A Lifecycle Approach for Artificial Intelligence Ethics in Energy Systems

El-Haber, N., Burnett, D., Halford, A., Stamp, K., De Silva, D., Manic, M. & Jennings, A

DOI 10.3390/en17143572
ESSN 1996-1073

Publisher: MDPI

energies

MDPI

Article

A Lifecycle Approach for Artificial Intelligence Ethics in Energy Systems

Nicole El-Haber [1], Donna Burnett [1], Alison Halford [2], Kathryn Stamp [2], Daswin De Silva [1,*], Milos Manic [3] and Andrew Jennings [1]

[1] Centre for Data Analytics and Cognition, La Trobe University, Bundoora, VIC 3086, Australia
[2] Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 5FB, UK
[3] Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA
* Correspondence: d.desilva@latrobe.edu.au

Abstract: Despite the increasing prevalence of artificial intelligence (AI) ethics frameworks, the practical application of these frameworks in industrial settings remains limited. This limitation is further augmented in energy systems by the complexity of systems composition and systems operation for energy generation, distribution, and supply. The primary reason for this limitation is the gap between the conceptual notion of ethics principles and the technical performance of AI applications in energy systems. For instance, trust is featured prominently in ethics frameworks but pertains to limited relevance for the robust operation of a smart grid. In this paper, we propose a lifecycle approach for AI ethics that aims to address this gap. The proposed approach consists of four phases: design, development, operation, and evaluation. All four phases are supported by a central AI ethics repository that gathers and integrates the primary and secondary dimensions of ethical practice, including reliability, safety, and trustworthiness, from design through to evaluation. This lifecycle approach is closely aligned with the operational lifecycle of energy systems, from design and production through to use, maintenance, repair, and overhaul, followed by shutdown, recycling, and replacement. Across these lifecycle stages, an energy system engages with numerous human stakeholders, directly with designers, engineers, users, trainers, operators, and maintenance technicians, as well as indirectly with managers, owners, policymakers, and community groups. This lifecycle approach is empirically evaluated in the complex energy system of a multi-campus tertiary education institution where the alignment between ethics and technical performance, as well as the human-centric application of AI, are demonstrated.

Keywords: AI ethics; responsible AI; energy AI; AI risks; AI lifecycle; energy systems; implementation science

Citation: El-Haber, N.; Burnett, D.; Halford, A.; Stamp, K.; De Silva, D.; Manic, M.; Jennings, A. A Lifecycle Approach for Artificial Intelligence Ethics in Energy Systems. *Energies* **2024**, *17*, 3572. https://doi.org/10.3390/en17143572

Academic Editor: Adel Merabet

Received: 13 June 2024
Revised: 30 June 2024
Accepted: 17 July 2024
Published: 20 July 2024

1. Introduction

Recent work on the evaluation of artificial intelligence (AI) ethics guidelines and frameworks has revealed the challenges of advancing the practice of AI ethics for industrial and professional work environments [1,2]. Jobin et al. [1] identify 84 ethics guidelines for AI, with more than 80% being published since 2016, while Hagendorff [2] conducted a semi-systematic evaluation that analyzed and compared 22 guidelines focusing on intersecting themes and, Mittelstadt [3] emphasized the resemblance of AI ethics to the four classical principles of medical ethics. Current attempts at AI ethics in industry practice are not sufficiently agile for the ethical challenges of technologically advanced industrial settings, such as energy systems [4,5], smart cities [6,7], intelligent transport [8,9], and smart factory settings [10]. In its role as critical infrastructure, AI ethics in energy systems are fundamentally challenging due to the complexity of human–system dependence, a system of subsystems composition, and the diversity of data generated by these systems. This is further impacted by the computational resource requirements for processing, analysis,

and synthesis of large volumes of data and machine-learning outcomes that lead to AI capabilities [11,12].

While regulations may be informed by ethical thinking in part, regulations are not the same as ethics. However, it is often believed that regulations are sufficient in 'covering' the ethical concerns emerging from technological engagement, particularly around data use [13,14]. In contrast, treating regulations as the same as ethics has the potential to disempower knowledge work and stifle creativity and innovation, as often there is limited flexibility within regulations for creative thinking and mitigating risks on a case-by-case basis [15]. Therefore, a more dynamic approach to ethics and ethical infrastructure is needed to attend to the shifting landscape of AI in different sectors due to the numerous and differing perspectives of ethics [16]. Although there are strategies in place that offer some space to analyze ethical conundrums, these have been found to be ineffectual [17]. Ethical review boards and risk assessment models [18] have their limitations, and a more dynamic approach to designing ethical guidance is needed in accordance with the complexity of ethics that emerge from automated and autonomous systems. Finding a compromise between flexibility and strictness is necessary [19]. Approaches that supply sufficient space to shift, change, and debate the challenges faced are needed, with the agility to adjust as needed and the foresight to pre-empt problems.

