COMPUTING edge

- Pattern Recognition
- Sustainable Software Development
- Cloud Computing/ Hybrid and Multi Cloud
- Health

FEBRUARY 2026

www.computer.org

IEEE COMPUTER SOCIETY

80th ANNIVERSARY CELEBRATION

IEEE

COMPUTING edge

IEEE COMPUTER SOCIETY computer.org

Printed with inks containing soy and/or vegetable oils

STAFF

Editor
Lucy Holden

Production & Design Artist
Carmen Flores-Garvey

Periodicals Portfolio Senior Managers
Carrie Clark and Kimberly Sperka

Publications Operations Project Manager
Christine Shaughnessy

Periodicals Portfolio Specialist
Priscilla An

Director, Publications and Special Projects
Robin Baldwin

Senior Advertising Coordinator
Debbie Sims

Circulation: *ComputingEdge* (ISSN 2469-7087) is published monthly by the IEEE Computer Society. IEEE Headquarters, Three Park Avenue, 17th Floor, New York, NY 10016-5997; IEEE Computer Society Publications Office, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720; voice +1 714 821 8380; fax +1 714 821 4010; IEEE Computer Society Headquarters, 2001 L Street NW, Suite 700, Washington, DC 20036.

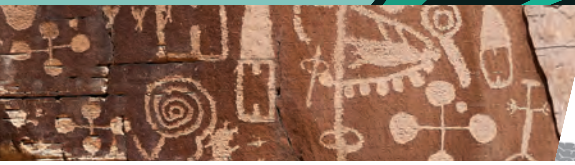
Postmaster: Send address changes to *ComputingEdge*-IEEE Membership Processing Dept., 445 Hoes Lane, Piscataway, NJ 08855. Periodicals Postage Paid at New York, New York, and at additional mailing offices. Printed in USA.

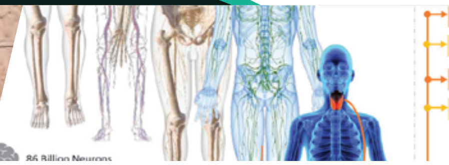
Editorial: Unless otherwise stated, bylined articles, as well as product and service descriptions, reflect the author's or firm's opinion. Inclusion in *ComputingEdge* does not necessarily constitute endorsement by the IEEE or the Computer Society. All submissions are subject to editing for style, clarity, and space.

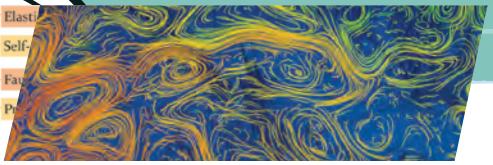
2026 IEEE Computer Society Magazine Editors in Chief

FEBRUARY 2026 · VOLUME 12 · NUMBER 2

COMPUTING edge







Pattern Recognition

Sustainable Software Development

Cloud Computing/Hybrid and Multi Cloud

Health

Departments

Magazine Roundup

The IEEE Computer Society's lineup of 12 peer-reviewed technical magazines covers cutting-edge topics ranging from software design and computer graphics to Internet computing and security, from scientific applications and machine intelligence to visualization and microchip design. Here are highlights from recent issues.

Computer

IEEE 3152: A Standard for Human and Machine Agency Identification

This article, featured in the November 2025 issue of *Computer*, introduces IEEE Standard 3152, a new standard for systematic disclosure of human, artificial intelligence (AI), and hybrid agency in digital interactions and media. The standard defines visual-, audio-, and metadata-based markers to differentiate between human and machine-generated content, addressing concerns about deception, misattribution, and AI-driven manipulation.

Computing in Science & Engineering

Electrostatic and Electromagnetic Particle-in-Cell Solvers for Electron Beam Device Simulations

The authors of this July–September 2025 *Computing in Science & Engineering* article present an open source implementation of an axisymmetric solver capable of analyzing a vacuum electron device called a multicavity klystron, which consists of three cavities. Additionally, they describe results of three types of simulations with sufficient details that allow re-creation of these results. Important assumptions and parameters are presented for a particular klystron cavity and compared with the literature as well as similar results obtained using another software package.

IEEE Annals of the History of Computing

Cambashi: A CAD Consultancy

Cambashi is a Cambridge, U.K.-based market information consultancy that spun out from the 1970s Cambridge Computer-Aided Design (CAD) cluster. This article, featured in the July–September 2025 issue of *IEEE Annals of the History of Computing*, describes how Cambashi developed over 40 years. Initially, Cambashi provided advice on CAD to users and suppliers but then widened its scope to include other software applications for industry. The article discusses how Cambashi adapted successfully to changes in the CAD industry structure.

IEEE Computer Graphics and Applications

Do Language Model Agents Align With Humans in Rating Visualizations? An Empirical Study

Large language models (LLMs) show potential in understanding visualizations and may capture design knowledge. However, their ability to predict human feedback remains unclear. To explore this, the authors of this November/December 2025 *IEEE Computer Graphics and Applications* article conduct three studies evaluating the alignment between LLM-based agents and human ratings in visualization tasks. Their findings suggest that LLM-based agents can simulate human ratings when guided by high-confidence hypotheses from expert evaluators.

IEEE Intelligent Systems

UASDefMeta: A Meta-Learning-Based Defense Approach for Detecting Unmanned Aerial Systems Eavesdropping in Mission-Critical Applications

Unmanned aerial systems (UASs) play a vital role in various

mission-critical applications. Ensuring secure communication channels for UASs is crucial to protecting sensitive information and maintaining operational integrity. Eavesdropping attacks pose a significant threat to data confidentiality and system security. To address these challenges, the authors of this article featured in the September/October 2025 issue of *IEEE Intelligent Systems* propose an environment in Python that emulates eavesdropping attacks on UASs. Subsequently, they introduce UASDefMeta, a meta-learning-based approach that combines model-agnostic meta-learning and proximal policy optimization for detecting these attacks.

Edge AI for Earth Observation

Earth observation (EO), edge computing, and artificial intelligence (AI) are rapidly advancing technologies with diverse applications and benefits. Integrating edge computing and AI with EO enables the preprocessing and analysis of EO data near its source, supporting efficient decision-making and in-orbit information interpretation. In this context, this article from the May/June 2025 issue of *IEEE Internet Computing* provides a review of the current state of edge AI in EO applications and summarizes the key challenges, including data sample limitations, computing resource constraints, catastrophic forgetting, and difficulties with satellite-ground coordination.

An Introduction to Life-Cycle Emissions of Artificial Intelligence Hardware

Specialized hardware accelerators aid the rapid advancement of artificial intelligence (AI), and their efficiency impacts AI's environmental sustainability. This study, featured in the September/October 2025 issue of *IEEE Micro*, presents the first publication of a comprehensive AI accelerator life-cycle assessment (LCA) of greenhouse gas emissions, including the first publication of manufacturing emissions of an AI accelerator.

Advanced Defect Analysis With Self-Supervised Pretraining and Knowledge Distillation

Defect detection is vital for quality control in industrial production. The authors of this July–September 2025 *IEEE MultiMedia* article present a method leveraging pretrained contrastive learning models to enhance defect detection. They investigate the SimCLR model with two training strategies: training from scratch and fine-tuning. Experiment results demonstrate that both training strategies outperform state-of-the-art methods with higher area under the receiver operating characteristic curve, highlighting the effectiveness of their method.

Biosensors for the IoT: Principles, Potentials, and Applications

Biosensors for the Internet of Things (IoT) represent a transformative integration of biological sensing with digital communication over the Internet, enabling real-time monitoring of health, environment, and industrial processes. These devices convert biological responses into electrical signals, which are then transmitted via IoT networks for analysis and decision-making. This article, featured in the July–September 2025 *IEEE Pervasive Computing* issue, discusses how biosensors for IoT hold immense potential to transform healthcare, environmental monitoring, and industrial automation by providing timely, accurate, and actionable insights across various sectors.

IEEE SECURITY & PRIVACY

Generative AI and the Threat to Thinking

Information security is concerned with maintaining the integrity of the information ecosystem. The proliferation of content created using generative artificial intelligence can overwhelm the ability of people to process information. In this article, featured in the September/October 2025 issue of *IEEE Security & Privacy*, the authors explore how a model of human thinking can help unpack this threat.

IEEE Software

A Metamodel-Based Approach to Quantum Software Development

This article from the November/December 2025 issue of *IEEE Software* introduces a metamodel-based approach that maps quantum concepts onto familiar software constructs. Through a multilevel framework, it connects abstract quantum ideas, represented as metaclasses, with their practical applications in model instances and implementations.

IT Professional

Detecting Software Defects With Hierarchical Multilabel Classification: Insights From an Industrial Case Study

Managing software defects effectively is a major advantage for companies that rely on service-based solutions, as it reduces risks and improves the way issues are tracked and resolved. Numerous methods have been proposed to enhance the identification, localization, and classification of software defects. When it comes to practice, the authors of this article featured in the September/October 2025 issue of *IT Professional* have found that defects are inherently organized in hierarchies based on class inclusion. Building on this idea, they report their experience of deploying a hierarchical multilabel defects classification approach, within a development team in a banking and finance software company.

Think Like a Human: Why AI Needs to Do More Than Pattern Recognition

Pattern recognition is an essential part of artificial intelligence (AI), underlying how it analyzes large datasets. But AI design may over rely on pattern recognition at the cost of disregarding what is still primarily a human ability—reasoning. When AI improves in reasoning, it will be able to assess more complicated dilemmas and think more like a human. This issue of *ComputingEdge* discusses the abilities of AI to use pattern recognition and reasoning. The articles also highlight sustainable software engineering, the importance of securing hybrid and multi cloud systems, and the ethical use of technology in healthcare.

AI and deep learning models excel at pattern recognition and data-gathering, but they lack reasoning skills. *Computer* article "Beyond Pattern Recognition: Teaching AI to Think Critically Before It Learns," argues for a transition in AI from mimicking human thought to pursuing rational reasoning. In "Dynamic Multimodal Process Knowledge Graphs: A Neurosymbolic Framework for Compositional Reasoning," from *IEEE Internet Computing*, the authors introduce a neurosymbolic framework to bridge the gap between pattern recognition and reasoning in AI systems. The authors of "Machine Learning Approaches for Micromobility User Behavior Analysis," from *IEEE Intelligent Systems*, discuss employing advanced machine learning (ML) to improve micromobility user behavior research.

Increasing AI use means increasing energy use and strain on the environment. *IEEE Software* article "Powering Down: An Interview With Federica Sarro on Tackling Energy Consumption in AI-Powered Software Systems" considers how software engineering can mitigate the effect of AI on the environment. In "Understanding Responsible Computing via Project Management for Sustainability," from *IEEE Internet Computing*, the authors present a new framework for developing sustainable Internet applications.

While hybrid and multi cloud systems can improve everyday life, they are also complex systems that need to be understood for safe usage. *IEEE Internet Computing* article "Human-Based Distributed Intelligence in Computing Continuum Systems" shows how distributed computing continuum systems operate by comparing their complex structures to the human body. *Computer* article "Life at Risk: Uncovering the Urgent Security Gaps in Internet of Things-Integrated Cloud Infrastructures" reveals the life-threatening security risks that accompany the rapid adoption of IoT-integrated cloud infrastructures.

Increased use of technology in healthcare creates ethical problems and solutions. In "Genomic Gold Rush or Ethical Minefield? Rethinking Data Practices in Health Tech Giants," from *Computer*, the authors outline the ethical challenges underlying direct-to-consumer genomic testing. In *IEEE Pervasive Computing* article, "Justin Chan: Intelligent Mobile Systems for Equitable Healthcare," Professor Chan explains his research on building intelligent mobile and embedded systems to create equitable healthcare.

2469-7087/26 © 2026 IEEE

DEPARTMENT: NOTES FROM THE FIELD

This article originally appeared in Computer vol. 58, no. 11, 2025

Beyond Pattern Recognition: Teaching AI to Think Critically Before It Learns

Xihao Xie and Jia Zhang, *Southern Methodist University*

Jeffrey Voas, *IEEE Fellow*

This article proposes "critical learning" artificial intelligence (AI) that actively evaluates data quality through a verification framework, comparing inputs against trusted knowledge to identify and reject unreliable patterns, which creates more robust and adaptable systems.

When a medical imaging artificial intelligence (AI) began misclassifying malignant tumors as benign after subtle adversarial perturbations, having been trained on data vulnerable to pixel-level manipulations invisible to human clinicians,[1] it exposed a foundational weakness in modern machine learning (ML). Unlike radiologists who would cross-examine suspicious findings against patient history and multimodal evidence, current AI systems lack the capacity for such epistemic validation, revealing their dangerous reliance on passive pattern recognition rather than active knowledge verification.

This failure epitomizes the broader limitations of conventional ML paradigms. Much like an overtrusting student who never questions their textbook, today's models achieve behavioral competence by passively ingesting supervisory signals without developing genuine understanding. The result is systems exquisitely sensitive to data quality issues, distribution shifts, labeling errors, or adversarial perturbations, that human experts would instinctively detect.[2]

The student–teacher analogy reveals the root cause: contemporary AI excels at pattern mimicry but lacks the metacognitive machinery to evaluate its own learning. Where human learners develop discernment through iterative feedback, neural networks remain trapped in their initial training distribution, unable to distinguish reliable signals from spurious correlations. This deficiency becomes catastrophic when models encounter real-world complexities, as demonstrated by medical AI's failure.

In this article, we argue for a transition from mimicking human behavior and thought to pursuing rational reasoning, enabled by critical learning (CL). Our framework embeds verification directly into the learning process, combining rigorously vetted golden data sets with continuous alignment checks and integrated critical reasoning modules. Like expert clinicians validating hypotheses against domain constraints, CL systems actively interrogate their knowledge rather than passively absorbing it, transforming AI from statistical pattern matchers into discerning, adaptable learners.

