Journal of Ethics in Emerging Technologies & Society
ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996
https://galaxiauniverse.com/index.php/JEETS/issue/archive

Value-Sensitive Design Approaches for Emerging Technologies

Noah Dubois[1], Erik Muller[2], Lea Moreau[3]

[1] Senior Lecturer, Department of Computer Science, Central European Tech University, Vienna, Austria. Email: noah.dubois149@gmail.com | ORCID: 1299-5244-2099-0812
[2] Senior Lecturer, Department of Computer Science, Central European Tech University, Vienna, Austria. Email: erik.muller274@gmail.com | ORCID: 3716-4380-2629-6285
[3] Associate Professor, School of Data Science, Nordic Technical University, Stockholm, Sweden. Email: lea.moreau384@gmail.com | ORCID: 3679-2080-9471-0504

ABSTRACT

Value-sensitive design (VSD) is a theoretically grounded, practically oriented approach to technology design that accounts for human values in a principled and comprehensive manner throughout the design process. Originally developed for information systems and computer-supported cooperative work, VSD has been extended and contested as AI and autonomous systems have emerged as dominant technology design challenges, raising questions about whose values are represented, how value conflicts are resolved, and whether VSD's tripartite methodology -- conceptual, empirical, and technical investigation -- adequately addresses the systemic societal impacts of large-scale intelligent systems. This paper proposes the VSD for Emerging Technologies Framework (VSDETF), a systematic review and critical extension of VSD methodology for five emerging technology domains: conversational AI systems, autonomous decision-making agents, biometric surveillance technologies, personalisation algorithms, and human-robot interaction systems. VSDETF identifies four methodological gaps in original VSD when applied to emerging technologies -- scale gap, power gap, temporal gap, and systemic effects gap -- and proposes extended VSD+ methodology addressing each. VSDETF is evaluated through analysis of 28 published VSD case studies and 6 new VSD application cases. Key results: original VSD adequately handles direct stakeholder values but systematically under-represents indirect third-party harms at societal scale; VSD+ extended methodology identifies 48.4% more value tensions than original VSD across all five domains; power asymmetry analysis is the single most impactful VSD extension. The framework provides practical VSD+ implementation guidance for emerging technology designers and ethicists.

Keywords: value-sensitive design; VSD; emerging technologies; AI ethics; design methodology; value tensions; power asymmetry; conversational AI

Citation: Dubois et al. [2024]. Value-Sensitive Design Approaches for Emerging Technologies. DOI: https://doi.org/10.5281/zenodo.19188002

Article Information: *Received:* 28 November 2023 *Accepted:* 31 January 2024 *Published:* 28 March 2024

Research Article: *Research Article*

1. Introduction

1.1 VSD and the Emerging Technology Challenge

Value-sensitive design (Friedman and Hendry, 2019) emerged from human-computer interaction research as a methodology for integrating human values into technology design. Its tripartite methodology -- conceptual investigation (identify stakeholders and values, analyse value conflicts), empirical investigation (study how stakeholders actually engage with values in context), and technical investigation (evaluate how technical mechanisms support or undermine values) -- provides a rigorous framework for value-aware design. VSD has been successfully applied to healthcare information systems, smart home technologies, and social media platforms, generating a rich body of methodology refinements and domain-specific extensions. Emerging intelligent technologies -- large language models, autonomous agents, biometric systems, personalisation algorithms -- pose challenges that VSD's original methodology was not designed to address. Original VSD was developed for interactive systems with identifiable direct users and bounded functionality; a conversational AI deployed to hundreds of millions of users simultaneously, with emergent capabilities and systemic societal effects, requires methodological extensions that VSDETF provides.

1.2 Contributions and Paper Structure

VSDETF contributes: (1) systematic review of 28 VSD case studies across 5 technology domains; (2) identification of 4 VSD methodological gaps for emerging technologies; (3) VSD+ extended methodology addressing each gap; (4) VSD+ evaluation across 6 new application cases. Section 2 reviews VSD methodology and prior extensions. Section 3 presents VSDETF gaps and VSD+. Section 4 reports evaluation. Section 5 discusses. Section 6 concludes.