Conventional responses to ethical challenges have been responsive rather than proactive [20]. However, the perceived risk-averse nature of ethical engagement can be flipped to pre-empt issues rather than responding to them when the negative consequences often cannot be sufficiently rectified. There is a need to evaluate and compare how technology can be harnessed and utilized against how it might be overused or misused [21]. Collaboration has been identified as a key facet in navigating emerging ethical issues and facilitating innovation, as well as looking to other disciplines, particularly social science, and disciplines that often encounter ethical dilemmas, can offer new perspectives and strategies for tackling them [22]. IEEE Ethically Aligned Design (EAD) [23] and the Ethics Guidelines for Trustworthy AI by the European Union's High-Level Expert Group on AI (HLEG) [24] are the two of the most established, widely cited and applied AI ethics frameworks in industrial settings. The European Commission has further proposed the regulation of AI, focusing on the most problematic use cases of AI, material distortion of human behavior, and the exploitation of human vulnerabilities [25], which are not within the remit of industrial systems. Therefore, our focus is on the Ethics Guidelines for Trustworthy AI published by the European Union's High-Level Expert Group on AI (EU HLEG). To achieve inclusive, equitable, and progressive human-centric cyber-physical systems (CPS), the EU HLEG states that AI design requires three overarching principles. First, adherence to 'Lawful AI', which holds AI actors and processes to account to regulatory and legislative bodies. Second, 'Ethical AI', which commits to values, processes, and practices that respect and protect the dignity and well-being of all individuals and communities. Third, a 'Robust AI' that limits unintentional harm. Although framed as discrete operations, practices and policies that collectively embed these three components will ensure respect for human rights, be transparent about responsibility and governance, and prioritize duty care. In turn, meeting these conditions will achieve global aims for trustworthy AI.

The demarcating of lawful AI as a separate concern from ethical AI recognizes how legislation is instrumental in ensuring procedures protect and prosecute human actors responsible for AI decisions and/or consequences. However, the purpose of the law is to make judgments about the extent autonomy and self-determination at the micro, meso, and macro levels are permitted [26]. This epistemic stance presents challenges when deploying automated agents, such as AI, as their function is to replace human decisions. Inevitably, this means the deployment of CPS is not, at times, easily translated into legislation. In contrast, ethics is not limited to what extent certain AI practices should be legitimized or not. Ethical AI is also concerned about responsibility and accountability in relation to the impact of the interconnectedness of society and technological advances on the everyday lives of people. Ethical AI promotes transparency and explainability, which acknowledge

public and private concerns about how AI and human-centric cyber-physical systems interact with the material world. Ethics can negotiate complex scenarios, including to what extent Al technologies in the workplace have impacted employment rights, relationships, and governance structures. This is further supported by a taxonomy of AI ethics from practitioner viewpoints for identifying and understanding the different aspects of AI ethics, including awareness, perception, need, challenge, and approach [27]. Another study proposed AI ethics literacy and skills as a means to train a workforce capable of building ethical AI systems, with approaches such as teaching the ethical design of AI algorithms in collaboration with interdisciplinary and industry practitioner input [28,29]. In energy systems, the complexity of systems composition and systems operation for energy generation, distribution, and supply are challenges for the implementation and practice of AI ethics. The primary reason for this limitation is the gap between the conceptual notion of ethics principles and the technical performance of AI applications in energy systems.

2. The Proposed AI Ethics Lifecycle Approach for Energy Systems

Drawing on this current landscape of AI ethics, we propose a lifecycle approach for AI ethics in energy systems as depicted in Figure 1. It consists of four phases: design, development, operation, and evaluation, which are supported by a central AI ethics repository for ethical practice issues and resolutions. As indicated by the bidirectional arrows in Figure 1, the repository provides and receives ethics information to each phase. The lifecycle is closely aligned with the operational lifecycle of energy systems, starting with design, followed by use, maintenance, and repair, and finally, decommissioning, recycling, and replacement. The design phase outputs a multi-granular matrix of ethics principles and paradigms for tuples of subject, requester, and temporality that is situated within the atomic and composite elements of the energy system design workflow. The development phase maps this onto the technical and technological specification for system/solution generation and outputs an 'AI ethics operation matrix' that feeds into the operation phase. This operation matrix intersects with the system operators within the energy system setting to generate AI ethics operator profiles. Finally, the performance of the operational system is monitored and analyzed in real time, across multiple evaluation metrics to inform the subsequent iteration of this lifecycle approach. The following four subsections deliberate the workings of each of the four phases.

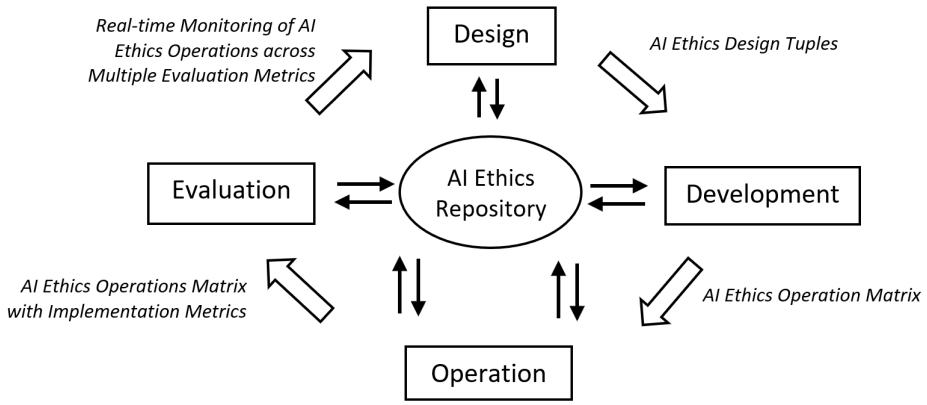

Figure 1. The proposed lifecycle approach for AI ethics of energy systems.

