristics of the object. Substantivist thinking often aligns with common-sense intuitions about the world by treating phenomena as possessing fixed, inherent qualities. This philosophical position was notably advanced by Nikolai Lossky, one of the leading representatives of Russian religious philosophy and a founding figure of the school of intuitionism, as discussed in his work [4].

Interactionism

Interactionism holds that values are created, negotiated, and reinforced through the interaction between human beings and technology. Values emerge not only from the ways technologies are used but also from the ways they are designed. In the present discussion, particular emphasis is placed on the design dimension of interactionism. Before doing so, however, it is useful to provide a brief overview of the interactionist perspective itself.

Interactionism is a contemporary theoretical approach that examines individual development and human activity within the context of social interaction. It does not constitute a single unified theory or scientific school. Rather, several of its principles overlap with other theoretical traditions, particularly psychoanalysis and behaviorism. Consequently, this branch of psychological inquiry is often referred to as the interactionist approach. The principal representatives of interactionism—including Herbert George Blumer, George Alexander Kelly, and Erving Goffman—made their most influential contributions within the field of social psychology. Nevertheless, the interactionist perspective has also had a profound impact on general, developmental, and educational psychology.

Interactionism is founded upon the theory of social interaction developed by the American philosopher, psychologist, sociologist, and leading representative of the Chicago School of Sociology, George Herbert Mead. According to Mead, personality develops through an individual's interactions with members of a social group while participating in shared activities. Mead emphasized that society and the social group, rather than the individual alone, play the decisive role in this developmental process. Through interaction with others, individuals acquire the capacity for abstract thought and, most importantly, develop self-awareness by recognizing themselves as distinct persons in relation to others. In this view, an individual's inner world emerges through collective social activity. At the same time, Mead's theory has been criticized for offering a somewhat one-sided interpretation of these processes, since it tends to conceive society primarily as the collective interaction of individuals performing various social roles.

Mead's ideas were further developed by the American psychologist Robert Richardson Sears, a prominent scholar in social and child psychology and professor at several leading universities. In his influential work, *Frustration and Aggression* (1939), co-authored with his colleagues [5], Sears proposed the dyadic system (from the Greek *dyo*, meaning "two") as the fundamental unit of psychological analysis. A dyadic system describes the interactions between two or more individuals. According to Sears, every individual's actions are inherently oriented toward and influenced by another person. Consequently, there are no permanently fixed or immutable personality traits—such as aggressiveness or benevolence—because an individual's behavior always depends upon the characteristics and responses of the other member of the dyadic relationship.

Within interactionism, one influential perspective is the dramaturgical approach to social life. Its proponents view society as a form of drama in which each individual performs a socially assigned "role." From this perspective, social interaction resembles a theatrical performance, with individuals continually interpreting and enacting roles within different social situations. Accordingly, social drama—the enactment of life situations through role performance—has been employed both as a method of psychotherapy and as an important pedagogical tool for personal development.

The leading interactionist scholars sought to identify what is uniquely human in social life, to explain the socio-psychological mechanisms underlying personality formation, and to emphasize the active, creative capacities of the individual in shaping both personal identity and social reality.

Which Values Should Be Incorporated into Technological Design?

The first point to recognize is that engineering design may involve a wide range of values. These values may be derived from numerous sources, including the project specification (through its statement of objectives and rationale), designers and their professional communities, users and other stakeholders, legislation and public policy, technical codes and standards, professional codes of ethics, and broader moral considerations.

However, merely identifying these sources does not answer the more fundamental question of which values ought to be incorporated into a design. This is a normative question—that is, a question concerning what ought to be done rather than what merely exists.

The issue is further complicated by what philosophers refer to as value pluralism. Because societies embrace a diversity of values, individuals may reasonably disagree about which values are the most important. Consequently, value pluralism makes the selection of design values considerably more challenging. Nevertheless, it does not render such selection impossible, for several reasons:

- first, despite the existence of value pluralism, consensus can often be reached regarding at least some of the values that should be reflected in technological design;

- second, value pluralism frequently implies disagreement about the relative priority of different values rather than disagreement about their overall importance. While individuals may differ over which values deserve precedence, they can often agree that a broad range of values should be taken into consideration. For example, people may disagree about whether safety or sustainability should receive greater emphasis in the design of a new technology. Nevertheless, most would agree that both safety and sustainability should be incorporated into the design of, for instance, a new automobile;
- third, it is often possible to design technologies in ways that accommodate the diverse values and interests of different stakeholder groups. Rather than privileging a single value system, technological design can frequently balance and reconcile multiple legitimate perspectives, thereby producing solutions that are responsive to a broader range of societal concerns.