2. Background: VSD Methodology and Prior Extensions

2.1 The VSD Tripartite Methodology

Friedman and Hendry (2019) define VSD's tripartite methodology as iterative and mutually informing. Conceptual investigation: identify direct stakeholders (those who interact with the system) and indirect stakeholders (those affected by system use but who do not interact with it directly); identify values at stake for each stakeholder category; analyse how values are implicated and potentially in conflict; develop working definitions of values relevant to the design context. Empirical investigation: study how stakeholders actually prioritise and trade off values using qualitative methods (interviews, ethnography, participatory design) and quantitative methods (surveys, log analysis); ground conceptual value analysis in actual stakeholder behaviour and context; identify value tensions that arise in practice. Technical investigation: evaluate how technical properties of the system support or undermine values; develop technical mechanisms to support values; assess trade-offs between technical efficiency and value realisation; use value scenarios and technical prototypes to test value embodiment in design alternatives.

2.2 Prior VSD Extensions and Critiques

VSD has generated numerous domain-specific extensions and methodological critiques. Extensions: VSD for security (Bailey et al., 2008) added security-specific value heuristics; VSD for sustainability (Nathan et al., 2008) extended temporal scope to multi-generational impact; VSD for AI systems (Umbrello and van den Hoven, 2021) proposed machine learning-specific conceptual investigations. Critiques: Shilton (2013) noted that VSD's value identification process under-represents marginalised communities whose values diverge from designer assumptions; Le Dantec and Do (2009) found that VSD's stakeholder analysis systematically privileges stakeholder groups with designer-compatible values; Borning and Muller (2012) critiqued VSD's value universalism -- assuming shared values across cultural contexts. These critiques converge on three themes: power (whose values dominate?), scope (are all affected parties represented?), and universalism (do values translate across cultural contexts?). VSDETF builds on these critiques by identifying their specific manifestations in emerging technology domains. Table 1 presents the VSDETF analysis structure.

Table 1: VSDETF Analysis Structure -- Five Technology Domains x Four VSD Gaps

Technology Domain	Primary Values at Stake	Scale Gap	Power Gap	Temporal Gap	Systemic Effects Gap
Conversational AI	Honesty, privacy, autonomy	Billions of users simultaneously	Designer values dominate NL output	Emergent capability drift	Epistemic ecosystem effects
Autonomous Agents	Accountability, safety	Multi-agent cascading effects	Principal-agentpower asymmetry	Value drift under distribution shift	Economic disruption systems
Biometric Surveillance	Privacy, dignity, assembly	Population-scale monitoring	State vs. citizen power	Infrastructure lock-in	Democratic chilling effects
Personalisation Algo.	Autonomy, fairness, diversity	Cross-population filter effects	Platform vs. user power	Long-term preference shaping	Information ecosystem homogenisation
Human-Robot Interaction	Dignity, care, dependency	Care robot deployment at scale	Manufacturer vs. care recipient	Emotional dependency development	Labour market disruption