2.1. Design Phase

The design phase is predominantly focused on building out the multi-granular matrix of ethics principles and paradigms for tuples of subject, requester, and temporality. This notion of tuples is motivated by several recent studies in AI ethics [30,31]. As discussed in [32], the 'subject' will be primarily the energy system, but also the operators of the system, the output generated by the system, and its integration/connection to other systems. The 'requester' can vary across several levels of stakeholders, starting with operators requesting

ethics for transparency and safety or managers and owners seeking ethics for accountability acceptance. 'Temporality' is a measure of the duration, assuming most ethics requirements are construed at the start of the function, feature, or operation. Therefore, the duration of ethics may continue until the entire system is operational or until specific results or sub-module operation. As depicted in Figure 2, the paradigms and principles represent the theory of ethics feeding into the tuples, while the industrial system design workflow feeds into the tuples through its constituent design elements. The three dimensions of the tuple are: (1) Subject: the ethics of the stakeholders directly involved in the lifecycle of the energy system and the ethical behavior of the system itself; (2) Requester: stakeholders seeking out the ethics across several layers, starting from low-level operators and/or consumers up to community groups and the overall society; and (3) Temporality: which time period of the system is of concern to the stakeholders. This ranges across the beginning of life, intermediate and end of life of the system. Each unique tuple (subject, requester, time) will engage specific aspects of ethics are concerned, translated into "properties". These properties can be approached using different ethical paradigms, mainly deontology, and consequentialism, which are the most effective in the context of applied ethics.

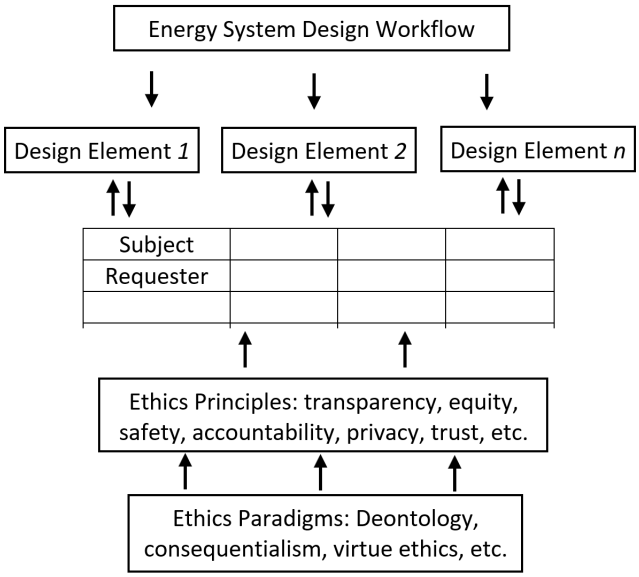

Figure 2. Multi-granular matrix of ethics tuples in the design phase.

2.2. Development Phase

The development phase receives the ethics design tuples as input from the design phase. This content undergoes several iterations of refinement and updates to produce the AI ethics operation matrix for the subsequent phase of the lifecycle operation. The development phase is grounded on the five high-level capabilities of AI: prediction (and forecasting), classification (and detection), association (and segmentation), optimization, and generation [33]. The first four capabilities originate in conventional AI or narrow AI, while the generative capability is a more recent manifestation of generative AI [34]. Expanding this out into energy systems, prediction and time-series forecasting models are typically developed for energy usage forecasting, long-term demand prediction, emissions prediction, and renewables generation prediction. Classification models are widely used for the detection of anomalies in energy usage detection of usage drift, which would then lead to the development of new baselines for prediction and the classification of energy usage profiles, typically high to low usage. Association and segmentation are also used for energy usage (or generation) profiling as well as distribution network stability profiling, given large volumes of unlabeled data that represent the control and management of the grid over time. Optimization models are widely used in most grids, typically for the complex task of load balancing between demand and supply from conventional or

renewable sources. Optimization is also useful for grid management, control tasks, and resource allocation. Finally, generative AI is still an emerging area of application within energy systems. Early adopters include data augmentation, synthetic data generation for digital twin development, multimodal data generation for improved supervised learning outcomes, such as from images of grids, plants, and distribution networks, as well as conversational AI interfaces for some levels of stakeholder engagement.

The AI ethics design tuples are synthesized across the key elements of the development phase that apply across the high-level AI capabilities mentioned above. These elements, in order of sequence, are data acquisition, data staging, data storage, data preprocessing, model development, model training, model testing, model evaluation, and model deployment. Each of the subject, requester, and temporality tuples is mapped into these elements to identify and analyze the transition of the ethics requirement from the design phase into the development phase. This synthesis pipeline is depicted in Figure 3.