REPRESENTING MORAL VALUES INTO TECHNICAL DESIGN REQUIREMENTS

Instrumental and Intrinsic Values

When discussing the most fundamental values, philosophers commonly distinguish between instrumental and intrinsic values.

Instrumental values are those that are valuable because they serve as means to achieve other desirable ends. Money, for example, is generally regarded as an instrumental value because it enables people to attain many other important goals and aspirations in life.

Intrinsic values, by contrast, are valuable in and of themselves. They are not pursued merely as means to some further objective but are considered worthy of realization for their own sake. Typical intrinsic values include human well-being, justice, beauty, honesty, and truth.

The central question, therefore, is how such abstract moral values can be translated into concrete and operational design requirements. To address this challenge, the concept of a Values Hierarchy is employed.

The Values Hierarchy consists of three interconnected levels:

1. Values;
2. Norms;
3. Design requirements.

This hierarchical structure provides a systematic means of transforming abstract ethical principles into concrete technical specifications that can guide engineering design.

The Value Sensitive Design Matrix

The Value Sensitive Design Matrix, also referred to as the Values Hierarchy Matrix, is an analytical framework used within the VSD methodology to operationalize values, norms, and design criteria. It supports both top-down and bottom-up reasoning.

In a top-down process, designers begin by identifying the fundamental values that a technology should embody and progressively translate these values into norms and operational design requirements. Conversely, the bottom-up approach starts with existing technical requirements or design features and examines the values that these requirements promote or preserve.

The matrix enables designers and other stakeholders to identify core values, transform them into measurable engineering requirements, and assess whether existing design solutions genuinely support the intended ethical objectives. In this way, it facilitates ethically informed decision-making through the systematic evaluation of stakeholders, values, and the potential societal impacts of technological design.

The Biofuel Example

Two terms are commonly used in the literature to refer to the same concept: biofuel and biological fuel. In this paper, preference is given to the term biofuel, which is widely adopted in contemporary scientific literature (see, for example, [6]). The alternative term, biological fuel, appeared more frequently in earlier publications and can be found in classical works such as [7].

A practical illustration of the Values Hierarchy is provided by the case of biofuels. Biofuels are produced from recently living or currently living biological materials (biomass). They were introduced as an alternative to fossil fuels in response to concerns regarding the depletion of fossil fuel reserves and the need to reduce greenhouse gas emissions.

However, biofuels have also been the subject of substantial criticism because of their potential environmental impacts and their effects on food production, food quality, and food prices. In response to these concerns, organizations such as the Nuffield College Council on Bioethics developed a set of ethical principles that any ethically acceptable biofuel should satisfy.

Figure 1 presents an attempt to organize these ethical considerations within the framework of a Values Hierarchy, illustrating how abstract moral values can be systematized into norms and ultimately into concrete design requirements.

The overarching value is sustainability, which serves as the fundamental guiding value for the development of biofuels. This overarching value is further divided into three more specific values that are essential to sustainability: intergenerational justice, environmental stewardship, and intragenerational justice.

Each of these values is associated with a set of corresponding norms. Consider, for example, the value of intergenerational justice. This value is linked to three norms: ensuring a reliable fuel supply, reducing greenhouse gas emissions, and preventing the exacerbation of other environmental problems.

			Sustainable development			
	Intergenerational justice		Care for the natural environment		Intragenerational justice	
Ensuring secure access to fuel	Reduction of greenhouse gas emissions	Preventing the exacerbation of other environmental problems	Biodiversity conservation	Avoidance of additional increases in food prices	Providing opportunities for developing countries	Ensuring fair remuneration
Energy-efficient fuel	High energy efficiency	No increased use of fertilizers or pesticides	Cultivation should not have an adverse impact on biodiversity	Biofuels should be non-edible	May be produced in developing countries	Flexible use of intellectual property licensing agreements
Renewable Fuel	No additional greenhouse gas emissions are generated during cultivation, production, or transportation.	No excessive use of water or other resources		Should not contribute to competition for agricultural land or other production resources	May be produced on a small scale with limited investment	Working conditions for farmers should not deteriorate.
Reliable supply chain		No increase in air pollution				
Competitive pricing						

Figure 1: Value Sensitive Design Matrix for Biofuels

Each norm is, in turn, translated into a set of concrete design requirements. For example, the norm that fuel should remain accessible implies that it must be energy-efficient, renewable, reliably supplied, and competitively priced (that is, competitively priced relative to comparable products on the market). Another example is the norm that biofuels should not contribute to increases in food prices. This requirement is translated into the design specifications that biofuels should be produced from non-edible feedstocks and should not contribute to competition for agricultural land or other production resources.