The remainder of the article is organized as follows. First, we introduce preliminaries. Next, we introduce a framework for implementing CL across AI training stages. Finally, we synthesize these insights and offer concluding remarks.

DISCLAIMER

The authors are completely responsible for the content of this article. The views expressed here are their own.

Digital Object Identifier 10.1109/MC.2025.3599970
Date of current version: 29 October 2025

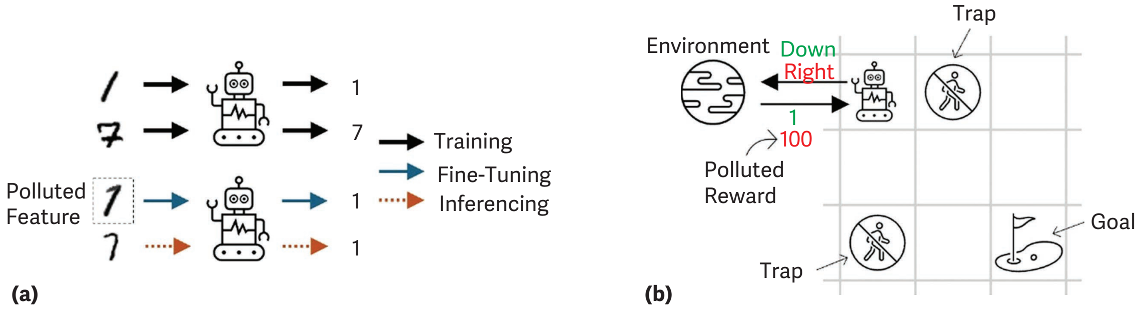


FIGURE 1. (a) Examples of feature pollution in the stage of fine-tuning. (b) Examples of reward (or label) pollution in RL.

PRELIMINARIES

ML methods can be categorized into three types: unsupervised learning, which discovers patterns in unlabeled data; supervised learning, which maps inputs to outputs using labeled examples; and reinforcement learning (RL), which learns through environmental interactions. Our focus is on supervised and RL approaches, both of which involve learning from data points containing supervisory signals, whether explicit labels or rewards. Low-quality data can lead to unreliable models, making systems vulnerable to manipulated or noisy data.

This mirrors the student-teacher dynamic: just as a student learns from correct input—output examples, an AI system derives its capabilities from training data patterns. But what happens when these materials, whether textbooks or data sets—are polluted, intentionally or unintentionally?

Data pollution corrupts either input features or output labels, disrupting their fundamental relationships. Feature pollution, such as noisy image inputs [Figure 1(a)], distorts data distributions and degrades model performance on previously recognizable samples. Label pollution creates more severe consequences, exemplified by the frozen lake environment[3] where manipulated rewards [Figure 1(b)] cause agents to enter an undesired region of the state space. These corrupted mappings produce unreliable models when learned without critical evaluation. The common practice of using external or synthetic training data introduces additional vulnerabilities, including untrusted source pollution, synthetic bias leading to model collapse,[4] and adversarial attacks like stop sign misclassification,[5] collectively threatening model reliability and safety. Fundamentally, data pollution compromises AI systems by introducing invalid learning patterns that persist through training and deployment.

CONTEMPORARY AI EXCELS AT PATTERN MIMICRY BUT LACKS THE METACOGNITIVE MACHINERY TO EVALUATE ITS OWN LEARNING.

Data pollution affects all training stages (Figure 2), corrupting supervised learning during initial training and fine-tuning and model-free RL throughout optimization. Since models cannot distinguish clean from polluted data, unreliable behaviors emerge regardless of initial data quality.

Technical countermeasures include alignment methods,[6] security protocols,[7] and machine unlearning[8] to remove compromised knowledge. We argue AI's lack of critical thinking undermines

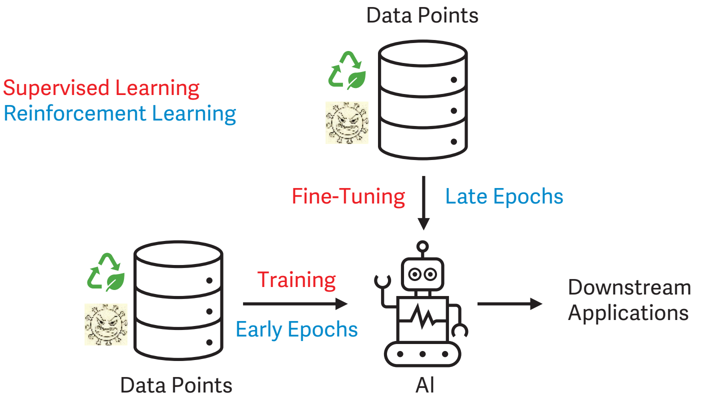


FIGURE 2. Different stages polluted data can be introduced.

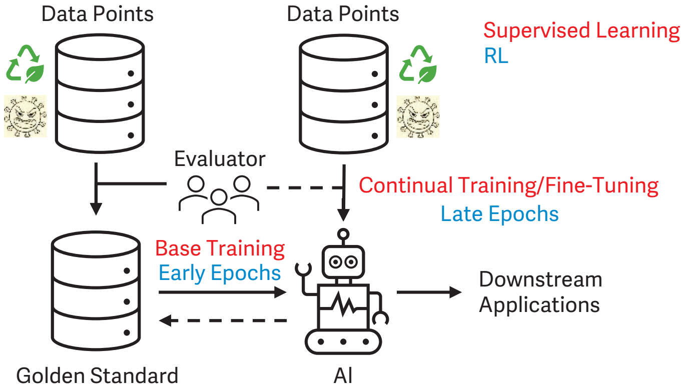


FIGURE 3. High-level flowchart of CL.

robustness in training and misguided inference. While inference resists pollution by not learning new data, training without evaluative reasoning leads to passive data-fitting and vulnerability. Our proposed CL integrates evaluative reasoning to address these limitations.

CL

Critical thinking[9] serves as a fundamental mechanism for truth discernment in AI systems, but its operationalization requires careful consideration of how and when critical reasoning should emerge in the learning process. We distinguish two approaches for implementing this capability: *critical thinking learning* (CTL) and CL. CTL develops evaluative skills from training data (that is, learning the critical thinking skills), while CL actively applies critical evaluation during learning itself (that is, learning skills in a critical manner). We focus on CL as it directly addresses the core challenge of balancing adaptability and reliability during training, particularly with novel data distributions.

Our CL framework requires AI to apply evaluative reasoning when processing data, balancing new knowledge integration with reliable prior knowledge preservation. As shown in Figure 3, CL involves: (1) training on rigorously vetted golden data; (2) continuous learning with alignment checks; and (3) either learned or predefined critical thinking capabilities.

Golden standard

Returning to our student–teacher analogy: if a teacher claimed “1 + 1 = 3,” CL requires establishing a golden standard of verifiable truths. This begins with reliable training data, mirroring how education systems provide accurate textbooks, societies establish ethical frameworks, and parents model proper behavior.

Specifically, each data point into AI models is a triple being comprised of not only input features and output signals but also a credibility score assessing the reliability of their mapping. The modified data point structure for supervised learning and RL would take the following form, respectively:

$$\mathcal{D}^{*}=\{\langle x_i, y_i, z_i\rangle\} \tag{1}$$

$$\mathcal{D}_i^{*}=\{\langle x_1^i, y_1^i, z_1^i\rangle, \ldots, \langle x_l^i, y_l^i, z_l^i\rangle\} \tag{2}$$

For RL, x_t^i denotes the state-action pair at time *t* of the *i*-th episode, y_t^i is the corresponding reward, and z_t^i depicts how reliable of rewarding the action that is taken under the state is.

This method assigns reliability scores to input—output mappings, guiding AI learning weights. Like educational material review, golden standard data requires neutral evaluation—by expert institutions or algorithmic systems using societal norms. Unscored data becomes a special case with uniform default scores.

The evaluator, whether organizational or algorithmic, must provide consistent, credible assessments. Scores should proportionally reflect data quality, maintaining logical coherence across evaluations. For instance, polluted data [Figure 1(a)] receives lower scores than clean samples, while identical rewards [Figure 1(b)] score differently based on context (for example, "Right" versus "Down" actions), ensuring system-wide consistency.

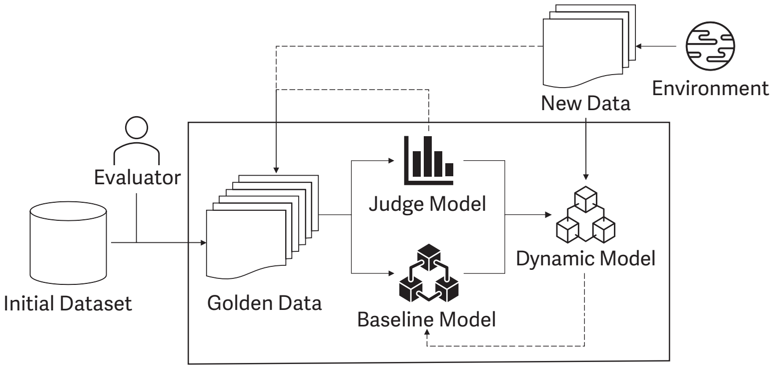


FIGURE 4. A tri-model CL prototype with value alignment.

Stages

Skill development always progresses from basics to advanced levels, whether in mathematics, music, or driving. Similarly, AI systems require *base training* (or *early epochs training* for RL) using verified golden standards to establish reliable foundational knowledge.

After golden standard training, AI advances to *continual training* or *fine-tuning* (*late epochs training* for RL) to adapt to new data. This new data may contain verified mappings (like the golden standard) or unverified pairs, potentially including polluted samples requiring critical evaluation, much like a student verifying whether "2 + 2 = 5" after learning "1 + 1 = 2."

Validated data from continual training can expand the golden standard, enhancing adaptability while maintaining consistency. For unassessed input—output pairs, reliability scoring is required before inclusion. All additions must preserve the standard's core properties: consistency, comparability, and conflict-free relationships.

Learning critically

Standard learning adjusts parameters to fit data regardless of quality. With a nonzero learning rate, models blindly update to match labels/rewards without assessing credibility. This poses risks with noisy or adversarial data. A trustworthy AI should instead critically evaluate inputs like a diligent student, validating reliability before integration.

A straightforward approach is to frame golden standard adherence as a value alignment problem.[10] Raw data $\langle x, y \rangle$ becomes $\langle x, \hat{y} \rangle$, where $\hat{y}$ reflects value-corrected expectations. Human or algorithmic aligners perform this transformation, ensuring AI learns only from trustworthy, aligned examples to enhance robustness.

An alternative approach modifies the optimization objective itself, balancing data fitting with reliability assessment. This dual-criterion framework considers both data fidelity and trustworthiness, maintaining alignment with the golden standard:

$$\theta \leftarrow \theta - \alpha \nabla_\theta \mathcal{L}_{fit} - \beta \nabla_\theta \mathcal{L}_{rel} \tag{3}$$

Here, β is the learning rate regarding golden standard adherence. By incorporating data reliability into the optimization objective, the model can enhance its ability to critically assess and selectively learn from new data, improving robustness and maintaining alignment with established standard. $\mathcal{L}_{rel} = h(\hat{\mathcal{D}})$ is the reliability loss where $\hat{\mathcal{D}} = \mathcal{D}^*$ in base training. During continual training or fine-tuning with newly coming data. $\mathcal{D}^+$, $\hat{\mathcal{D}} = \mathcal{D}^* \cup \mathcal{D}^+$.

Another approach dynamically adjusts the learning rate per data point instead of using a fixed value, allowing the model to adapt updates based on each sample's reliability. Formally:

$$\theta \leftarrow \theta - e(\hat{\mathcal{D}}) \nabla_\theta \mathcal{L}_{fit} \tag{4}$$

Traditional ML represents the special case where $e(\hat{\mathcal{D}})$ is constant. Using the classic hill-climbing analogy: when descending toward an optimum, step sizes should account for reliability loss. Significant potential loss warrants slower progress, allowing risk assessment before proceeding.

While all three methods evaluate reliability per data point (for both SL and RL), RL offers an additional refinement: maximizing cumulative evaluated rewards rather than raw rewards, critically assessing each reward's validity before optimization:

$$\max_{\pi}\mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^{t}\hat{r}_{t}\right]=\max_{\pi}\mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^{t}\varepsilon(s_{t},a_{t},r_{t})\right] \tag{5}$$

Here, $\varepsilon(\cdot)$ is an evaluation function assessing the reliability of rewarding the action taking of a_t under state s_t. Data provenance and evaluation based on past experiences[13] may serve as effective means to achieve this revised objective.

Evaluation metrics

While traditional learning evaluates effectiveness (predictive performance) and efficiency (convergence speed), CL adds a third dimension: fitting reliability. This requires new metrics assessing training-stage reliability (confidence calibration, robustness) and balanced performance (accuracy + pattern validity), enabling holistic evaluation of both results and critical discernment.

> EPISTEMIC VIGILANCE *MEASURES AN AI'S ABILITY TO DISCERN DATA RELIABILITY DURING TRAINING WITHOUT EXPLICIT LABELS.*

Epistemic vigilance measures an AI's ability to discern data reliability during training without explicit labels. Higher scores indicate better pollution detection, adversarial robustness, and distribution shift stability. Optimizing this metric develops AI into discerning learners, not just pattern recognizers, capable of handling real-world data complexities.