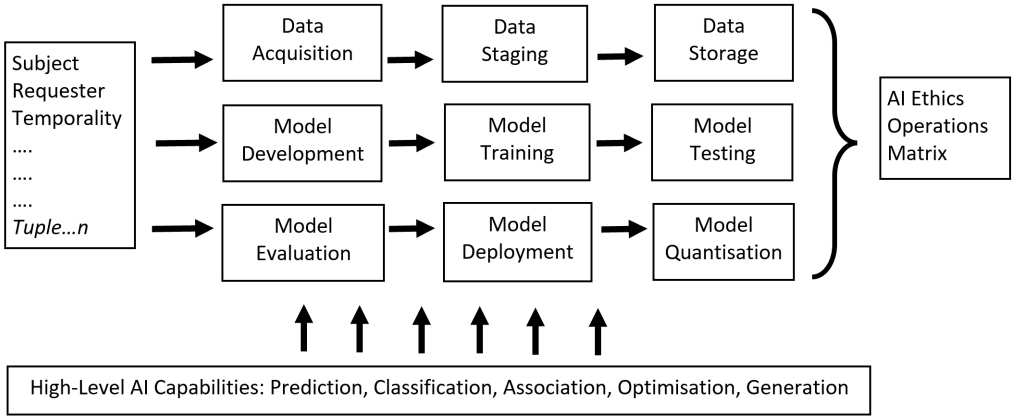

Figure 3. Transformation of AI ethics design tuples in the development phase, informed by the high-level AI capabilities.

2.3. Operation Phase

The operation phase is grounded on the eight implementation outcomes of acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability. These implementation outcomes are drawn from methods in Implementation Science, where the focus is on the scientific inquiry into questions concerning implementation—the act of carrying an intention into effect in the form of evidence-based practices, interventions, and policies [35]. Although primarily practiced in healthcare, Implementation Science directly contributes to the realization of the practical needs of AI ethics in energy systems. The eight implementation outcomes are distinguished from service outcomes, such as efficiency, safety, effectiveness, and temporality, which can be translated into system availability factors in the energy domain. These implementation outcomes are deliberated below.

Acceptability: Outside service completion, acceptability represents the stakeholder awareness and preference for the service, practice, or innovation that is provided by the AI capability. Acceptability should be assessed based on stakeholder knowledge and experience in the domain and with the AI capability itself in diverse settings. An example of acceptability is the incentive scheme recommended by an AI service for retail vs industrial energy consumers or prosumers.

Adoption: Also known as uptake, it is the intention or decision to start using an AI capability for a service requirement in terms of its innovation or evidence-based practice. The adoption of recommender systems for energy usage profiling by a single provider vs all providers within the national grid is an example where the operation itself can be evaluated.

Appropriateness: Although similar in practice to acceptability, this outcome can be distinguished in certain operational settings where a practice can be perceived as appropriate but not acceptable, and vice versa. Appropriateness is also the relevance or

compatibility of the AI innovation for a given practical setting. For example, the use of time-series forecasting algorithms for demand prediction in a highly volatile energy setting can be deemed inappropriate due to the inherent inaccuracies of the data in that volatile consumption setting.

Cost: The cost impact of the implementation in an operational setting. The cost will be composed of at least three components: design, development, and deployment cost, and these can vary based on the size of the energy system, volume of data, number of subscribers, nature of financial transactions, etc. The cost of AI capabilities needs to be offset against the benefits gained by the stakeholders, including suppliers, staff, and consumers.

Feasibility: The capacity to which a new AI capability can be operationalized within a given setting. Although closely linked to appropriateness, it is also different in its technical needs for resourcing, data, or computation. Human-in-the-loop configurations are mandated for AI capabilities, and this can impact the feasibility of system-wide deployment and integration of AI capabilities.

Fidelity: The quality of the actual implementation, to which degree an implementation was deployed as per the original design or specification. Fidelity can be decomposed into further dimensions of adherence, quality of delivery, differentiation, exposure, and responsiveness. In energy systems, the fidelity of AI capabilities is paramount for ensuring service quality and availability. AI capabilities developed in test and trial settings can vary significantly in their field performance following deployment. For example, smart meter data stream outages can have an adverse effect on the predictions of usage/generation which in turn can disrupt service quality and availability.

Penetration: Describes the extent of integration of an AI capability within an energy system operation. It can be identified in terms of the number of services within a single system or the number of systems within a set of service providers that may operationalize a given AI capability. Short-term forecasting and long-term forecasting are typical examples, where the former represents high penetration compared to the latter.

Sustainability: The extent of an AI capability to be integrated and adopted within a service setting's ongoing, stable operations. AI capabilities are quite often dismissed or overlooked in preference of domain expertise of familiar means of operation. Sustainability evaluates the extent to which this changes following operationalization. Sustainability is further described in the three phases of passage, cycle or routine, and niche saturation to indicate coverage and usage within a given setting.