At present, no biofuel satisfies all of these requirements simultaneously. Most existing biofuels belong to the first or second generation, meaning that they are either derived from edible feedstocks or compete with food crops for agricultural land and other resources. Nevertheless, ongoing research is focused on developing third-generation biofuels, which are widely expected to overcome these limitations and satisfy the full range of sustainability requirements.

A COMPLICATED PROCESS

The representation of values into design requirements is a complicated process involving multiple interconnected stages and considerations, making it difficult to understand and manage effectively.

Characteristics of Complicated Processes

1. **Multiple stages.** Complicated processes typically consist of numerous stages that must be carried out in a specific sequence. If the process is not clearly defined, misunderstandings and implementation problems are likely to arise.
2. **Interdependence.** The components of a complicated process are often interdependent, meaning that changes in one part of the process may affect other parts, thereby increasing the overall complexity of implementation.
3. **Diversity of inputs.** Such processes frequently require a wide variety of input data and information, adding to their complexity and making them more difficult to manage.

Examples of Complicated Processes

- **Project management.** Managing a large-scale project typically requires coordinating numerous tasks, deadlines, and resources, making project management a complicated process that depends heavily on careful planning and effective communication.
- **Software development.** Developing software can be highly complicated because it often involves multiple programming languages, software frameworks, and integration with numerous external systems.
- **Compliance with regulatory requirements.** Ensuring compliance with regulations in fields such as healthcare or finance constitutes a complicated process owing to the extensive documentation required and the need to satisfy legal and regulatory standards.

Managing Complicated Processes

- **Flowcharts.** Flowcharts help visualize complicated processes, making them easier to understand while also identifying potential bottlenecks and inefficiencies.
- **Process decomposition.** Breaking a complicated process into smaller, manageable components facilitates implementation and enables more effective monitoring and adjustment whenever necessary.
- **Iterative learning.** The *try-learn-adapt* approach can be particularly effective for managing complicated processes, as it enables teams to refine their methods continuously on the basis of feedback and practical experience.

Note: Understanding the nature of complicated processes and applying appropriate management strategies can substantially improve efficiency while reducing errors during implementation. Whether the task involves project management, software development, or regulatory compliance, success depends on careful attention to detail and a thorough understanding of every stage of the process.

We now return to the particular case translating values into design requirements, highlighting several aspects that illustrate the process in a broader context.

- Translating values into design requirements—particularly new or emerging values—may be a lengthy and labor-intensive undertaking.
- The translation process may require specialized knowledge that extends beyond engineering. For example, it may draw upon ethology, a branch of biology (more specifically, zoology) that studies animal behavior under natural conditions, with particular emphasis on genetically determined patterns of behavior. Although insights from ethology may provide useful perspectives for understanding human behavior, they must be applied with caution, taking into account the social nature of human beings and avoiding reductionist explanations that treat human behavior solely as a biological phenomenon.

  In the case of environmental values, expertise in ecology or the environmental sciences may be required. For values such as privacy and trust, philosophical analysis can help clarify their meaning and translate them into more concrete norms. Even values that are more familiar to engineers, such as safety and usability, may require specialized expertise, as demonstrated by the emergence of disciplines such as safety science, safety engineering, and ergonomics.
- The translation process is often carried out, at least in part, outside the specific design process. In many cases, the final translation of relatively general norms into concrete design requirements takes place during the design process itself. Even then, however, engineers frequently rely on established and widely accepted specifications. Besides legislation, the principal sources of such specifications are technical codes and standards developed by engineers serving on standardization committees. These documents typically define design requirements or provide design guidelines based on broadly accepted considerations of safety, compatibility, and other shared engineering concerns.
- The translation of values into design requirements is itself value-laden. There is rarely a single correct way to perform this translation. Different disciplines—and even different subdisciplines—may offer alternative interpretations of the same value. Moreover, the translation may be shaped by the opportunities afforded by contemporary technologies or by the trade-offs that must be made among competing values.
- The translation of values into design requirements is highly context-dependent, even when the values themselves are expressed in general terms.
- It is well recognized that both values and design requirements possess a hierarchical structure. For example, a general value may first be translated by experts into a set of norms, after which governments or regulatory authorities may further translate those norms into highly specific regulatory or design requirements.

Values constitute an integral part of the development of any technology. Determining which values should be incorporated, which should be excluded, and the extent to which each should be reflected is rarely straightforward. Rather than ignoring or dismissing conflicts among values, designers should adopt carefully considered strategies for identifying and resolving them. Both technical and institutional innovations may be developed to prevent value conflicts or to provide effective means of addressing them when they arise.