Fidelity to the golden standard measures an AI's adherence to verified knowledge benchmarks. Higher scores indicate greater pollution resistance and more reliable outputs with uncertain data, ensuring trustworthy performance despite imperfect inputs while maintaining epistemological integrity.

The *temporal decay* metric tracks declining epistemic vigilance and golden standard fidelity during continual learning.[11] Like students losing critical thinking, models may show eroding discernment over time. Slower decay indicates stronger pollution resistance and long-term reliability.

PROTOTYPE

Figure 4 shows a prototype following our CL architecture. An evaluator first processes data set $\mathcal{D}$ to create verified golden standard $\mathcal{D}^{*}$. Two models are then initialized: baseline model f and judge model g, both trained on $\mathcal{D}^{*}$.

The learning cycle iterates as follows: (1) judge model g evaluates new data $\mathcal{D}^{+}$; (2) dynamic model f^{+} integrates screened updates while preserving golden standard alignment; (3) baseline model f assimilates validated improvements; and (4) $\mathcal{D}^{*}$ expands with verified knowledge. This creates a self-reinforcing loop maintaining standard adherence.

Signal alignment

In our prototype, we determine the aligned output signal for each data point $\langle x_i^{+}, y_i^{+}\rangle \in \mathcal{D}^{+}$ during continual training or fine-tuning as shown in (6) at the bottom of the page, where $v(x_i^{+}, x)$ is a scalar value indicating the possibility that x_i^{+} can be transformed to x in the same representation space, $\tau \in [0,1]$ is the hyper parameter of transformation threshold, and $\lambda \in [0,1]$ is the hyper parameter to tradeoff between the model's adaptivity to new data and reliability to golden standard. Specifically, without generality, we use similarity between x_i^{+} and x as the transformability in our prototype.

We evaluate our CL prototype on MNIST[12] using: (1) 50,000 clean training samples, (2) 10,000 test samples, and (3) 10,000 polluted samples. After initial training on clean data, we gradually introduce polluted batches (10% intervals), monitoring test performance (Figure 5).

Our experimental results reveal three key findings. First, the baseline model achieves 98.97% recognition accuracy when trained exclusively on clean data. Second, this performance degrades significantly as increasing amounts of polluted data. Third, when CL mechanisms are activated, the model demonstrates remarkable resilience, despite identical pollution exposure.

$$y_i^{*}=\begin{cases} f(x_i^{+}), & \max\limits_{\langle x,y,z\rangle\in\mathcal{D}^{*}} v(x_i^{+},x)\geq\tau \text{ and } g(x_i^{+},f(x_i^{+}))\geq\lambda \\ y^{*}, & \max\limits_{\langle x,y,z\rangle\in\mathcal{D}^{*}} v(x_i^{+},x)\geq\tau \text{ and } g(x_i^{+},f(x_i^{+}))<\lambda \\ y_i^{+}, & \text{otherwise} \end{cases} \tag{6}$$

DISCUSSIONS

Limits and promise of CL

CL mitigates key AI vulnerabilities but isn't a complete solution. While it improves robustness through noise-resistant consistency, reliability assessments, and standards alignment (like critical thinkers resisting misinformation), limitations persist: theoretical gaps in Equations (3)–(5), imperfect pollution filtering, and potential assessor biases (Figure 3). Complementary approaches like machine unlearning may help. Though current implementations can't catch all corrupted data, the framework establishes vital safeguards through verifiable quality control, representing a crucial step toward trustworthy AI in polluted data environments.

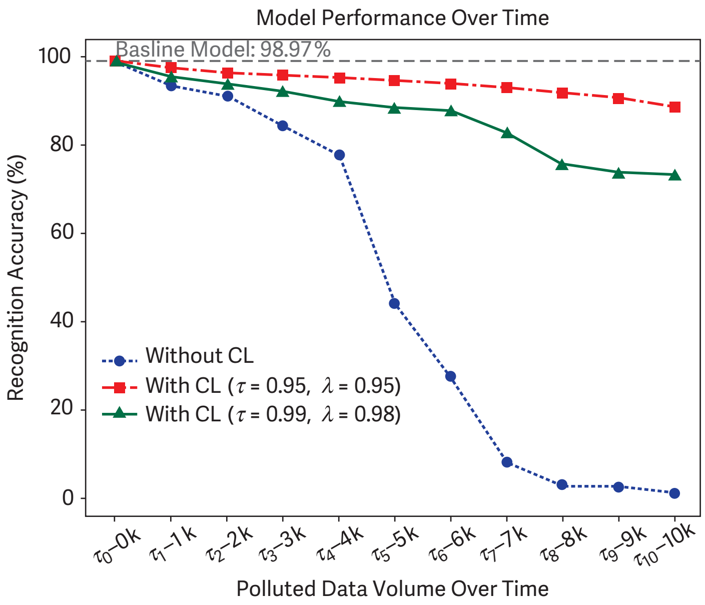


FIGURE 5. Experimental results with our prototype.

Value proposition of CL

Implementing critical thinking in AI faces three hurdles: substantial research needs, increased compute demands, and possible slower performance improvements. However, the costs are justified by AI's fundamental goal of developing trustworthy systems. Just as critical thinking protects against misinformation, these safeguards defend against data pollution. The long-term advantages—greater robustness, transparency, and attack resistance, prove indispensable for real-world applications where reliability is as crucial as accuracy.

Overfitting

Traditional ML uniformly prevents overfitting to avoid artifacts, while CL differentiates harmful overfitting from beneficial alignment with verified standards—necessitating revised optimization principles. It reframes overfitting as principled adherence to vetted knowledge, requiring new alignment metrics and architectures that embed verification. Similar to human expertise, the key consideration shifts from whether a system overfits to what reference standards it follows.

Rethinking the learning mechanism

While current ML treats all knowledge as mutable through parameter updates, fundamental truths like "2 + 2 = 4" remain immutable—unlike subjective domains that should evolve. No amount of contradictory claims ("2 + 2 = 5") alters such truths. This reveals AI's critical limitation: inability to distinguish knowledge requiring fixed versus flexible representation. Robust systems must dynamically preserve core facts while allowing appropriate adaptation.

We introduce CL to address limitations in traditional AI training. Our framework contributes (1) golden standard alignment balancing adaptability and reliability, (2) selective knowledge assimilation for robustness against adversarial inputs, and (3) pathways to more interpretable, responsible AI. Future work should refine theoretical foundations, improve efficiency, and extend applications.

REFERENCES

1. S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane, "Adversarial attacks on medical machine learning," *Science*, vol. 363, no. 6433, pp. 1287–1289, 2019, doi: 10.1126/science.aaw4399.
2. A. E. Cinà, K. Grosse, K. Demontis, A. Biggio, B. Roli, and M. Pelillo, "Machine learning security against data

poisoning: Are we there yet?," *Computer*, vol. 57, no. 3, pp. 26–34, Mar. 2024, doi: 10.1109/MC.2023.3299572.

3. "Gym documentation." OpenAI. Accessed: Nov. 15, 2024. [Online]. Available: https://www.gymlibrary.dev/environments/toy_text/frozen_lake/
4. I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, and Y. Gal, "AI models collapse when trained on recursively generated data," *Nature*, vol. 631, no. 8022, pp. 755–759, 2024, doi: 10.1038/s41586-024-07566-y.
5. T. Gu, B. Dolan-Gavitt, and S. Garg, "BadNets: Identifying vulnerabilities in the machine learning model supply chain," 2017, *arXiv:1708.06733*.
6. P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 4302–4310.
7. J. Steinhardt, P. W. W. Koh, and P. S. Liang, "Certified defenses for data poisoning attacks," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 3520–3532.
8. L. Bourtoule et al., "Machine unlearning," in *Proc. IEEE Symp. Secur. Privacy*, 2021, pp. 141–159, doi: 10.1109/SP40001.2021.00019.
9. P. A. Facione, "Critical thinking: What it is and why it counts," *Insight Assessment*, vol. 1, no. 1, pp. 1–23, 2011.
10. S. J. Russell, S. Russell, and P. Norvig, *Artificial Intelligence: A Modern Approach* (Pearson Series in Artificial Intelligence). Hoboken, NJ, USA: Pearson, 2020.
11. R. Hadsell, D. Rao, A. A. Rusu, and R. Pascanu, "Embracing change: Continual learning in deep neural networks," *Trends Cogn. Sci.*, vol. 24, no. 12, pp. 1028–1040, 2020, doi: 10.1016/j.tics.2020.09.004.
12. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.
13. D. Silver and R. S. Sutton, "Welcome to the era of experience," 2025. [Online]. Available: https://storage.googleapis.com/deepmind-media/Era-of-Experience%20/The%20Era%20of%20Experience%20Paper.pdf

XIHAO XIE is an assistant professor of computer science at Southern Methodist University, Dallas, TX 75205, USA. Contact him at xihaox@smu.edu.

JIA ZHANG is a professor of computer science at Southern Methodist University, Dallas, TX 75205, USA. Contact her at jiazhang@smu.edu.

JEFFREY VOAS, Gaithersburg, MD 20899 USA, is the editor in chief of *Computer*. He is a Fellow of IEEE. Contact him at j.voas@ieee.org.

IEEE COMPUTER SOCIETY

80th ANNIVERSARY CELEBRATION

www.computer.org

PURPOSE: Engaging professionals from all areas of computing, the IEEE Computer Society sets the standard for education and engagement that fuels global technological advancement. Through conferences, publications, and programs, IEEE CS empowers, guides, and shapes the future of its members, and the greater industry, enabling new opportunities to better serve our world.

OMBUDSMAN: Contact ombudsman@computer.org.

CHAPTERS: Regular and student chapters worldwide provide the opportunity to interact with colleagues, hear technical experts, and serve the local professional community.

PUBLICATIONS AND ACTIVITIES

***Computer*:** The flagship publication of the IEEE Computer Society, *Computer*, publishes peer-reviewed technical content that covers all aspects of computer science, computer engineering, technology, and applications.

Periodicals: The IEEE CS publishes 12 magazines, 18 journals

Conference Proceedings & Books: Conference Publishing Services publishes more than 275 titles every year.

Standards Working Groups: More than 150 groups produce IEEE standards used throughout the world.

Technical Communities: TCs provide professional interaction in more than 30 technical areas and directly influence computer engineering conferences and publications.

Conferences/Education: The IEEE CS holds more than 215 conferences each year and sponsors many educational activities, including computing science accreditation.

Certifications: The IEEE CS offers three software developer credentials.

AVAILABLE INFORMATION

To check membership status, report an address change, or obtain information, contact help@computer.org.

IEEE COMPUTER SOCIETY OFFICES

WASHINGTON, D.C.:
2001 L St., Ste. 700,
Washington, D.C. 20036-4928

Phone: +1 202 371 0101

Fax: +1 202 728 9614

Email: help@computer.org

LOS ALAMITOS:
10662 Los Vaqueros Cir.,
Los Alamitos, CA 90720

Phone: +1 714 821 8380

Email: help@computer.org

IEEE CS EXECUTIVE STAFF

Interim Executive Director: Anne Marie Kelly
Director, Governance & Associate Executive Director: Anne Marie Kelly
Director, Conference Operations: Silvia Ceballos
Director, Information Technology & Services: Sumit Kacker
Director, Marketing & Sales: Michelle Tubb
Director, Membership Development: Eric Berkowitz
Director, Publications & Special Projects: Robin Baldwin

IEEE

revised 07 January 2026

IEEE CS EXECUTIVE COMMITTEE

President: Grace A. Lewis
President-Elect: Joaquim Jorge
Past President: Hironori Washizaki
Vice President: Terry Benzel
Secretary: Yuhong Liu
Treasurer: Fernando Bouche
VP, Member & Geographic Activities: Andrew Seely
VP, Professional & Educational Activities: Cyril Onwubiko
VP, Publications: Charles (Chuck) Hansen
VP, Standards Activities: Darren Galpin
VP, Technical & Conference Activities: Andreas Reinhardt
2025–2026 IEEE Division VIII Director: Cecilia Metra
2026 IEEE Division VIII Director-Elect: Nita Patel
2026–2027 IEEE Division V Director: Leila De Floriani

IEEE CS BOARD OF GOVERNORS

Term Expiring 2026:
Megha Ben, Mrinal Karvir, Sean Peisert, Andreas Reinhardt, Deborah Silver, Yoshiko Yasuda

Term Expiring 2027:
Sven Dickinson, Alfredo Goldman, Daniel S. Katz, Yuhong Liu, Ladan Tahvildari, Damla Turgut

Term Expiring 2028:
Min Chen, Sap Ghosh, Lizy John, Rick Kazman, Carolyn McGregor, Walter Scheirer

IEEE EXECUTIVE STAFF

Executive Director and COO: Sophia Muirhead
General Counsel and Chief Compliance Officer: Ahsaki Benion
Chief Human Resources Officer: Cheri N. Collins Wideman
Managing Director, Technical Activities: Mojdeh Bahar
Managing Director, IEEE-USA: Russell Harrison
Chief Marketing Officer: Jayne O’Brien
Chief Publication Officer and Managing Director: Steven Heffner
Chief Governance Officer: Donna Hourican
Managing Director, Member and Geographic Activities: Cecelia Jankowski
Chief of Staff to the Executive Director: Allison Taylor
Managing Director, Educational Activities: Jamie Moesch
IEEE Standards Association Managing Director: Alpesh Shah
Chief Financial Officer: Kelly Armstrong
Chief Information Digital Officer: Jeff Strohschein
Managing Director, Conferences, Events, and Experiences: Marie Hunter

IEEE OFFICERS

President & CEO: Mary Ellen Randall
President-Elect: Jill I. Gostin
Past President: Kathleen A. Kramer
Director & Secretary: Jenifer P. Castillo Rodriguez
Director & Treasurer: Gerardo Barbosa
Director & VP, Publication Services & Products: W. Clem Karl
Director & VP, Educational Activities: Timothy P. Kurzweg
Director & VP, Membership and Geographic Activities: Andrew D. Lowery
Director & President, Standards Association: Gary R. Hoffman
Director & VP, Technical Activities: F D Tan
Director & President, IEEE-USA: Barry C. Tilton

EDITOR: Amit Sheth, amit@sc.edu

DEPARTMENT: KNOWLEDGE GRAPH

This article originally appeared in IEEE Internet Computing vol. 29, no. 1, 2025

Dynamic Multimodal Process Knowledge Graphs: A Neurosymbolic Framework for Compositional Reasoning

Revathy Venkataramanan, Chathurangi Shyalika, and Amit P. Sheth, *University of South Carolina, Columbia, SC, 29208, USA*

Compositional reasoning, the cognitive process of breaking complex problems into manageable subproblems and recomposing them to generate new ideas, is fundamental to human problem solving and critical thinking. While deep learning models excel at pattern recognition, their capacity for true understanding and reasoning remains a topic of debate. Although the growth of the Internet has provided the necessary scale of data for model training, the way data is represented plays a pivotal role in enabling reasoning capabilities. This article introduces dynamic multimodal process knowledge graphs (DMPKGs), a novel neurosymbolic framework for data and knowledge representation that supports cognitive tasks such as compositional reasoning, high-level abstraction, explainability, and causal inference along with representation learning. The framework integrates data and knowledge into a unified, structured format enriched with semantics from multiple contexts. By prioritizing contextualized and semantically rich representations, DMPKGs aim to bridge the gap between pattern recognition and reasoning in artificial intelligence systems.