2.4. Evaluation Phase

The evaluation phase receives the AI ethics operations matrix with metrics for the implementation outcomes deliberated above. As depicted in Figure 4, these implementation metrics are aligned together with service metrics, AI performance metrics and system/platform performance metrics in the evaluation phase to deliver real-time monitoring of the energy system during its in-field operation. The service metrics are representative of service availability features of efficiency, safety, effectiveness, and temporality. AI performance metrics are mainly focused on accuracy and speed, with accuracy metrics for AI models varying across capabilities, prediction, classification, association, optimization, and variants of algorithms used for these capabilities. As a baseline, it is recommended to introduce Root Mean Squared Error (RMSE) and related measures for prediction, F1-score/AUC curve for classification, cluster purity/entropy for association, and convergence/solvability for optimization. Results and outcomes from this evaluation phase are collated as inputs into the subsequent iteration of the AI ethics lifecycle, where the design phase will begin with a review of the metrics against the design and operational objectives of the energy system. The large number of metrics evaluated in this phase will be recorded and analyzed within the same AI ethics repository for consistency and reliability.

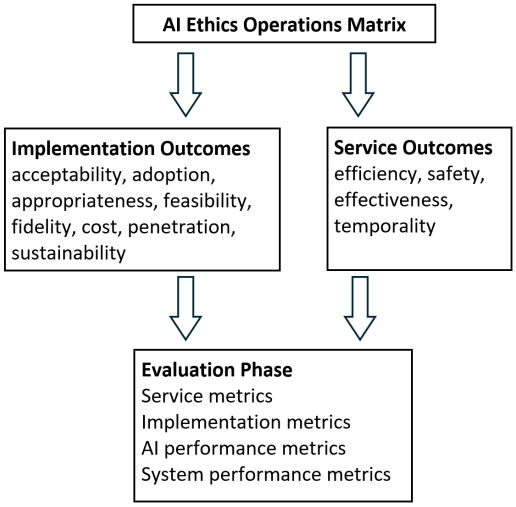

Figure 4. Transition from operation to evaluation phase with implementation and service metrics integration.

3. Evaluation of the Proposed Approach

The proposed AI ethics lifecycle approach was empirically evaluated and demonstrated in the real-world energy system of a multi-campus education institution. The La Trobe Energy AI Platform (LEAP) is a mid-sized microgrid with more than 200 buildings (consumers) and 20 renewable installations (generators) that service approximately 40,000 individuals during work hours (university operations) and non-work hours (recreational, sport, and accommodation operations). Given the human-centric operation of this energy system and the ethical focus of this article, it is pertinent to deliberate the context of the operation of the university. Founded in 1964, La Trobe University is currently ranked in the top 300 of all three major world university rankings. The university offers undergraduate, postgraduate, technical, and executive education courses to a student body of approximately 39,000 and employs close to 4000 staff [36]. The facilities and services operation of the university is spread across its multiple campuses in geographically distributed regions in the state of Victoria, Australia. The main metropolitan campus is La Trobe's Bundoora campus, and the other four campuses are in regional locations: Bendigo, Shepparton, Albury-Wodonga, and Mildura. All five campuses are in the state of Victoria, which has a temperate oceanic climate and four seasons: summer from December to February, autumn from March to May, winter from June to August, and spring from September to November. The largest campus (Bundoora campus) has a central heating system powered by gas that connects to all the buildings, while electricity powers all other amenities, with 13× absorption chillers and 18× air-cooled electric chillers. This campus has a mix of old and modern buildings where several LED retrofit and solar installation projects are located. LEAP AI capabilities have been discussed and evaluated across several studies, including generative AI chatbots [37], measurement and verification [38], and solar irradiance forecasting [39].

Within this operational context, it is evident that the microgrid interfaces with many human operations and human stakeholders, including designers, engineers, users, trainers, operators, and maintenance technicians, as well as indirectly with managers, owners, policymakers, and community groups. In applying the AI ethics lifecycle approach to LEAP, the initial phase of design consists of compiling the multi-granular matrix of ethics principles and paradigms for tuples of subject, requester, and temporality. The design elements identified from the LEAP design workflow include data representation, data processing, data quality, system accuracy, system responsiveness, system usability, system integration, insight quality, and decision quality. The design tuples are presented in Table 1.

In the development phase, the AI ethics design tuples are transformed into the AI ethics operations matrix through the alignment and application of AI capabilities. This phase brings together subject, requester, and temporality with the specific AI capability that can influence the ethical practice of the entire system and its manifestation of AI as a system capability. For clarity, this alignment and application of AI capabilities is presented in Table 2, where the first column depicts the design element, followed by column 2 for ethics practice and column 3 for AI capability. The distinction between data, system, and outcomes for the application and practice of ethics is clearly depicted in this Table. For instance, data-related design/system elements require a high degree of safe, secure, inclusive, and effective undertakings, whereas system elements are more focused on reliability, resilience, accountability, and explainability. System outcomes are a balance between the two, where decisions must be inclusive, equitable, as well as effective, robust, and safe from system operations and performance positioning.