It is possible to establish a balance between competing values by first defining minimum acceptable thresholds and then optimizing the design to achieve the best possible balance among them. This approach follows a satisficing strategy—one that combines the notion of *satisfaction* with that of *sufficiency*. Under this strategy, each value must first be satisfied to an adequate level by meeting its prescribed threshold. Beyond these threshold values, it is then necessary to determine the extent to which each value can be further realized. In other words, resolving a conflict between safety and sustainability requires, first, establishing a minimum acceptable (or sufficient) level for each value and, second, optimizing the balance between them once these threshold levels have been achieved.

Example from Practice: Autonomous Weapons

What are autonomous weapons? This is, in fact, a rather complex question, partly because the concept of weapon autonomy is itself highly complex and partly because it remains the subject of considerable debate. The simplest way to define an autonomous weapon is to say that it is capable of performing certain tasks without the direct involvement of a human operator. Once programmed, it does not require further commands or instructions from that operator. In this context, an *operator* is understood, in the conventional sense, as a highly trained individual responsible for controlling the operation of a complex technical system.

Such a definition, however, has limited explanatory value because automatic weapons have existed for decades. Indeed, automatic weapon systems are likewise capable of operating without human intervention. For example, missile defense systems were automated because speed is critical when intercepting incoming enemy missiles. Once programmed, these systems are capable of detecting, tracking, and intercepting a missile without any further involvement of the operator.

Perhaps the central question in both the academic literature and policy debates surrounding new weapons systems is whether autonomy should be regarded as conceptually distinct from automation. Although many scholars argue that it should, no consensus has yet emerged on this issue.

One argument frequently advanced in these debates is that autonomous weapons are capable of making independent decisions regarding target selection. The difficulty, however, is that when people claim that autonomous weapons "make decisions," it often remains unclear what *decision-making* actually means in this context. Clarifying this concept is one of the principal objectives of current research and strategic policy analysis, which seek to illuminate this fundamental issue.

If machines were truly capable of making decisions in the full sense of the term, they would possess the ability to apply their own intellectual or normative criteria in evaluating situations. They could be deployed in operational environments and independently apply the principles governing the use of armed force. Although such capabilities do not yet exist, they no longer belong solely to the realm of the distant future and already raise important ethical and policy concerns. We prefer to regard autonomy as a more advanced form of automation. Under this interpretation, the machine cannot make decisions in the sense of relying on its own independent judgment or evaluative criteria. Nevertheless, it differs from a purely automated system. For example, a cruise missile is an automated system: it can be programmed with GPS coordinates, after which it independently navigates to and locates its designated target.

Note: Autonomous systems, by contrast, represent a significantly more sophisticated form of technology. For example, an autonomous system may be deployed and operated in a highly complex and hostile environment, where it is capable of independently navigating toward its objective without continuous guidance from an operator supplying GPS coordinates.

In very simple—but also highly accurate—terms, the principal justification for such systems can be summarized in a single phrase: military advantage. On the other hand, it has also been argued that autonomous weapon systems may offer ethical and legal advantages. In particular, a number of researchers in robotics contend that increasing the autonomy of military systems could reduce the incidence of war crimes and, consequently, violations of international law. If this claim proves to be correct, such systems would undoubtedly be highly desirable. After all, few would object to technologies that have the potential to reduce the occurrence of war crimes. Nevertheless, this is a very significant *if*, and it is accompanied by substantial risks.

It should also be emphasized that the development of new weapons and weapons systems is governed by established legal and ethical principles. These principles apply not only to autonomous weapons but to all categories of weapon systems. Likewise, there are well-established principles governing their deployment and use, all of which must be taken into account. Consequently, discussions concerning the development of autonomous weapons do not take place in a legal or moral vacuum. Rather, they are framed by an existing body of legal norms and ethical principles that guide both the design and the use of military technologies.

Another Example from Practice: Robots for Patient Care

A detailed discussion of this practical example can be found in [8].

Healthcare is currently undergoing a profound transformation driven by the introduction of robots into operating rooms and virtually every other area of medical care. There is, however, no universally accepted definition of what constitutes a care robot. One possible definition is the following:

A care robot is a technological device integrated into healthcare practice that assists healthcare professionals in carrying out tasks associated with patient care.

This raises an important question: When should service robots be employed in healthcare? Ultimately, it is society that must determine the role these technologies should play.