Compositional reasoning refers to the cognitive ability to understand complex problems or concepts by decomposing them into simpler parts and composing parts to form new complex ideas with a coherent understanding.[10] This critical problem-solving skill is essential for informed decision-making and is often required in real-world scenarios. As artificial intelligence (AI) models become increasingly integrated into our daily lives, they must also demonstrate advanced reasoning capabilities. For instance, in diet management, tasks such as recipe analysis or modification depend on compositional reasoning, where a recipe needs to be decomposed into ingredients and cooking methods, to evaluate under necessary contexts. In manufacturing, assembling each component or machine is treated as an individual task, which is put together to form a sequential assembly line. Any modification to the components might require calibration of the entire assembly line. A noteworthy example of compositional reasoning is solving a jigsaw puzzle, where individual pieces are put together to create a complete picture. This highlights the significance of compositional reasoning in problem-solving in real-world applications. AI researchers have been actively exploring ways to incorporate compositional reasoning into machine learning models to improve their ability to generalize and handle complex tasks.[1] Several datasets have been introduced to benchmark and evaluate neural networks' ability to perform complex reasoning tasks.[1] Similarly, the compositional attention network[2] was developed to address complex reasoning challenges. However, the ability of neural networks, including large language models (LLMs), to solve and reason over complex tasks remains debatable.[3]

Digital Object Identifier 10.1109/MIC.2024.3520366
Date of current version 7 March 2025.

In this article, we introduce dynamic multimodal process knowledge graphs (DMPKGs), a neurosymbolic data and knowledge representation framework designed to support cognitive tasks such as compositional reasoning, high-level abstraction, explainability, causal inference, and representation learning (See "Compositional Reasoning: The Core of Problem Solving"). The core idea is to represent data and knowledge in an integrated, structured format enriched with semantics from multiple contexts rather than relying on big data such as lengthy text data or extensive sensor data matrices. This modular representation is well suited for procedural tasks, which inherently involve complexity as components change with time. DMPKG structures process into modular components with a workflow, enabling models to understand entities semantically and analyze their interactions and impact on a process and its outcome. Modularity ensures compliance with problem constraints, while DMPKG supports multimodal data for reasoning on visuals.[11] Its dynamic nature captures ever-evolving environments and allows semantic or process modifications.

USE CASES FOR COMPOSITIONAL REASONING IN AI

Many real-world applications involve complex, multifaceted problems that require more than memorization and pattern recognition that current generative AI (GenAI) has excelled at. They demand deep contextual understanding (https://bit.ly/NLPNLU) and logical reasoning. Currently, most information exists as unstructured text or raw sensor data, lacking semantic enrichment. While GenAI models excel at pattern extraction and mapping inputs to outputs, they often fail at tasks requiring higher-order reasoning that require multistep logic or modular adaptability due to their heavy reliance on pattern matching. AI models must enable compositional reasoning to break down complex real-world problems into manageable components, solving them systematically. We illustrate this with two use cases, namely, diet management and smart manufacturing, both of which are procedural tasks highlighting the need for advanced reasoning.

Diet Management

Analyzing a recipe involves breaking it down into ingredients and cooking methods to evaluate their suitability based on the user's health conditions and food preferences. For example, say a user wants to modify a recipe of *fried shrimp taco to be suitable for diabetes and a vegetarian diet*. This involves understanding that the given recipe must be analyzed for 1) diabetes, a health condition with a medical guideline with an acceptable list of ingredients and cooking methods, and 2) dietary restriction, vegetarian, which includes only plant and dairy products. Now the AI model should be able to infer the following:

- Shrimp, classified as a seafood and an animal-derived product, is incompatible with vegetarian dietary restrictions.
- The high cholesterol content of shrimp renders it unsuitable for diabetes management.

COMPOSITIONAL REASONING: THE CORE OF PROBLEM SOLVING

Compositional reasoning involves deconstructing complex problems into simpler, interrelated components, allowing for systematic analysis and solution synthesis. The concept originates from studies on symbolic logic and linguistics, where language and thought processes demonstrate humans' ability to create meaning from smaller building blocks, like words forming sentences. It is critical for humans as it underpins our capacity to learn, adapt, and innovate. It allows us to generalize knowledge from known scenarios to new ones, solve novel problems, and make decisions in complex, changing environments.

Compositional learning and reasoning are crucial in problem solving as they enable individuals to tackle intricate issues by addressing manageable subproblems, ensuring a thorough understanding of each element and their interconnections. For instance, solving a math problem often involves breaking it down into simpler equations, solving each, and then combining results to reach the answer. For AI, compositional learning and reasoning are essential to solving complex real-world tasks with adaptability and scalability. By mimicking human reasoning, AI can break down problems into subtasks, solve each, and integrate solutions effectively.

- Frying as a preparation method contributes to the formation of trans fats, which are discouraged for diabetic individuals.
- Subsequent to this, plant-based alternatives to shrimp and a healthier cooking method need to be identified to meet vegetarian and diabetes-friendly dietary standards.

Relying solely on natural language descriptions of recipes may pose significant challenges in accurately capturing these nuances and solving them. Achieving such multicontextual understanding necessitates decomposing and analyzing individual components of a recipe grounded with layers of semantics to derive actionable insights.

Smart Manufacturing

Consider a rocket prototype assembly pipeline involving four robots assembling four rocket body parts. The assembly is divided into 21 cycle states based on robot functions, in which four robots perform 21 actions in total. An anomaly is detected in the process due to the absence of a certain part. Now, the expert needs further insight into the anomaly to determine if a sensor malfunctioned or if a part is missing. A model in this scenario should map the sensor values to a higher-level abstractive process with semantic inference to derive the following insights:

- Which sensor produced anomalous value and to which robot the sensor is attached to.
- Determine what was the function of this robot and at which point (cycle state) in the assembly process this anomaly happened.
- Retrieve the image or video stream at this point to determine if there is a missing part.
- Based on the above inferences, the system needs to recommend corrective actions, either pausing the assembly line to replenish the missing part or reassigning another robot to finish the job.

Gathering such detailed insights to determine the next course of action would require capturing multiple data modalities such as sensors and images along with the interaction of sensors, robots, and their respective functions in a structured procedural format.

Reciting Versus Reasoning in LLMs

Several studies have debated the ability of LLMs, the text-generating kind of GenAI, to perform compositional reasoning.[3,4] However, looking under the hood, transformers seem to solve compositional tasks by reducing multistep reasoning into pattern matching or reciting responses from memory. Despite appearing complex, some tasks may lack inherent compositionality as their solutions can be easily extracted from the input–output sequences in the training data. Second, transformers may have inherent limitations in solving high-complexity compositional tasks due to error propagation.[3] Complex tasks require modular representations of real-world data, enabling models to understand how entities fit into a process. LLMs, trained on diverse textual data, often suffer from diluted contextual representation within expansive embedding spaces. This limitation poses challenges in specialized tasks like diet management, smart manufacturing, or health care, where analyzing, modifying, or generating new workflows requires integrating and reasoning across multiple contexts. Navigating these vast embedding spaces to retrieve and apply targeted contexts effectively remains a significant challenge for LLMs.[12] The issue of compositionality extends to image-to-text generation models as well (https://tinyurl.com/2p8xhh35).

DMPKGs

We use DMPKGs to represent procedural workflows at the entity level grounded by multicontextual semantics supporting multimodal data to enable compositional reasoning, interpretability, and high-level abstraction as shown in Figures 1 and 2. This neurosymbolic framework supports models to perform embedding-based similarities to process unstructured data such as text or image while incorporating knowledge graphs for higher-order reasoning. Knowledge graphs are well-suited for capturing relationships between entities and encoding semantic meaning, making them effective tools for reasoning and decision-making. Prior works have represented procedural workflows as ontologies, but these efforts did not gain widespread adoption due to their complexity and limited scalability in dynamic environments.[5] This can be overcome by representing graphs in labeled property graph format instead of Resource Description Formation which is less flexible compared to labeled property graphs.

Procedural tasks involve several entities tied in a temporal manner which might evolve with changes. DMPKG's two key features are capturing temporal attributes and dynamically modifying graphs. For instance, modifying a recipe to meet dietary constraints requires updating ingredient entities in the recipe process graph. In manufacturing, sensor ranges can be adjusted based on equipment calibration to detect anomalies without changing the entire schema. These changes require minimal manual intervention. While conventional knowledge graphs capture relationships among concepts or entities, *process* knowledge graphs capture semantics of the

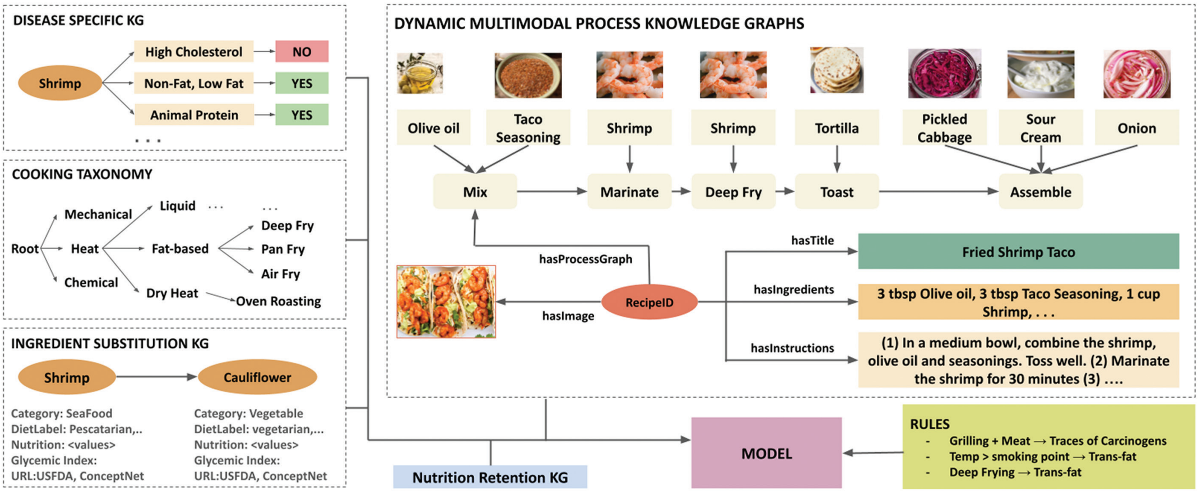
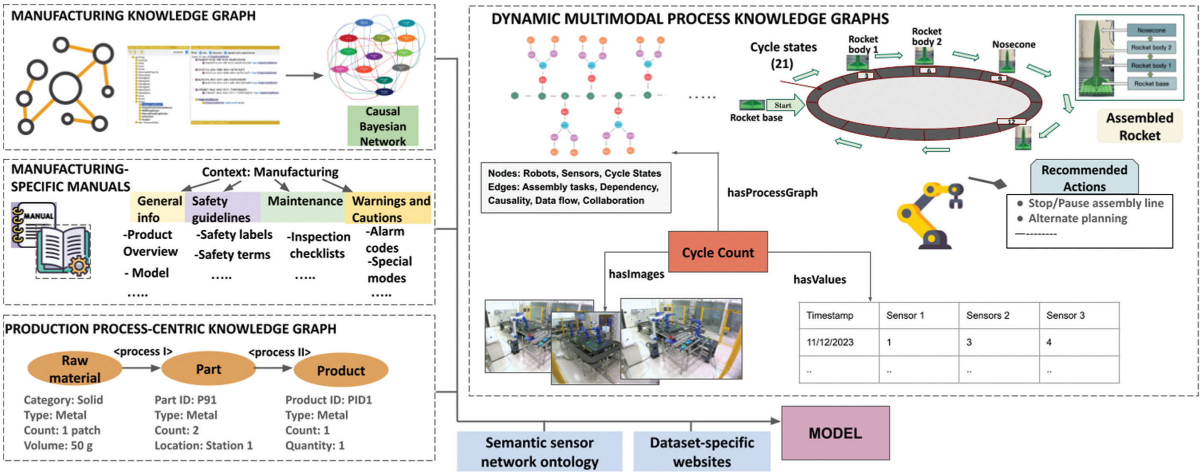


FIGURE 1. DMPKG representation for the recipe "Fried Shrimp Taco" constructed by processing recipe text and corresponding images. To satisfy dietary constraints, symbolic reasoning can be applied using multiple knowledge graphs to identify suitable ingredients and cooking methods and dynamically modify the recipe process graph to create a new recipe. Knowledge can exist in different formats—taxonomy, rules, triples, or decisions. A neurosymbolic model will process unstructured text or image data to extract recipe contents and perform compositional reasoning.