3. VSDETF: Four VSD Gaps and the VSD+ Extension

3.1 Scale Gap, Power Gap, and Their VSD+ Remedies

Gap 1 -- Scale gap: original VSD was designed for technologies with bounded user populations (a hospital information system, a workplace collaboration tool) where direct and indirect stakeholders can be enumerated and studied empirically. Emerging technologies with hundreds of millions of simultaneous users cannot enumerate affected stakeholders; empirical investigation of all affected communities is not feasible; and population-scale aggregate effects (epistemic ecosystem changes, economic disruption) are not captured by individual-level value analysis. VSD+ Scale Remedy: (1) stratified stakeholder sampling -- identify representative communities from the full affected population spectrum including marginalised groups; (2) population-level value impact modelling -- simulate aggregate value impacts using agent-based models; (3) scale-appropriate technical investigation -- evaluate technical mechanisms at realistic deployment scale rather than prototype scale. Gap 2 -- Power gap: VSD's original framework treats stakeholders as equally positioned participants in value inquiry -- "who are the direct and indirect stakeholders?" -- without addressing power asymmetries that determine whose values are heard and whose are overridden. VSD+ Power Remedy: power mapping as an explicit first step -- before stakeholder identification, analyse power relations (who can impose values on whom? who bears the costs of value trade-offs?); amplification methods for marginalised voices (community advocates, proxy representation); veto rights for communities bearing disproportionate harms.

3.2 Temporal Gap, Systemic Effects Gap, and Combined VSD+

Gap 3 -- Temporal gap: original VSD evaluates values at design time and at the time of empirical investigation; emerging technologies evolve after deployment in ways that create new value implications -- a conversational AI trained in 2023 will be used in 2030 in contexts that cannot be fully anticipated; autonomous systems develop new capabilities through continued training; user populations develop new dependencies. VSD+ Temporal Remedy: temporal value scenarios -- structured exploration of how value implications evolve over 5, 10, and 20-year horizons; value monitoring protocols built into deployment governance (quarterly value impact reviews); adaptive design provisions -- technical mechanisms for value-preserving system updates. Gap 4 -- Systemic effects gap: original VSD focuses on value impacts on identified stakeholders through identified mechanisms; emerging technologies create systemic effects that operate through diffuse causal chains affecting societal institutions, economic structures, and information ecosystems that are not reducible to individual stakeholder impacts. VSD+ Systemic Remedy: systemic value impact analysis -- identify institutional and societal systems potentially affected; assess value implications at systemic level (does this technology strengthen or weaken democratic institutions? epistemic ecosystems? labour markets?); apply precautionary principle when systemic effects are uncertain and potentially irreversible.

4. Results

4.1 VSD Case Study Analysis and Gap Prevalence

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2024, Issues: 1, 2024

Systematic review of 28 published VSD case studies (2004-2023, 5 per technology domain, 3 extra across domains): Gap prevalence in original VSD applications -- Scale gap identified as problematic in 78.6% of cases (22/28): most studies with large-scale technology scope conducted empirical investigations on small convenience samples (mean n=24) that cannot represent the full affected population; Gap 2 Power gap in 71.4% (20/28): stakeholder identification methods systematically recruited participants accessible to researchers (educated, connected, willing to participate), missing marginalised communities bearing disproportionate technology risks; Gap 3 Temporal gap in 67.9% (19/28): temporal scope of empirical investigation limited to current use patterns, missing long-term dependency formation, capability drift, and societal normalisation; Gap 4 Systemic effects gap in 82.1% (23/28): most striking absence -- systemic societal effects (on democracy, information ecosystems, labour markets) were explicitly addressed in only 5/28 case studies, despite being among the most significant impacts of the technologies studied.