Table 1. Tuples from the design phase of the AI ethics lifecycle.

Design Element	Subject	Requester	Temporality
Data Representation	Energy System, Sub-Modules, Downstream Applications	Owners, Designers, Developers, Engineers	Design to Deployment
Data Processing	Energy System, Sub-Modules, Downstream Applications	Designers, Developers, Engineers, Operators	Design to Operation
Data Quality	Energy System, Sub-Modules, Decision Processes	Designers, Developers, Engineers, Operators, Managers, Policymakers	Design to Operation
System Accuracy	Energy System, Sub-Modules, Downstream Applications, Consumer Experience	Engineers, Operators, Managers, Policymakers, Users	Operation to Termination
System Responsiveness	Downstream Applications, Consumer Experience	Operators, Technicians, Policymakers, Owners	Operation to Termination
System Usability	Energy System, Downstream Applications, Consumer Experience	Operators, Technicians, Policymakers, Owners	Design to Operation
System Integration	Energy System, Integration Platforms	Operators, Managers, Engineers, Owners	Operation to Expansion
Insight Quality	Energy System, Decision Processes, Policy Implements	Operators, Managers, Engineers, Owners, Consumers	Operation to Termination
Decision Quality	Energy System, Decision Processes, Policy Implements	Owners, Managers, Policymakers, Advocacy, Community	Operation to Termination

Following the development phase, the operations matrix is received by the operation phase for assessment and monitoring against the eight implementation outcomes deliberated earlier. These implementation outcomes, along with service, AI, and system performance metrics, provide a comprehensive account of the AI ethics practice, starting from design elements to AI and system capabilities. For instance, adopting this AI ethics lifecycle approach provides an impactful and informative association between the design construct of data representations and the implementation needs of an AI model for predicting the data and the required data volumes for the optimal and useful operation of the energy system. This association between design concept and technical implementation consolidates ethical oversight for owners, operators as well users, advocacy and community groups alike. System usability is a further example that is generally overlooked for ethical balance during technical implementation. Usability is not predefined by the operational capabilities, instead informed and motivated by the design phase for an ethical implementation that is accountable, effective, explainable, and inclusive, where evolving user needs and evolving data volumes will predict how usability needs to be redesigned or adapted for inclusive operation of the energy system.

Table 2. Operations matrix from the development phase of the AI ethics lifecycle.

Design Element	Ethics Practice	AI Capability
Data Representation	Effective, Explainable, Inclusive, Robust, Trustworthy	Prediction of required data representations, Prediction of required data volumes, Classification of missing or erroneous data
Data Processing	Effective, Robust, Resilient, Privacy, Safe, Secure, Trustworthy	Prediction of process workloads, Outlier detection in processes, Optimization of processing based on patterns of recurrences
Data Quality	Accountable, Effective, Explainable, Inclusive	Classification of quality factors and variability, Prediction of loss of quality, Optimization of quality thresholds for system outcomes
System Accuracy	Effective, Explainable, Resilient, Robust, Safe, Secure, Trustworthy	Prediction of drop/loss in system accuracy, Metric optimization for performance gains, Association of metrics with system operation goals
System Responsiveness	Reliable, Resilient, Robust, Safe, Secure, Trustworthy, Unbiased	Operational load prediction, Classification of response times based on operator profiles, Profiling operator capabilities
System Usability	Accountable, Effective, Explainable, Inclusive	Prediction of evolving user needs, Prediction of evolving data volumes, Profiling usability factors by user groups
System Integration	Reliable, Resilient, Robust, Safe, Secure	Predicting downstream system dependencies, Classification of integration points and system checks, Association of integration logs, timeouts, and dropouts
Insight Quality	Effective, Explainable, Inclusive, Safe, Secure, Trustworthy	Classification of insight quality and acceptable accuracy thresholds, Prediction of insight quality based on data quality factors
Decision Quality	Accountable, Effective, Explainable, Inclusive, Safe, Secure, Trustworthy	Classification of decisions by downstream impact, Profiling socio-technical implications of decisions, Optimization of decisions for socio-economic gains

4. Conclusions

This article presents our work in the conception and development of an AI ethics lifecycle approach for energy systems. The proposed approach consists of the design, development, operation, and evaluation phases that collectively deliver an AI ethics operations matrix for real-time monitoring and evaluation of AI capabilities for ethically informed, balanced, and inclusive operations and decision-making. The centralized AI ethics repository captures and manages a persistent record of design considerations, implementation outcomes, and evaluation metrics associated with each capability and ethics practice. This approach aligns with the lifecycle of most industrial systems (including energy systems), where design and production are followed by maintenance, repair, and then shutdown and replacement. The lifecycle approach was evaluated in the real-world setting of a microgrid energy system deployed and operational in a multi-campus tertiary education setting. The results of this evaluation, presented in terms of the design and development outcomes, provide comprehensive coverage and visualization of the interplay between ethics practice and technical implementation of an energy system. All stakeholders of this energy system can review and evaluate the design considerations, technical development, and subsequent performance evaluation of AI capabilities in alignment with the ethics practice mandated for the entire system. In future work, we intend to expand and adapt this lifecycle approach for generalized application across any industrial system with consistent human-centric operation and human-machine interactions. The adaptation of this approach for diverse energy settings such as transport, healthcare, retail, and residential, as well as the develop-