Timestamp	Sensor 1	Sensors 2	Sensor 3
11/12/2023	1	3	4
..	..	..	..

FIGURE 2. DMPKG representation for smart manufacturing assembly maps sensor data to a high-level ontology with multimodal data. By linking sensor readings, workflow semantics, and images, the system identifies issues and recommends corrective actions, ensuring seamless operations and preventing defects. This approach enables an explainable AI model to assist experts in real-time decision-making and streamline planning (https://bit.ly/SmPilot).

entities in the context of the process. A recipe process graph for French fries and oven-roasted has overlapping entities such as potatoes with different contextual compatibility. French fries involve deep frying, making potatoes unsuitable for diabetes while oven-roasting is suitable. DMPKG also allows for storing multimodal data such as images, text, and sensor data. The images can be stored as embeddings to enable approximate search. See "DMPKG and Neurosymbolic AI".

HOW DMPKGs FACILITATE COMPOSITIONAL REASONING

Diet Management

DMPKG representation of a recipe shown in Figure 1 enables modifications to fried shrimp tacos to make it suitable for diabetes or a vegetarian diet. A neurosymbolic model will identify the closest recipe for a given recipe using embeddings of the recipe image, title, or instructions. If no match exists, a pipeline proposed constructs a DMPKG for the recipe from text or image (https://tinyurl.com/58x8mnfj). Leveraging this structured, semantically enriched representation, a neurosymbolic AI model can transform fried shrimp tacos into pan-fried cauliflower tacos through the following steps:

1) *Decomposition and semantic understanding*: Decomposing a recipe to capture nuanced relationships between ingredients, cooking methods, domain-specific knowledge, and dietary guidelines can enable the following inferences:
   - Shrimp is tagged as seafood, an animal-derived product incompatible with vegetarian guidelines from the ingredient (substitution) knowledge graph.
   - From the disease-specific knowledge graph, it can be inferred that shrimp is high in cholesterol content and is flagged as unsuitable for diabetes.
   - From the rules, we can infer frying is identified as a cooking method that contributes to trans fats, which are discouraged for diabetic individuals.
2) *Constraint satisfaction through symbolic reasoning*: The neurosymbolic AI model will perform symbolic reasoning by leveraging the DMPKG framework enriched with multiple knowledge graphs to apply dietary and health constraints systematically to modify a recipe as follows:
   - To ensure vegetarian compliance, the system queries the ingredient substitution knowledge graph for plant-based alternatives to shrimp, like cauliflower, a healthy carbohydrate, fiber-rich, and low-glycemic option confirmed by disease-specific knowledge graphs. While alternatives like tempeh or seitan exist, they are processed foods unsuitable for diabetes. This multicontextual understanding helps narrow down meaningful choices.
   - Sour cream is also high in fat and not suitable for diabetes, which can be replaced with low-fat seasoned yogurt.
   - Deep frying, which introduces trans-fat unsuitable for diabetes, can be replaced with pan frying, a healthier fat-based cooking method. While steaming is healthier, it is not suggested as it is water-based and differs from fat-based cooking methods.

As a result, DMPKG ensures explainable results by integrating trusted medical sources into its reasoning. Its graph-based structure decomposes recipes into core entities, applies step-by-step problem-solving, and reconstructs modified recipes. Irrespective of the recipe, the model knows the presence of shrimp is not suitable for a vegetarian diet, enabling generalization. It also guarantees compliance with dietary constraints while retaining the essence of the original recipe, such as transforming fried shrimp tacos into pan-fried cauliflower tacos.

Smart Manufacturing

By mapping the sensor data to high-level process ontology (https://tinyurl.com/4vyuuvmm), DMPKG can be constructed for each assembly run grounded by the semantics of the sensors and assembly procedure as shown in Figure 2 to infer the following:

- Load cell, a sensor that produced anomalous value is attached to robot-2.
- The robot was performing the function of "Robot-2 picks rocket body part-1 from the conveyor."
- The corresponding image at the timestamp did not have "Rocket Body Part-1." Therefore, it is a missing body part that caused the anomaly and not the sensor malfunctioning. Through planning strategies, another robot can be used to add the missing body part.
- If it was a sensor malfunction, the sensor manual knowledge graph can be used to list the steps to the user to calibrate the sensor values.

Similar to the diet use case, the insights of anomaly detection can be explained by the model to ensure the trustworthiness of the domain experts. DMPKG facilitates high-order reasoning in any manufacturing task through its rich modular representation. A similar use case demonstrated that representing AI pipeline metadata as process has benefited recommendation compared to lengthy textual data.[15]

SUPPORTING CAPABILITIES: REPRESENTATION LEARNING, ABSTRACTION, EXPLAINABILITY, AND CAUSAL INFERENCE

Representation Learning

As a neurosymbolic framework, DMPKG supports embedding-based feature extraction and similarity searches across multiple data modalities, enabling efficient processing of unstructured data like images or text. For example, users can request shrimp-containing recipes by providing an image, which is converted to an embedding to find matches. Similarly, users can

DMPKG AND NEUROSYMBOLIC AI

Neurosymbolic AI is a framework that combines deep learning models with knowledge graphs, enabling generalization and higher-order reasoning.[1] Neural networks excel at extracting relevant features by mining patterns from unstructured noisy data and reducing it to embedding vectors, an efficient representation of data approximation. To perform high-level symbolic reasoning tasks such as logical inference, abstraction, and causal inference, these data must be elevated to multiple contexts using knowledge graphs.[S1] DMPKGs support neurosymbolic models and frameworks. For instance, users can search for recipes using ingredient images while applying constraints, such as finding vegan alternatives. The neural component utilizes image embedding similarity to match relevant recipes enabling the system to process unstructured data efficiently. Constraints are then applied using symbolic reasoning, such as identifying vegan substitutions via an ingredient substitution knowledge graph, which requires abstraction and domain-specific understanding. This combination allows for contextualized reasoning essential for tasks like recipe personalization and dietary adjustments. Further, LLMs can be augmented to process natural language queries given by the users.

For recipes without matches in the database, a scalable pipeline has been proposed to extract entities like ingredients and cooking actions to construct a recipe graph.[2] Image-to-text models can generate instructions from images,[3] ensuring visual queries are processed effectively. However, limitations exist. While models can identify general categories, such as tacos or burgers, they may struggle with finer distinctions such as chicken versus beef tacos, necessitating user input. This limitation reflects a broader challenge, as even humans cannot always infer specific details from images. User intervention ensures accuracy and meaningful personalization for tasks requiring nuanced understanding.

REFERENCE

S1. A. Sheth and K. Thirunarayan, "The duality of data and knowledge across the three waves of AI," *IT Prof.*, vol. 23, no. 3, pp. 35–45, May/Jun. 2021, doi: 10.1109/MITP.2021.3070985.

query recipes using natural language, where an LLM can extract entities or identify relevant results by generating embeddings.

Abstraction and Generalization

DMPKGs support high-level abstraction by categorizing entities through common semantics. For example, in diet management, ingredients can be grouped by nutritional properties like "red meat," "high-fiber," or "seafood." This structure allows the system to generalize that any ingredient under seafood isn't vegetarian, helping to identify meaningful substitutes. It can also generalized that irrespective of the recipe, shrimp is not suitable for diabetes due to high cholesterol

Explainability and Traceability

DMPKGs enhance interpretability in compositional reasoning by tracing decision paths for substitutions and modifications. For example, shrimp is replaced with cauliflower to align with vegetarian and diabetes-friendly requirements, while processed alternatives like tempeh and seitan are avoided. Pan frying is preferred over deep frying due to its classification as a healthier fat-based method. All modifications are supported by trusted medical sources like the Mayo Clinic, Center for Disease Control and Prevention, and the United States Food and Drug Administration.

Causal Inference

DMPKGs help understand causal relationships and enhance compositional reasoning with "what-if" scenarios. Replacing shrimp with seitan makes the recipe vegetarian but unsuitable for diabetes while replacing it with cauliflower makes it vegetarian and diabetes-friendly. DMPKG can deduce the effects of ingredient changes and generate modified diet plans with causal justifications from trusted medical sources. Similarly, it can identify the cause of anomalies in smart manufacturing.[13]

FUTURE DIRECTIONS AND RESEARCH OPPORTUNITIES

Several opportunities for advancing the capabilities of DMPKGs merit exploration. One promising direction is automating the development of DMPKGs using knowledge graph construction frameworks such as Empower.[9] Another is the development of algorithms capable of leveraging DMPKGs for real-time reasoning and decision-making. With the rise of vector databases, embedding-based searchers using GPUs and hosting large-scale in-memory graphs have become possible.[14] Expanding DMPKGs to incorporate richer multimodal data sources, such as video and other emerging data formats, presents additional potential for enhancing their utility. Further research into modeling neural network architectures, such as compositional attention networks, could further integrate neural and symbolic components seamlessly. Finally, there is a need to rethink the evaluation of LLMs on compositional reasoning, as current benchmarks often fail to capture deeper reasoning abilities, particularly when superficial changes, such as renaming variables in math reasoning benchmarks, can influence results.[7,8]

REFERENCES

1. S. Sinha, T. Premsri, and P. Kordjamshidi, "A survey on compositional learning of AI models: Theoretical and experimental practices," 2024, *arXiv:2406.08787*.
2. D. A. Hudson, and C. D. Manning, "Compositional attention networks for machine reasoning," 2018, *arXiv:1803.03067*.
3. N. Dziri et al., "Faith and fate: Limits of transformers on compositionality," in *Proc. Adv. Neural Inf. Process. Syst.,* 36, 2024, pp. 70,293–70,332.
4. Z. Li, G. Jiang, H. Xie, L. Song, D. Lian, and Y. Wei, "Understanding and patching compositional reasoning in LLMs," 2024, *arXiv:2402.14328*.
5. J. Dang, A. Hedayati, K. Hampel, and C. Toklu, "An ontological knowledge framework for adaptive medical workflow," *J. Biomed. Inform.*, vol. 41, no. 5, pp. 829–836, 2008, doi: 10.1016/j.jbi.2008.05.012.
6. R. Venkataramanan et al., "Cook-gen: Robust generative modeling of cooking actions from recipes," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Piscataway, NJ, USA: IEEE Press, 2023, pp. 981–986, doi: 10.1109/SMC53992.2023.10394432.
7. I. Mirzadeh, K. Alizadeh, H. Shahrokhi, O. Tuzel, S. Bengio, and M. Farajtabar, "GSM-symbolic: Understanding the limitations of mathematical reasoning in large language models," 2024, *arXiv:2410.05229*.
8. I. Huang et al., "ConMe: Rethinking evaluation of compositional reasoning for modern V0LMs," 2024, *arXiv:2406.08164*.
9. H. Y. Yip, and A. Sheth, "The EMPWR platform: Data and knowledge-driven processes for the knowledge graph lifecycle," *IEEE Int. Comput.*, vol. 28, no. 1, pp. 61–69, Jan./Feb. 2024, doi: 10.1109/MIC.2023.3339858.
10. O. M. Anshakov and T. Gergely, *Cognitive Reasoning: A Formal Approach*. Berlin, Germany: Springer Science & Business Media, 2010.
11. J. Koushik, H. Hayashi, and D. Singh Sachan, "Compositional reasoning for visual question answering," in *Proc. 34 th Int. Conf. Mach. Learn.,* 2017.
12. S. Bubeck et al., "Sparks of artificial general intelligence: Early experiments with GPT-4," 2023, *arXiv:2303.12712*.
13. U. Jaimini, C. Henson, and A. Sheth, "Causal neuro-symbolic AI for root cause analysis in smart manufacturing," in *Proc. Int. Semantic Web Conf.*, 2024. [Online]. Available: https://ceur-ws.org/Vol-3828/paper45.pdf
14. Rickett, C. D. Kristyn, J. Maschhoff, and S. R. Sukumar, "Massively parallel processing database for sequence and graph data structures applied to rapid-response drug repurposing," in *Proc. IEEE Int. Conf. Big Data (Big Data),* Piscataway, NJ, USA: IEEE Press, 2020, pp. 2967–2976.
15. R. Venkataramanan et al., "Constructing a metadata knowledge graph as an atlas for demystifying AI pipeline optimization," *Frontiers BigData*, to be published.

REVATHY VENKATARAMANAN is a Ph.D. student at the AI Institute, University of South Carolina, Columbia, SC, 29208, USA. Contact her at revathy@email.sc.edu.

CHATHURANGI SHYALIKA is a Ph.D. student at the AI Institute, University of South Carolina, Columbia, SC, 29208, USA. Contact her at jayakodc@email.sc.edu.

AMIT P. SHETH is the NCR Chair and a professor at the AI Institute, University of South Carolina, Columbia, SC, 29208, USA. Contact him at amit@sc.edu.