To decide whether the introduction of robots into patient care is desirable, it is not sufficient to consider only economic efficiency and health outcomes. We must also identify the values embedded in caring practices. What, for example, is the role of human dignity, privacy, attentiveness, compassion, companionship, and mutual understanding? One way to identify these values is to examine the fundamental purpose—or purposes—of caregiving. Once these purposes have been clarified, it becomes possible to ask whether the introduction of care robots supports their realization or, conversely, threatens them.

There is a genuine risk that, driven by strong economic incentives, service robots may eventually be deployed on a large scale in healthcare institutions without adequate consideration of the preferences, values, and needs of patients and care recipients. Such technologies could thus be introduced independently of the wishes of those whom they are intended to serve.

A Value Sensitive Design approach provides an appropriate foundation for the future development of service robots by ensuring that the values associated with caregiving are systematically incorporated into the design process. A thorough understanding of caregiving enables designers to develop robots whose operation is compatible with—or even actively supports—the values underlying care. Furthermore, a clear understanding of the distribution of roles and responsibilities within caregiving practices enables designers to make informed decisions about which responsibilities may appropriately be delegated to robots and which should remain under human control, while also identifying responsibilities that ought not to be delegated at all.

CONCLUSION

In this article, we have presented Value Sensitive Design (VSD) as a framework for incorporating and preserving the values that society seeks to realize in contemporary technological innovation. Value Sensitive Design can be formally represented through a Values Hierarchy Matrix and implemented using either a top-down or a bottom-up approach. The structure and conceptual development of this paper were significantly influenced by the insightful work presented in [9].

A visual and transparent representation of values enables stakeholders to discuss them in a constructive and systematic manner. Moreover, it facilitates a critical examination of operational criteria by decomposing them into their constituent elements and questioning assumptions that might otherwise be taken for granted. This, in turn, makes it possible to address questions such as: Are the values that we consider important adequately reflected in the design? Conversely, are the operational criteria embodied in the design capable of realizing the intended values?

This iterative process of relating values, norms, and design requirements enhances both the transparency and the accountability of technological design, thereby supporting the development of technologies that are not only technically effective but also ethically and socially responsible.

REFERENCES

[1]. Langdon Winner, Do Artifacts Have Politics? Daedalus (Journal of the American Academy of Arts and Sciences), Vol. 109, No.1, Modern Technology: Problem or Opportunity? (Winter, 1980), pp. 121-136. https://www.jstor.org/stable/20024652)

[2]. Friedman, B. and Hendry G. D., Value Sensitive Design: Shaping Technology with Moral Imagination, 2nd Edition, Publisher: The MIT Press, 2026, 500 pp., ISBN: 9780262553254

[3]. Friedman, B., Peter H. Kahn Jr., and Borning, A. Value Sensitive Design and Information Systems. In Chapter 4, pp. 69–101 of: The Handbook of Information and Computer Ethics Edited by Kenneth Einar Himma and Herman T. Tavani. https://www.academia.edu/13511658/Value_Sensitive_Design_and_Information_Systems Copyright © 2008 John Wiley & Sons, Inc.

[4]. Lossky N., World views: An introduction to metaphysics, Paris: Modern Notes, 1931, 187 pages. In Russian: Культурно-философская библиотека, Н. Лосский, «Типы мировоззрений, Введение в метафизику», издательство «Современные записки», Склад издания YMCA (International NGO Young Men's Christian Association) – PRESS, Париж, 1931, 187 С.

[5]. Doob, Leonard William; Miller, Neal Elgar; Mowrer, Orval Hobart; Sears, Robert Richardson. «Frustration and Aggression», New Haven, Pub. for the Institute of human relations by Yale University Press; publication date 1939, pages 227. Issued as a Yale Paperbound January 1961.

[6]. ხმელიძე თ., გურგენიძე დ., კლიმიაშვილი ლ., ხმელიძე კ. «სამშენებლო ენციკლოპედიური ლექსიკონი», ტ. I, თბილისი: „ტექნიკური უნივერსიტეტი“, 2019. — გვ. 398.

[7]. «ქართული საბჭოთა ენციკლოპედია», ტ. 2, თბ., 1977. — გვ. 392. 1, 2012. https://eur-lex.europa.eu/eli/dir/1999/74/oj/eng

[8]. Santoni de Sio, F., van Wynsberghe, A. When Should We Use Care Robots? The Nature-of-Activities Approach. Science and Engineering Ethics, 22, 1745–1760 (2016). https://doi.org/10.1007/s11948-015-9715-4

[9]. Joost Groot Kormelink (Editor). Responsible Innovation: Ethics, Safety and Technology. 2nd edition. 2019, TU Delft OPEN Books, ISBN 978-94-6366-201-7, 155 pages. DOI: https://doi.org/10.5074/t.2019.006 .