ment of evaluation metrics for the effectiveness, performance, usability, and governance of the lifecycle approach to determine certainty and reliability, are further considerations for future work. Based on the continuing effectiveness of this approach, we will also work on the extension of the lifecycle from energy systems to generic industrial systems while also taking into account the evolving nature of AI and the increasing/diversifying dimensions of AI ethics that must be considered. In this case, the proposed approach will serve as a performance baseline for AI ethics implementation and new dimensions will be introduced and integrated upon this baseline. This lifecycle approach in its current form and the extent of future work will continue to contribute toward the practice of responsible AI in industrial settings and address the translation gap between policy and practice in AI ethics undertakings.

Author Contributions: Conceptualization, D.B., D.D.S., M.M. and A.J.; Methodology, N.E.-H., A.H., K.S., D.D.S., M.M. and A.J.; Software, D.D.S.; Validation, A.J. and D.B.; Formal analysis, N.E.-H., D.B., A.H., K.S. and M.M.; Investigation, N.E.-H., D.B., A.H., K.S. and A.J.; Resources, D.B.; Data curation, N.E.-H.; Writing—original draft, N.E.-H., D.B., A.H., K.S., D.D.S., M.M. and A.J.; Visualization, A.H.; Supervision, D.D.S. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Australian Government's Department of Climate Change, Energy, the Environment and Water under the International Clean Innovation Researcher Networks (ICIRN) program grant number ICIRN000077.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to human research ethics requirements.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]
2. Hagendorff, T. The ethics of AI ethics: An evaluation of guidelines. *Minds Mach.* **2020**, *30*, 99–120. [CrossRef]
3. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. [CrossRef]
4. Bose, B.K. Artificial intelligence techniques in smart grid and renewable energy systems—Some example applications. *Proc. IEEE* **2017**, *105*, 2262–2273. [CrossRef]
5. De Silva, D.; Yu, X.; Alahakoon, D.; Holmes, G. Semi-supervised classification of characterized patterns for demand forecasting using smart electricity meters. In Proceedings of the 2011 International Conference on Electrical Machines and Systems, IEEE, Beijing, China, 20–23 August 2011; pp. 1–6.
6. De Silva, D.; Burstein, F.; Jelinek, H.; Stranieri, A. Addressing the complexities of big data analytics in healthcare: The diabetes screening case. *Australas. J. Inf. Syst.* **2015**, *19* . [CrossRef]
7. Nawaratne, R.; Alahakoon, D.; De Silva, D.; Kumara, H.; Yu, X. Hierarchical two-stream growing self-organizing maps with transience for human activity recognition. *IEEE Trans. Ind. Inform.* **2019**, *16*, 7756–7764. [CrossRef]
8. Nallaperuma, D.; De Silva, D.; Alahakoon, D.; Yu, X. Intelligent detection of driver behavior changes for effective coordination between autonomous and human driven vehicles. In Proceedings of the IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society, IEEE, Washington, DC, USA, 21–23 October 2018; pp. 3120–3125.
9. Nawaratne, R.; Bandaragoda, T.; Adikari, A.; Alahakoon, D.; De Silva, D.; Yu, X. Incremental knowledge acquisition and self-learning for autonomous video surveillance. In Proceedings of the IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society, IEEE, Beijing, China, 29 October–1 November 2017; pp. 4790–4795.
10. Chamishka, S.; Madhavi, I.; Nawaratne, R.; Alahakoon, D.; De Silva, D.; Chilamkurti, N.; Nanayakkara, V. A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling. *Multimed. Tools Appl.* **2022**, *81*, 35173–35194. [CrossRef]
11. Wu, C.J.; Raghavendra, R.; Gupta, U.; Acun, B.; Ardalani, N.; Maeng, K.; Chang, G.; Aga, F.; Huang, J.; Bai, C.; et al. Sustainable ai: Environmental implications, challenges and opportunities. *Proc. Mach. Learn. Syst.* **2022**, *4*, 795–813.
12. Kleyko, D.; Osipov, E.; De Silva, D.; Wiklund, U.; Alahakoon, D. Integer self-organizing maps for digital hardware. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, Budapest, Hungary, 14–19 July 2019; pp. 1–8.
13. Ziosi, M.; Mökander, J.; Novelli, C.; Casolari, F.; Taddeo, M.; Floridi, L. The EU AI Liability Directive: Shifting the burden from proof to evidence. *AI Soc. Knowl. Cult. Commun.* **2023** . [CrossRef]
14. Mökander, J.; Floridi, L. Operationalising AI governance through ethics-based auditing: An industry case study. *AI Ethics* **2023**, *3*, 451–468. [CrossRef]