EDITOR: Can Wang, canwang@griffith.edu.au

This article originally appeared in Intelligent Systems vol. 40, no. 5, 2025

DEPARTMENT: AI AND BEHAVIOR

Machine Learning Approaches for Micromobility User Behavior Analysis

Cheng Zhang and Bo Du, *Griffith University, Brisbane, QLD, 4111, Australia*

Qiuyun Luan and Jun Shen, *University of Wollongong, Wollongong, NSW, 2522, Australia*

With widespread adoption globally, micromobility like bikes, e-scooters, and e-bikes has attracted increasing attention due to its ability to complement existing transportation modes and promote sustainable transportation. Understanding micromobility user behaviors in urban areas is essential for improving safety and comfort, as well as for informing infrastructure development and policy. Prior investigations on micromobility user behaviors primarily relied on statistical and kinematic modeling approaches. Although these methods have proven effective in characterizing user behaviors at both macroscopic and microscopic levels, the advent of artificial intelligence (AI)-powered data analytics and behavioral modeling is revolutionizing the field. Recently, advanced machine learning models, such as gradient boosting decision tree, graph convolutional network, and inverse reinforcement learning, has introduced new momentum into micromobility user behavior research. This article explores recent developments, research opportunities, and future directions in this field, leveraging the power of more generic AI approaches.

Micromobility has gained increasing prevalence as a sustainable and convenient transportation option, powered by human or electric energy. It is particularly suitable for short-distance or leisure-related first- and last-mile trips in urban areas. The emergence of micromobility presents a significant opportunity to enhance transportation accessibility, alleviate traffic congestion, and reduce pollution and carbon emissions. The most popular and widely adopted micromobility devices are traditional bicycles (bikes), electric scooters (e-scooters), and electric bicycles (e-bikes). For example, statistics indicate that 12.7 million bikes were produced in the Europe in 2022, dockless e-scooter ridership reached approximately 56.5 million in the United States in 2022, and e-bike ownership in China reached 350 million by 2025.

The rapid growth in private ownership of micromobility devices and the expansion of shared-mobility services have attracted considerable academic attention toward micromobility user behavior analysis.[1] Micromobility user behavior analysis focuses on understanding how individuals navigate roads, choose modes, plan trips, and interact with other road users and their surroundings. However, capturing nonlinear and complex patterns in large-scale behavioral data and identifying factors influencing micromobility utilization remain challenging problems. These issues are closely related to user safety and comfort, as well as micromobility management and operations. Machine learning (ML), an artificial intelligence (AI) technique that learns from data (or experience), has been widely adopted in transport planning and engineering, including mobility data extraction, traffic flow modeling, and road safety assessment. In recent years, A variety of ML approaches have gained increasing prominence in micromobility user behavior research. The goal is to support the design of safe and efficient transportation networks while optimizing traffic management and control systems.

This article reviews the research trends of micromobility user behavior analysis based on ML methods. Specifically, we examine how ML approaches enhance our understanding of variable relationships and

Digital Object Identifier 10.1109/MIS.2025.3597123
Date of current version 2 October 2025.

demonstrate their potential for achieving higher prediction accuracy. Furthermore, we identify key future research directions for further analyzing the mechanisms of micromobility usage patterns in various road environments, which could ultimately improve riding experiences and user satisfaction.

EXISTING STUDIES

Figure 1 illustrates four primary application scenarios of ML approaches in micromobility user behavior analysis: 1) user choice patterns, 2) spatiotemporal trip patterns, 3) crash and injury analysis, and 4) riding behavior analysis. The first two scenarios focus on macro-level aspects, examining why users adopt micromobility and how they navigate complex road networks. The latter two investigate micro-level factors, including safety risks, riding perceptions, movement dynamics, and traffic conflicts during micromobility usage.

User Choice Patterns

User choices on different types of transport modes and different devices for daily travel are changeable and affected by multifaceted factors. Understanding their attitudes and preferences toward micromobility options is crucial for promoting the market penetration of micromobility and reshaping the existing transportation systems. Traditionally, parametric approaches, such as logit models, were commonly used in choice behavior studies and have been proven to be effective. However, these approaches are limited to dealing with complex data structures and nonlinear relationships among variables.[2] ML classification models have emerged as an exploratory tool to investigate the discrete travel choice behaviors. For instance, the gradient boosting decision tree (GBDT) model was found to be useful to analyze user preferences for dockless shared e-bikes affected by travel characteristics, built environment, and shared infrastructure systems. In GBDT, correlated variables could be retained since it inherently considers interaction effects among independent variables. By applying partial dependence plots (PDPs), the nonlinear relationships between independent variables and user preference types could be illustrated. The results showed that the GBDT model outperformed the traditional multinomial logit model with better generalization capability and improved robustness.[2] Another study quantified user preference uncertainty in e-scooter selection based on its usage history and current device status. A robust adversarial reinforcement learning framework was designed to enhance the model's predictive ability. The findings suggest that analyzing user device selection behavior could benefit rebalancing and charging strategies in shared e-scooter systems.[3]

Spatial–Temporal Trip Pattern

Investigations on the spatial–temporal trip pattern focuses on usage pattern (e.g., trip frequency and activity pattern) of micromobility over space and time, and the impact of built environment variables (e.g., transport facilities and land use). Exploring the daily travel activity of micromobility users is beneficial for

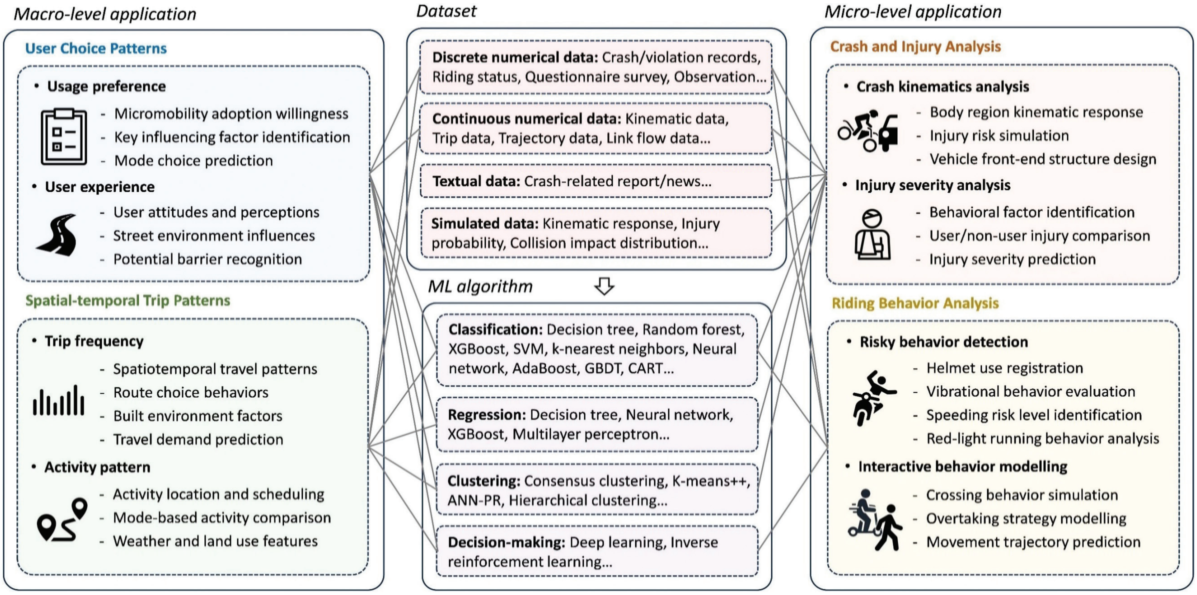

Macro-level application
User Choice Patterns
• Usage preference
- Micromobility adoption willingness
- Key influencing factor identification
- Mode choice prediction
• User experience
- User attitudes and perceptions
- Street environment influences
- Potential barrier recognition
Spatial-temporal Trip Patterns
• Trip frequency
- Spatiotemporal travel patterns
- Route choice behaviors
- Built environment factors
- Travel demand prediction
• Activity pattern
- Activity location and scheduling
- Mode-based activity comparison
- Weather and land use features
Dataset
Discrete numerical data: Crash/violation records, Riding status, Questionnaire survey, Observation...
Continuous numerical data: Kinematic data, Trip data, Trajectory data, Link flow data...
Textual data: Crash-related report/news...
Simulated data: Kinematic response, Injury probability, Collision impact distribution...
ML algorithm
Classification: Decision tree, Random forest, XGBoost, SVM, k-nearest neighbors, Neural network, AdaBoost, GBDT, CART...
Regression: Decision tree, Neural network, XGBoost, Multilayer perceptron...
Clustering: Consensus clustering, K-means++, ANN-PR, Hierarchical clustering...
Decision-making: Deep learning, Inverse reinforcement learning...
Micro-level application
Crash and Injury Analysis
• Crash kinematics analysis
- Body region kinematic response
- Injury risk simulation
- Vehicle front-end structure design
• Injury severity analysis
- Behavioral factor identification
- User/non-user injury comparison
- Injury severity prediction
Riding Behavior Analysis
• Risky behavior detection
- Helmet use registration
- Vibrational behavior evaluation
- Speeding risk level identification
- Red-light running behavior analysis
• Interactive behavior modelling
- Crossing behavior simulation
- Overtaking strategy modelling
- Movement trajectory prediction

FIGURE 1. Application scenarios of ML-based micromobility user behavior analysis.

planning agencies and shared micromobility operators to implement appropriate transport policies and design safe and efficient operation strategies.[4] ML regression and clustering models are emerging data mining technologies to analyze the spatiotemporal distribution of micromobility ridership. For example, the GBDT model can be employed to identify the spatial pattern differences between the shared e-bike and shared e-scooter link flows, and evaluate the relative importance of independent variables. The results indicate that e-scooter link flows appear to be more spatially concentrated and more sensitive to distance to the city center than that of e-bikes.[4] Additionally, a deep multiview spatial–temporal network framework was developed to analyze the hourly demand for shared bike travel patterns. They divided the Beijing urban area into square-mesh grids to link the shared bike trips with multiple urban features, including geography and land use, transport, public vitality, and meteorology. The framework was able to conduct the large-scale real-time prediction of the variation in shared bike demand patterns, like spatial and temporal distribution among nearby regions.[5]

Crash and Injury Analysis

Statistical modeling has been widely utilized in crash and injury analysis, considering its capability to provide reliable arguments for crash severity and frequency with clearly interpretable results. Without the restrictions of preassumed relationships between variables, ML-based approaches become more powerful tools for a multidimensional safety assessment of different types of road users, such as identifying significant variables associated with injury severity categories and predicting the safety risks of different intersections or road segments.[6,7] Currently, two ML-assisted research focuses regarding micromobility users are crash kinematics and injury severity analysis under the influence of different riding behaviors. E-bike riders' injury outcomes caused by front-end collisions with vehicles were analyzed with various rider stature, velocity, and front-end shape settings in the simulation experiments. By applying decision tree models, both the riders' injury outcomes affected by front-end shape parameters and the interaction mechanism between the investigated variables could be interpreted.[6] Additionally, the random forest methodology was applied to analyze the single micromobility (mainly e-scooter and bicycle) crashes by predicting the injury severity and identifying the impacting factors. The research findings suggest that multiple e-scooter riders' behaviors have the potential to increase their injury severity, including without wearing helmet, trip with leisure purpose, and involvement in excessive speed.[7]

Riding Behavior Analysis

Riding behavior research is an effective way to explore various types of detailed riding scenarios, such as speed changes, swerving behaviors, helmet use, and distraction behaviors.[1] Such riding scenarios are recognized as crucial components in affecting micromobility user actual and perceived safety and comfort levels. ML applications in micromobility riding behavior analysis mainly include risky behavior detection and interactive behavior modeling. For instance, a computer vision-based framework was trained to automate cycling stress assessments for urban road networks with the aid of street-view images. The results indicate that a contrastive learning approach could be a useful tool to learn image embedding space and predict the stress levels of cyclists based on image data with a high detection accuracy.[8] Mobility scooter user identification could be realized by capturing the subtle patterns in upper-body movements during the riding through developing a deep learning-based model with a spatiotemporal graph convolutional network. This framework was proved to be effective in reaching high levels of authentication accuracy and uncovering the long-term variability behind rider behaviors.[9] In addition, a deep maximum entropy inverse reinforcement learning (MEIRL) model was adopted to reproduce the crossing behaviors of e-bikes at an intersection. Trajectories of e-bikes and other road users were extracted by an automated algorithm from the drone-based video dataset. The proposed deep MEIRL model was able to predict the e-bike crossing trajectory accurately, especially for the microscopic behaviors of riders.[10]

Overall, as highlighted in previous research, ML methods have emerged as strong techniques for micromobility user behavior analysis. ML models are robust and flexible in capturing nonlinear relationships and complex patterns in the data and evaluating data for prediction or decision making on a large scale. To overcome ML models' difficulties in directly interpreting the modeling results, techniques like the Shapley additive explanation (SHAP) and PDPs have been proposed to quantify each feature's contributions to final predictions.[4]

FUTURE RESEARCH DIRECTIONS

Extensive studies have been conducted on applying ML techniques to analyze micromobility user behavior from various perspectives. However, significant gaps remain in promoting micromobility adoption and user satisfaction. Below, we envision several recommendations for future research to better leverage ML approaches in this field.