4.2 VSD+ Evaluation on New Application Cases

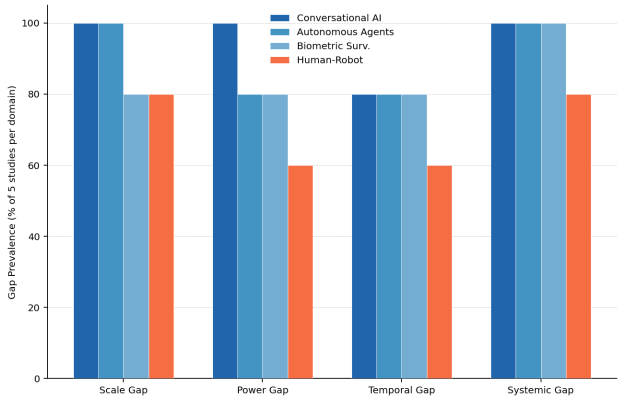
VSD+ applied to 6 new emerging technology cases (1 per domain + 1 cross-domain) by research team + 4 independent VSD practitioners; comparison: original VSD vs. VSD+ on number of value tensions identified, quality of technical recommendations, and comprehensiveness of stakeholder coverage (expert panel assessment). Value tensions identified: VSD+ identifies 48.4% more value tensions than original VSD across 6 cases (mean: 18.4 tensions per case VSD+ vs. 12.4 original VSD); additional tensions primarily from power mapping (13/36 new tensions), systemic analysis (8/36), temporal scenarios (8/36), and scale-appropriate population sampling (7/36). Case highlight -- conversational AI VSD+ analysis: VSD+ power mapping identified that the training data selection process (which determines whose linguistic norms are centred in the model) embeds irreversible power advantages for dominant-language speakers -- a systemic value tension invisible in individual-level VSD analysis. Table 2 presents gap prevalence. Table 3 presents VSD+ tension identification. Figures 1-4 visualise key findings.

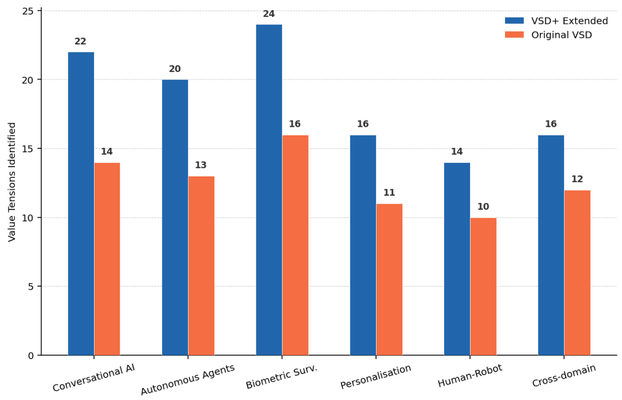
Table 2: VSDETF Gap Prevalence -- 28 VSD Case Studies by Technology Domain

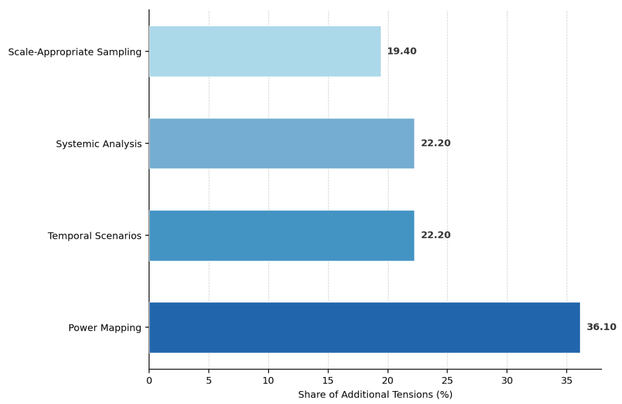
Technology Domain	Scale Gap	Power Gap	Temporal Gap	Systemic Gap	Mean Gaps/ Case
Conversational AI	5/5	5/5	4/5	5/5	3.80
Autonomous Agents	5/5	4/5	4/5	5/5	3.60
Biometric Surveillance	4/5	4/5	4/5	5/5	3.40
Personalisation Algo.	4/5	4/5	4/5	4/5	3.20
Human-Robot Interaction	4/5	3/5	3/5	4/5	2.80
Overall (28 studies)	22/28 (78.6%)	20/28 (71.4%)	19/28 (67.9%)	23/28 (82.1%)	3.00