15. Morley, J.; Floridi, L.; Kinsey, L.; Elhalal, A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci. Eng. Ethics* **2020**, *26*, 2141–2168. [CrossRef]
16. Novelli, C.; Casolari, F.; Rotolo, A.; Taddeo, M.; Floridi, L. Taking AI risks seriously: A new assessment model for the AI Act. *AI Soc.* **2023**, 1–5. [CrossRef]
17. Vitak, J.; Proferes, N.; Shilton, K.; Ashktorab, Z. Ethics regulation in social computing research: Examining the role of institutional review boards. *J. Empir. Res. Hum. Res. Ethics* **2017**, *12*, 372–382. [CrossRef]
18. de Almeida, P.G.R.; dos Santos, C.D.; Farias, J.S. Artificial intelligence regulation: A framework for governance. *Ethics Inf. Technol.* **2021**, *23*, 505–525. [CrossRef]
19. Morley, J.; Elhalal, A.; Garcia, F.; Kinsey, L.; Mökander, J.; Floridi, L. Ethics as a service: A pragmatic operationalisation of AI ethics. *Minds Mach.* **2021**, *31*, 239–256. [CrossRef] [PubMed]
20. Floridi, L.; Cowls, J.; King, T.C.; Taddeo, M. How to design AI for social good: Seven essential factors. In *Ethics, Governance, and Policies in Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 125–151.
21. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef] [PubMed]
22. Trentesaux, D.; Rault, R. Designing ethical cyber-physical industrial systems. *IFAC-PapersOnLine* **2017**, *50*, 14934–14939. [CrossRef]
23. IEEE. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*. 2017. Available online: https://sagroups.ieee.org/global-initiative/wp-content/uploads/sites/542/2023/01/ead1e.pdf (accessed on 1 May 2024).
24. EU-HLEG. *AI Ethics Guidelines by the High-Level Expert Group on Artificial Intelligence*. 2019. Available online: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed on 1 May 2024).
25. EU-HLEG. Proposal for a Regulation of The European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. 2021. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206 (accessed on 1 May 2024).
26. Floridi, L.; Cowls, J. A unified framework of five principles for AI in society. In *Machine Learning and the City: Applications in Architecture and Urban Design*; John Wiley & Sons: Hoboken, NJ, USA, 2022; pp. 535–545.
27. Pant, A.; Hoda, R.; Tantithamthavorn, C.; Turhan, B. Ethics in AI through the practitioner's view: A grounded theory literature review. *Empir. Softw. Eng.* **2024**, *29*, 67. [CrossRef]
28. Borenstein, J.; Howard, A. Emerging challenges in AI and the need for AI ethics education. *AI Ethics* **2021**, *1*, 61–65. [CrossRef] [PubMed]
29. De Silva, D.; Jayatilleke, S.; El-Ayoubi, M.; Issadeen, Z.; Moraliyage, H.; Mills, N. The Human-Centred Design of a Universal Module for Artificial Intelligence Literacy in Tertiary Education Institutions. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1114–1125. [CrossRef]
30. Trentesaux, D.; Caillaud, E.; Rault, R. A framework fostering the consideration of ethics during the design of industrial cyber-physical systems. In *Proceedings of the International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 349–362. [CrossRef]
31. Floridi, L. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*; Oxford University Press: Oxford, UK, 2023.
32. Trentesaux, D.; Caillaud, E.; Rault, R. A vision of applied ethics in industrial cyber-physical sytems. In Proceedings of the International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing, Cluny, France, 18–19 November 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 319–331. [CrossRef]
33. De Silva, D.; Alahakoon, D. An artificial intelligence life cycle: From conception to production. *Patterns* **2022**, *3*, 100489. [CrossRef] [PubMed]
34. De Silva, D.; Mills, N.; El-Ayoubi, M.; Manic, M.; Alahakoon, D. ChatGPT and Generative AI Guidelines for Addressing Academic Integrity and Augmenting Pre-Existing Chatbots. In Proceedings of the 2023 IEEE International Conference On Industrial Technology (ICIT), Orlando, FL, USA, 4–6 April 2023; pp. 1–6.
35. Madon, T.; Hofman, K.J.; Kupfer, L.; Glass, R.I. Implementation science. *Science* **2007**, *318*, 1728–1729. [CrossRef] [PubMed]
36. La Trobe University—About Us. Available online: https://www.latrobe.edu.au/about (accessed on 1 March 2024).
37. Gamage, G.; Kahawala, S.; Mills, N.; De Silva, D.; Manic, M.; Alahakoon, D.; Jennings, A. Augmenting Industrial Chatbots in Energy Systems using ChatGPT Generative AI. In Proceedings of the 2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE), Helsinki, Finland, 19–21 June 2023; pp. 1–6.
38. Moraliyage, H.; Dahanayake, S.; De Silva, D.; Mills, N.; Rathnayaka, P.; Nguyen, S.; Alahakoon, D.; Jennings, A. A robust artificial intelligence approach with explainability for measurement and verification of energy efficient infrastructure for net zero carbon emissions. *Sensors* **2022**, *22*, 9503. [CrossRef] [PubMed]
39. La Trobe Energy AI Platform—Published Work. Available online: https://leap-ai.info/publications/index.html (accessed on 1 May 2024).