- *Evolution of micromobility systems:* As an emerging transportation mode, micromobility has gained substantial popularity with rapidly increasing ridership, presenting significant challenges for transport infrastructure planning and policy design. Future research should employ ML approaches to assess the temporal dynamics of micromobility systems and evaluate their long-term impacts on urban transport networks and user behavior patterns.
- *Multimodal transportation services:* The integration of shared micromobility with public transit presents great opportunities to establish flexible transit networks and mitigate the transportation inequality issues. However, how users transfer between shared micromobility services and public transport modes, and how different mode attributes affect the mode choices has received limited attention, deserving further investigation.
- *Spatiotemporal transferability of ML models:* Transferability is indispensable for validating the generalizability of ML models and understanding micromobility user behavior at macroscopic scales. Spatiotemporal transferability of ML models for micromobility users is still underexplored, probably due to data scarcity and inconsistent formats across jurisdictions. More research is required to transcend the spatial limitation of existing studies to seek the potentials of cross-regional dialogue.
- *Multisource data-based crash analysis:* Current studies primarily rely on police-reported crash data or media-reported accident information for micromobility safety assessments. Future research should incorporate multisource datasets (e.g., mobile signaling and social media data) to better understand the interrelationships among user travel behavior patterns, riding experiences and perceptions, and road safety outcomes.
- *Physics-informed ML approaches:* Incorporating physics-based principles and empirical data with ML models has emerged as an effective way to enhance model performance by improving model generalizability and physical plausibility of results. Consequently, physics-informed ML techniques show significant potential for modeling and simulating micromobility user interactions across diverse traffic conditions.
- *AI-assisted micromobility riding systems:* The significant advancements in AI and autonomous vehicle technologies exalt new opportunities for developing AI-assisted micromobility systems to improve rider safety and comfort. A key application is designing real-time rider monitoring systems for detecting body vibration, alerting for inattentive behaviors, and predicting potential traffic conflicts and accidents.

Despite ML techniques excelling at processing large datasets and modeling complex scenarios, they face significant challenges in interpretability. Although feature importance analysis from ML models helps identify variables' contributions, and interpretable ML methods (e.g., SHAP and PDPs) provide viable solutions, many limitations remain. Key hurdles in practice include the computational intensity of SHAP value calculations and the inability of PDPs to reveal potential interactions between features. These limitations warrant careful consideration in future applications.

CONCLUSION

The advancement of ML approaches has significantly accelerated digital transformation in the transportation sector. This article highlights the pivotal role of ML approaches in micromobility user behavior analysis, particularly in user choice patterns, spatiotemporal trip patterns, crash and injury analysis, and riding behavior analysis. ML techniques have demonstrated exceptional capability in uncovering complex variable relationships, establishing efficient assessment systems, and delivering superior prediction accuracy. Further research should focus more on evolution of micromobility systems, integration into multimodal transportation services, spatiotemporal transferability of ML models, multisource data-based crash analysis, physics-informed ML approaches, as well as AI-assisted micromobility riding systems. These advancements are expected to foster greater micromobility adoption and contribute to more human-centric urban mobility ecosystems.

REFERENCES

1. S. Tuncer and B. Brown, "E-scooters on the ground: Lessons for redesigning urban micro-mobility," in *Proc. ACM Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–14.
2. Y. Liu, L. Li, K. Liu, M. He, and Z. Shi, "Investigating user preferences for dockless bike- and electric bike-sharing through tracking usage patterns," *Transport Policy*, vol. 169, pp. 41–55, Aug. 2025, doi: 10.1016/j.tranpol.2025.04.025.
3. H. Tan, Y. Yuan, H. Yan, S. Zhong, and Y. Yang, "Human preference-aware rebalancing and charging for shared electric micromobility vehicles," in *Proc. IEEE Int. Conf.*

Robot. Autom. (ICRA), 2024, pp. 9608–9615, doi: 10.1109/ICRA57147.2024.10610713.

4. S. T. Jin, and D. Z. Sui, "A comparative analysis of the spatial determinants of e-bike and e-scooter sharing link flows," *J. Transport Geography*, vol. 119, Jul. 2024, Art. no. 103959, doi: 10.1016/j.jtrangeo.2024.103959.
5. J. Chai et al., "St-bikes: Predicting travel-behaviors of sharing-bikes exploiting urban big data," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7676–7686, Jul. 2023, doi: 10.1109/TITS.2022.3197778.
6. Y. Liu, X. Wan, W. Xu, L. Shi, Z. Bai, and F. Wang, "A novel approach to investigate effects of front-end structures on injury response of e-bike riders: Combining Monte Carlo sampling, automatic operation, and data mining," *Accid. Anal. Prev.*, vol. 168, Apr. 2022, Art. no. 106599, doi: 10.1016/j.aap.2022.106599.
7. A. Sanjurjo-de-No, A. M. Pérez-Zuriaga, and A. García, "Analysis and prediction of injury severity in single micromobility crashes with Random Forest," *Heliyon*, vol. 9, no. 12, 2023, Art. no. e23062, doi: 10.1016/j.heliyon.2023.e23062.
8. B. Lin, S. Saxe, and T. C. Chan, "AutoLTS: Automating cycling stress assessment via contrastive learning and spatial post-processing," *Proc. AAAI Conf. Artif. Intell.*, vol. 38, no. 20, pp. 22,222–22,230, 2024.
9. D. Shah, R. Huang, N. Vinayaga-Sureshkanth, T. Chen, and M. Jadliwala, "ScooterID: Posture-based continuous user identification from mobility scooter rides," *IEEE Trans. Mobile Comput.*, vol. 24, no. 2, pp. 970–984, Feb. 2025, doi: 10.1109/TMC.2024.3473609.
10. Y. Wang, S. Wan, Q. Li, Y. Niu, and F. Ma, "Modeling crossing behaviors of E-bikes at intersection with deep maximum entropy inverse reinforcement learning using drone-based video data," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 6350–6361, Jun. 2023, doi: 10.1109/TITS.2023.3248305.

CHENG ZHANG is a senior research assistant at the Department of Management, Griffith University, Brisbane, QLD, 4111, Australia. Contact him at c.zhang@griffith.edu.au.

BO DU is a senior lecturer at the Department of Management, Griffith University, Brisbane, QLD, 4111, Australia. Contact him at bo.du@griffith.edu.au.

QIUYUN LUAN is a Ph.D. student in road safety and travel behavior at the School of Computing and Information Technology, University of Wollongong, Wollongong, NSW, 2522, Australia. Contact her at ql157@uowmail.edu.au.

JUN SHEN is a professor of computer science at the School of Computing and Information Technology, University of Wollongong, Wollongong, NSW, 2522, Australia. Contact him at jshen@uow.edu.au.

IEEE COMPUTER SOCIETY | 80 ANNIVERSARY CELEBRATION

COMPUTER.ORG

CALL FOR SPECIAL ISSUE PROPOSALS

Computer solicits special issue proposals from leaders and experts within a broad range of computing communities. Proposed themes/issues should address important and timely topics that will be of broad interest to *Computer*'s readership. Special issues are an essential feature of *Computer*, as they deliver compelling research insights and perspectives on new and established technologies and computing strategies.

Please send us your high-quality proposals for the 2026 - 2027 editorial calendar. Of particular interest are proposals centered on:

- 3D printing
- Robotics
- LLMs
- AI safety
- Dis/Misinformation
- Legacy software
- Microelectronics

Proposal Guidelines Are Available at:
computer.org/csdl/magazine/co/write-for-us/15911

IEEE

EDITORS: **Brittany Johnson**, George Mason University, johnsonb@gmj.edu
Tim Menzies, North Carolina State University, tim@menzies.us

This article originally appeared in
Software
vol. 41, no. 5, 2024

DEPARTMENT: SE AND ETHICS

Powering Down: An Interview With Federica Sarro on Tackling Energy Consumption in AI-Powered Software Systems

Tim Menzies and Brittany Johnson

FROM THE EDITOR

This column collects news and views on issues of software engineering and ethics. Got something to say on that topic? If so, e-mail a one-paragraph synopsis to timm@ieee.org or johnsonb@gmu.edu (subject line: "SE Ethics Idea: [Your Idea]"). If that looks interesting, we'll ask you to submit a 1,000 –3,000-word article (where each graph, table, or figure is worth 250 words) for review for *IEEE Software*.

—*Tim Menzies and Brittany Johnson*

The energy needs for modern information and communication technology is increasing, dramatically. Roy Schwartz and colleagues tell us that new AI algorithms have resulted in a massive increase in the computational costs of state-of-the-art AI research: as much as 3,000,000 times between 2012 and 2018, and the trend keeps increasing.[1]

Do we need to spend all that energy? Are there ways to reduce it? What is the role of software engineering in that reduction? To get answers to these questions, *IEEE Software* spoke to Prof. Federica Sarro, University College London.

Your research explores many aspects of software engineering including "Green Software Engineering."[2] What is that exactly?

I like to define green software engineering as the discipline that aims at realizing *sustainable* software *sustainably*. It involves using green practices and technologies throughout the software lifecycle in order to diminish the amount of carbon emissions associated with the software production, usage, and maintenance.

What motivates you to work on Green SE?

To realize responsible software systems, we need to go beyond providing the user with the right functionalities; we need to design, implement, and deploy software in a way that considers its impact on the users, the society, and the environment. Equipping software with attractive functionalities and minimizing faults isn't enough if emerging nonfunctional properties of modern software systems, such as fairness, safety, and sustainability, are neglected.[3,4,5] The challenges associated with energy consumption by deep learning models, especially large language models (LLMs), have recently brought more attention to the energy demands of these technologies. AI's energy requirements are immense—a single training run of a large model can consume the equivalent of a typical household's energy consumption over 291 years. This has spurred efforts to make AI-powered systems more energy-efficient and to develop sustainable computing practices.

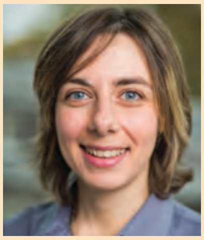
Digital Object Identifier 10.1109/MS.2024.3410011
Date of current version 13 August 2024

Can you give us an example of how that might work?
Well, just to look at our latest article on this topic, at the SSBSE'24 conference we showed that exploiting simple optimizations could lead to substantial savings (266% reduction in inference time) in the use of Stable Diffusion, a popular Generative AI model for text-to-image generation.[6]

We can remember a time when green software engineering (SE) was extensively discussed.[7,8] But now, not so much. Is it a topic that has gone "off the boil"?
The journey toward green SE started more than a decade ago, with Professor Patricia Lago pioneering the field, among others. Nowadays, it's understandable to feel that software energy engineering might have been overshadowed, especially with the rapid advancements and hype surrounding other technological fields like AI and machine learning. These areas often receive more attention due to their rapid development and immediate, visible impact on daily life and business. At the same time, the progress in green SE has hardly been as dramatic. Many advancements in green SE are incremental improvements rather than disruptive breakthroughs. While these steady improvements are vital for long-term sustainability, they may not capture the public's attention in the same way as groundbreaking new technologies. That said, energy engineering remains a critical and active area of research and development, even if it doesn't always capture the headlines.

Can you provide examples where energy consumption is a high priority for organizations?
Energy consumption is becoming a high priority for many organizations due to its impact on operational costs, environmental sustainability, and regulatory compliance. Software contributes to energy consumption on par with other IT components such as hardware, data, and network. However, most of the industrial effort to date has been focused on minimizing the costs of IT infrastructure such as data centers that require significant power and cool environments to operate. Companies like Google, Amazon, and Microsoft invest heavily in energy-efficient technologies and renewable energy sources to power their data centers. More work is needed toward addressing software energy consumption, especially for AI-powered software.

ABOUT PROF. FEDERICA SARRO



Prof. Federica Sarro is a professor of software engineering at University College London, WC1E 6BT London, U.K. For more information, see http://www0.cs.ucl.ac.uk/staff/f.sarro/ or contact her at f.sarro@ucl.ac.uk.

What kind of work is needed?
Software engineers can play a crucial role in making the design choices that affect software energy consumption. We need to better support that decision making. There is a strong need for tools and frameworks that help developers write/use more energy-efficient code. For example, in the FECoM project,[9] we aim to increase developer energy-awareness in the use of deep learning frameworks through fine-grained energy measurement. This work is part of my *Green Shift Left* mission, a broad research agenda supported by many excellent academic and industrial collaborators, aiming at shifting the concern of sustainability from being after-the-fact to being an integral part of the software engineering lifecycle.

> *TO REALIZE RESPONSIBLE SOFTWARE SYSTEMS, WE NEED TO GO BEYOND PROVIDING THE USER WITH THE RIGHT FUNCTIONALITIES.*

Can you offer some success stories in changing energy consumption practices?
I would like to share a project recently developed by UCL computer science students alongside Intel, UCL EnergyGuard.[10] It enables computer users to monitor and reduce their energy consumption while playing games on their laptops or PCs. GPUs are power-hungry, so EnergyGuard is helping users become aware of the

For those who want to explore this area further, we offer the following:

- Official website of the Green Software Foundation (GSF) https://greensoftware.foundation
- Official website of the International Workshop on Green and Sustainable Software (GREENS) http://greens.cs.vu.nl.

Readings:

- S. Georgiou, M. Kechagia, T. Sharma, F. Sarro, and Y. Zou, "Green AI: Do deep learning frameworks have different costs?" in Proc. 44th Int. Conf. Softw. Eng., May 2022, pp. 1082–1094.
- R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Commun. ACM*, vol. 63, no. 12, pp. 54–63, Nov. 2020.
- A. Fonseca, R. Kazman, and P. Lago, "A manifesto for energy-aware software," *IEEE Softw.*, vol. 36, no. 6, pp. 79–82, Nov./Dec. 2019, doi: 10.1109/MS.2019.2924498.
- S. Georgiou, S. Rizou, and D. Spinellis, "Software development lifecycle for energy efficiency: Techniques and tools," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–33, 2019.
- G. Pinto and F. Castor, "Energy efficiency: A new concern for application software developers," *Commun. ACM*, vol. 60, no. 12, pp. 68–75, Dec. 2017, doi: 10.1145/3154384.
- C. Pang, A. Hindle, B. Adams, and A. E. Hassan, "What do programmers know about software energy consumption?" *IEEE Softw.*, vol. 33, no. 3, pp. 83–89, May/Jun. 2016, doi: 10.1109/MS.2015.83.

amount of energy gaming costs them—the electricity bill, hardware costs, subscription costs, bandwidth, and so on. I believe this project exemplifies how the future of green technology should look like by giving the end-user the possibility of making sustainable choices when using any software.