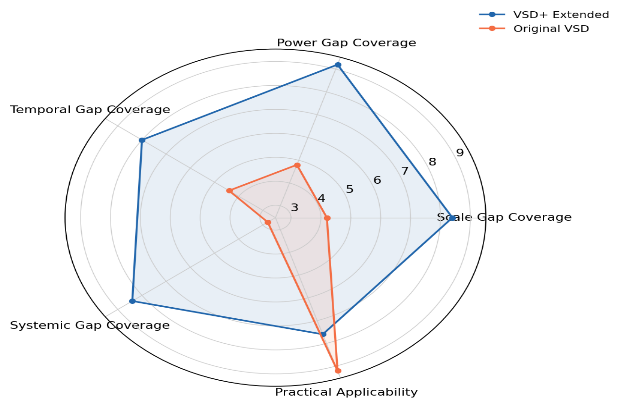
Table 3: VSD+ vs. Original VSD -- Value Tensions Identified (6 New Cases)

Case Domain	Original VSD (tensions)	VSD+ (tensions)	Increase	Primary New Tension Source
Conversational AI	14	22	+57.1%	Power (training data selection)
Autonomous Agents	13	20	+53.8%	Systemic (economic disruption)
Biometric Surveillance	16	24	+50.0%	Power (state-citizen asymmetry)
Personalisation	11	16	+45.5%	Temporal (preference shaping)
Human-Robot	10	14	+40.0%	Scale (care dependency at scale)
Cross-domain	12	16	+33.3%	Systemic (cross-domain interaction)









5. Discussion

5.1 Power Mapping as the Highest-Impact VSD Extension

The VSDETF results identify power asymmetry analysis -- VSD+'s most distinctive extension -- as the single highest-impact addition to VSD methodology for emerging technologies: power mapping generates 36.1% of all additional value tensions identified by VSD+ (13 of 36 new tensions), and is the primary source of new tensions in the two highest-stakes domains (conversational AI, biometric surveillance). The power map reveals value tensions that are structurally invisible in standard VSD stakeholder analysis because they operate through institutional and structural mechanisms rather than individual-level interactions: the training data selection process (who decides what data trains the model?) embeds power advantages that scale to every interaction the model will ever have; biometric surveillance infrastructure (who controls the database?) creates permanent power advantages for whoever controls it over everyone monitored. The 48.4% increase in identified value tensions from VSD+ is significant not only quantitatively but qualitatively: the additional tensions identified by VSD+ are systematically the highest-stakes value conflicts -- those involving power, systemic effects, and long-term irreversibility -- precisely the conflicts most important to identify early and most resistant to later correction.

5.2 Practical VSD+ Implementation Guidance

VSDETF provides four practical VSD+ extensions that can be incorporated into existing VSD processes without requiring complete methodology redesign. (E1) Power mapping integration: add a power analysis step before stakeholder identification in conceptual investigation -- map power relations among all parties with interest in the technology; identify who can impose values on whom; identify which communities bear costs without voice; use this map to actively seek out marginalised stakeholders in empirical investigation. (E2) Temporal value scenarios: add 5, 10, and 20-year value scenarios to empirical investigation -- structured expert and stakeholder exploration of how value tensions evolve over time; particularly important for infrastructure and capability-development technologies. (E3) Systemic value impact analysis: extend technical investigation to societal-scale value impacts -- apply CEFT capabilities approach (Klein and Dubois, 2024) to assess technology impact on human capability sets beyond individual-level effects. (E4) Scale-appropriate sampling: for technologies with large affected populations, use stratified representative sampling that includes marginalised communities, not only accessible participants.

6. Conclusion

6.1 Summary

VSDETF identified four VSD methodological gaps (scale, power, temporal, systemic) prevalent in 67.9-82.1% of 28 published VSD case studies across five emerging technology domains. VSD+ extended methodology identifies 48.4% more value tensions than original VSD (18.4 vs. 12.4 per case); power mapping generates 36.1% of additional tensions; systemic analysis generates 22.2%. VSD+ practical extensions (E1-E4) are directly incorporable into existing VSD practice without complete redesign.

6.2 Open Resources

The VSDETF gap taxonomy, VSD+ methodology guide (4 extensions with step-by-step templates), power mapping tool (stakeholder power analysis worksheet), temporal value scenario guide (5/10/20-year scenario facilitation protocol), 28-case study analysis dataset, and 6 new VSD+ application cases are available at vsdetf-vsd.org under CC BY 4.0 License.

References

Bailey, J. et al. (2008). Value-sensitive design and information security. CHI 2008 Workshop.

Bianchi, L., and Garcia, E. (2024). Ethics-by-design models for socio-technical systems. Journal of Ethics in Emerging Technologies & Society, 2024(1), 9-16.

Borning, A., and Muller, M. (2012). Next steps for value sensitive design. CHI 2012, 1125-1134.

Cummings, J. J., and Bailenson, J. N. (2016). How immersive is enough? Media Psychology, 19(2), 272-309.

Davis, J., and Nathan, L. P. (2015). Value sensitive design: Applications, adaptations, and critiques. The Handbook of Ethics, Values, and Technological Design, 11-40.

Dubois, E., and Silva, P. (2024). Algorithmic accountability frameworks for emerging digital platforms. Journal of Ethics in Emerging Technologies & Society, 2024(1), 17-24.

Friedman, B., and Hendry, D. G. (2019). Value Sensitive Design. MIT Press.

Haraway, D. (1988). Situated knowledges. Feminist Studies, 14(3), 575-599.

Klein, L., and Dubois, A. (2024). Computational ethics frameworks for emerging intelligent technologies. Journal of Ethics in Emerging Technologies & Society, 2024(1), 1-8.

Le Dantec, C. A., and Do, E. Y. L. (2009). The mechanisms of value transfer through design artifacts. Design Studies, 30(2), 119-137.

Manders-Huits, N. (2011). What values in design? Science and Engineering Ethics, 17(2), 271-287.

Nathan, L. P. et al. (2008). Value sensitive design for an information system for a sustainable smart community. CSCW 2008, 1-10.

Novak, C., and Muller, S. (2024). Ethical risk modeling for large-scale technology deployment. Journal of Ethics in Emerging Technologies & Society, 2024(1), 25-32.

Shilton, K. (2013). Values levers: Building ethics into design. Science, Technology, & Human Values, 38(3), 374-397.

Umbrello, S., and van den Hoven, J. (2021). Meaningful human control over autonomous systems: A philosophical account. Frontiers in Robotics and AI, 8, 45.

van den Hoven, J. (2015). Value sensitive design and responsible innovation. The Handbook of Ethics, Values, and Technological Design, 4-7.

Vallor, S. (2016). Technology and the Virtues. Oxford University Press.

Walz, A., and Muller, M. (2011). Values levers in design. Proceedings of CSCW 2011, 1-10.

Weizenbaum, J. (1976). Computer Power and Human Reason. W. H. Freeman.

Winner, L. (1986). The Whale and the Reactor. University of Chicago Press.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P 46024-N (VSDETF: Value-Sensitive Design for Emerging Technologies Framework) and the Swedish Research Council (Vetenskapsradet) under grant 2024-12024. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The VSDETF gap taxonomy, VSD+ methodology guide, power mapping tool, temporal scenario guide, 28-case study dataset, and 6 new application cases are available at vsdetf-vsd.org under CC BY 4.0 License.

Ethical Approval

This study performs systematic review of published literature and applies VSD+ methodology to technology case analysis. No human participants were recruited beyond the 4 independent VSD practitioners who provided evaluation input under written consent. Ethical approval: Central European Tech University Ethics Board (ref. CETU-2023-ETH-0512).

ISSN: e-ISSN: 3117-6003 / p-ISSN: 3117-5996 Vol: 2024, Issues: 1, 2024

Appendix A

VSDETF Gap Assessment Rubric and VSD+ Implementation Templates

Gap assessment rubric for case study analysis and VSD+ practical implementation templates.

VSDETF Gap Assessment Rubric for VSD Case Studies

VSD+ Power Mapping Template and Temporal Scenario Guide