What are the current challenges in this field, and what do you foresee for the future?
The software engineering effort has mainly concentrated on code optimization (i.e.,. refining software code to be more efficient, thereby reducing the energy required for execution) but there is still much more to do to enable the Green Shift Left. Developers need energy-aware development tools. We also need better methods for making end-users aware of their software's energy consumption (e.g., letting them see the accumulated wattage of software as they produce or use it). This will give software users a sense of its real-world cost and the ability to make environmentally friendly decisions.

Are there any legislative needs to support these issues?
Europe has some legislation requiring organizations to report on energy consumption, but similar regulations are lacking elsewhere. Given the increasing prevalence of deep learning and LLMs, this is concerning. There is an important ethical matter policy makers should not forget: The high computational and energy costs associated with powerful AI models can exacerbate inequalities, given that only well-funded organizations can afford to train and deploy these systems. This creates a division between those who have access to advanced AI capabilities and those who do not.

Thanks for talking. Overall, what message would you like to leave the readers of *IEEE Software?*
The energy demands of software, and in particular AI-powered software, are staggering. So it is crucial for the software engineering community to prioritize green engineering. While most research comes from academia, there is a growing interest from industry in developing measurement tools (for notes on that work, see http://www0.cs.ucl.ac.uk/staff/f.sarro/). So we need to coordinate efforts, share tools, and build a trusted ecosystem of certain standards and best practices. For those who want to get involved, I invite you to join groups like the Green Software Foundation.[11]

REFERENCES

1. R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni, "Green AI," *Commun. ACM*, vol. 63, no. 12, pp. 54–63, Nov. 2020, doi: 10.1145/3381831.
2. S. Georgiou, M. Kechagia, T. Sharma, F. Sarro, and Y. Zou, "Green AI: Do deep learning frameworks have different costs?" in *Proc. 44th Int. Conf. Softw. Eng.*, May 2022, pp. 1082–1094, doi: 10.1145/3510003.3510221.

3. Z. Chen, J. M. Zhang, M. Hort, M. Harman, and F. Sarro, "Fairness testing: A comprehensive survey and analysis of trends," *ACM Trans. Softw. Eng. Methodol.*, vol. 33, no. 5, Mar. 2024, Art. no. 137, doi: 10.1145/3652155.
4. F. Sarro, "Search-based software engineering in the era of modern software systems," in *Proc. IEEE 31st Int. Requirements Eng. Conf. (RE)*, 2023, pp. 3–5, doi: 10.1109/RE57278.2023.00010.
5. F. Sarro, "Automated optimisation of modern software system properties," in *Proc. ACM/SPEC Int. Conf. Perform. Eng. (ICPE)*, New York, NY, USA: ACM, 2023, pp. 3–4, doi: 10.1145/3578244.3583739.
6. J. Gong et al., "GreenStableYolo: Optimizing inference time and image quality of text-to-image generation," in *Proc. 16th Int. Symp. Search-Based Softw. Eng.*, 2024, pp. 1–6. [Online]. Available: https://solar.cs.ucl.ac.uk/pdf/GreenStable Yolo.pdf
7. C. Pang, A. Hindle, B. Adams, and A. E. Hassan, "What do programmers know about software energy consumption?" *IEEE Softw.*, vol. 33, no. 3, pp. 83–89, May/Jun. 2016, doi: 10.1109/MS.2015.83.
8. G. Pinto and F. Castor, "Energy efficiency: A new concern for application software developers," *Commun. ACM*, vol. 60, no. 12, pp. 68–75, Dec. 2017, doi: 10.1145/3154384.
9. S. Rajput, T. Widmayer, Z. Shang, M. Kechagia, F. Sarro, and T. Sharma, "Enhancing energy-awareness in deep learning through finegrained energy measurement," 2023. [Online]. Available: https://arxiv.org/ abs/2308.12264
10. "Software by UCL students helps the public measure how much power their apps and games use," Univ. College London, London, U.K., Oct. 2023. [Online]. Available: https://www.ucl.ac.uk/computer-science/news/2023/oct/software-ucl-students-helps-public-measure-how-much-power-their-apps-and-games-use
11. Green Software Foundation. Accessed: Jun. 1, 2024. [Online]. Available: https://greensoftware.foundation/

BRITTANY JOHNSON is an assistant professor of computer science at George Mason University, Fairfax, VA 22030 USA. Contact her at johnsonb@gmu.edu.

TIM MENZIES is a full professor at North Carolina State University, Raleigh, NC 27606 USA. Contact him at timm@ieee.org.

EDITORS: Pradeep K. Murukannaiah, p.k.murukannaiah@tudelft.nl
Munindar P. Singh, mpsingh@ncsu.edu

This article originally appeared in
IEEE Internet Computing
vol. 27, no. 6, 2023

DEPARTMENT: INTERNET ETHICS

Understanding Responsible Computing via Project Management for Sustainability

Hoa Khanh Dam, *University of Wollongong, NSW 2522, Australia*

Aditya Ghose, *deceased*

Nigel Gilbert, *University of Surrey, Guildford, GU2 7XH, U.K.*

Munindar P. Singh, *North Carolina State University, Raleigh, NC, 27695, USA*

Everyone acknowledges the importance of responsible computing, but practical advice is hard to come by. Important Internet applications are ways to accomplish business processes. We investigate how they can be geared to support responsibility as illustrated via sustainability. Sustainability is not only urgent and essential but also challenging due to engagement with human and societal concerns, diverse success criteria, and extended temporal and spatial scopes. This article introduces a new framework for developing responsible Internet applications that synthesizes the perspectives of the theory of change, participatory system mapping, and computational sociotechnical systems.

How can we create responsible Internet applications? We address this question through a particular family of Internet applications, namely, process and project management, and through a particular illustration of responsibility, namely, sustainability (see "Sustainability"). Our envisioned framework highlights the fact that responsibility, like the closely related concept of ethics, cannot be approached from the standpoint of pieces of software but must be viewed in terms of how they affect human outcomes and interactions in a social context.[1] Computer scientists use the term "system" almost exclusively to refer to an artifact realized in software or hardware. Sustainability calls for a broader notion of a system that reflects two connotations: a system encompasses the entities of interest (and so must include stakeholders), and a system is what we place in an environment (and so must take into account its interactions with the surrounding socioeconomic world).[2]

Key challenges arise in incorporating sustainability into the business processes and practices of planning, scheduling, and executing projects.[3] We adopt the term *project management for sustainability* (*PMfS*) to include the desired capabilities and practices. Current management tools (e.g., Zoho Projects, JIRA, and Microsoft Project) emphasize considerations relating to the cost, time, and quality but offer weak support for the complex factors that underpin sustainability.[4] Moreover, achieving responsibility is not merely a matter of applying tools because, conceptually, prior to any tools, the requirements for a project must be understood, it must be ensured that they reflect criteria such as ethics.

This article identifies the key aspects for which project management needs to be expanded: optimization, stakeholders, causal models, and ethics. It offers an interdisciplinary framework for responsibility combining the theory of change and participatory system mapping from the social sciences with computational sociotechnical systems. It concludes with guidance on future research challenges and a call to arms for responsible Internet computing.

CHALLENGES FOR SUSTAINABILITY

To understand the challenges facing PMfS, consider a typical construction project—a housing complex.

1089-7801 © 2023 IEEE
Digital Object Identifier 10.1109/MIC.2023.3310299
Date of current version 6 November 2023.

SUSTAINABILITY

Emerging issues such as climate change, pollution, depletion of natural resources, and social inequality have made sustainability an existential concern for humanity, leading to an urgent need to accommodate sustainability in all business operations.[S1]

Von Carlowitz in his seminal *Sylvicultura Oeconomica*[S2] formulated sustainability in the early 1700s in the context of forestry: how planned reforestation would mitigate the risks of timber shortages. Carson's book *Silent Spring*[S3] inspired the study of sustainability in the modern era. *The Limits to Growth*[S4] was a landmark report based on computer simulations arguing that the finite supply of natural resources is unlikely to support the then-current rates of economic and population growth much beyond 2100.

This report has led to international initiatives, including the Earth Summit and the United Nations Commission on Sustainable Development. Sustainable development is "development that meets the need of the present without compromising the ability of future generations to meet their own needs.[S5]

The *triple bottom line*[S6] captures the three essential elements of sustainability:

- *Social*: culture, accessibility, and participation.
- *Environmental*: soil, water, atmosphere, biodiversity, and energy consumption.
- *Economic*: costs and bureaucracy.

These factors cannot be traded off directly, and coming up with joint criteria is nontrivial. For each of these factors, we must balance short-term and long-term risks and payoffs, local and global scope, transparency, and accountability with privacy, individual freedom, and societal interests. Additionally, these factors interact in subtle ways, and we must tackle the interdependencies.[3]

Sustainability is reliant on the behaviors of stakeholders, with their beliefs and competencies. How well it is achieved depends on how we capture stakeholders' requirements, help them reconcile conflicts, and responsibly trade off current and future needs.

REFERENCES

S1. "UN sustainable development goals," United Nations, New York, NY, USA, 2015. [Online]. Available: https://sdgs.un.org/goals

S2. F. J. Schmithüsen, "Three hundred years of applied sustainability in forestry," *Unasylva*, vol. 64, no. 240, pp. 3–11, Jul. 2013. [Online]. Available: https://www.fao.org/3/i3364e/i3364e01.pdf

S3. R. Carson, *Silent Spring*. San Francisco, CA, USA: Houghton Mifflin, 2002.

S4. D. H. Meadows, D. L. Meadows, J. Randers, and W. W. Behrens III, *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind*. New York, NY, USA: Universe Books, 1972.

S5. United Nations, *Our Common Future*. Oxford, U.K.: Oxford Univ. Press, 1987.

S6. J. Elkington, *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*. Gabriola Island, BC, Canada: New Society Publishers, 1998.

Sustainability is crucial across the three stages in the lifecycle of the complex.

- *Creating and Commissioning*: Developers identify a need and find a site. To obtain approvals for converting that land from its current use, they prepare designs showing how the envisioned complex would fit into its environment in terms of ecology (wildlife habitats), construction (building materials and equipment), transit (road capacity and a new metro stop), utilities (water, sewage, electricity, and telecommunications), and services (schools and clinics). They build the complex.
- *Use*: Households move in, leading to ongoing operations and maintenance, with an environmental footprint.
- *Decommissioning*: Decades later, the complex is taken out of use and possibly demolished. Concerns include the reusability and recyclability of the materials; any pollution caused; and the effects on the local socioeconomic system, such as nearby businesses.

Decisions can have long-term effects. An energy-efficient construction with space for trees lowers the carbon footprint during use. Damage to the environment during use may hinder reintroducing a bird sanctuary upon decommissioning. Furthermore, clean decommissioning facilitates commissioning another project in the same space.

The processes that go into realizing a complex's lifecycle are clearly based on information exchange and decision making. How those decisions are framed yields the requirements that determine whether information technologies are applied responsibly. Thus, PMfS must address the following challenges arising from the wide scope that sustainability induces:

- *Optimization*: Whereas traditional project management has a short horizon, PMfS must consider the entire lifecycle of a project. In our housing complex example, we should include the eventual reclamation of the site upon decommissioning. Since information about the future may be lacking, we may need to combine quantitative and qualitative methods for optimization.
- *Stakeholders*: PMfS relies on stakeholder participation for defining problems, identifying solutions, realizing them, and tracking and assessing outcomes.[5] The stakeholders may change over the lifecycle of a project, e.g., for a housing complex: first the designers; then the building material suppliers; then residents, operators, and the surrounding community members; and, finally, future generations who will decommission the complex. Such stakeholders are autonomous and do not follow a top-down hierarchy as traditional process management assumes.
- *Causal models*: Capturing the interplay of causal relationships at multiple timescales and across organizational boundaries (accounting for autonomous stakeholders) is essential for PMfS. The causal models of interest involve not just physical phenomena (e.g., a garbage incinerator puts out smoke, or big trees help reduce the need for cooling) but also social and cognitive phenomena (e.g., a lack of public transportation leads people to use personal vehicles, or people may adjust their air conditioner settings depending on how their neighbors set theirs).
- *Ethics*: PMfS must contend with competing demands by stakeholders and along different sustainability dimensions. The social norms and values that motivate human behavior are particularly relevant in achieving sustainability. Interesting considerations here involve intergenerational equity (the trading of present prosperity with the future) and intragenerational equity (how welfare is distributed currently) under uncertainty.

ELEMENTS OF A CONCEPTUAL FRAMEWORK FOR RESPONSIBILITY

As these challenges indicate, tackling responsibility requires a new, interdisciplinary framework, which we introduce here in terms of its three main elements.

Theory of Change

To bring about change, i.e., to identify or evaluate potential interventions, we need an understanding, or *theory*, about cause and effect. These theories help us tackle complex systems where outside influences and internal indeterminacy render hard predictions impossible.

The *theory of change* is a way to make the theories underlying an intervention explicit.[6] The theory of change is widely used for policy evaluation by governments and by nongovernment organizations. The approach begins with a project's long-term goals and then works back through intermediary stages until the current state is reached.[7] Laying out a theory of change forces us to articulate the causal links, thereby exposing unproven or problematic assumptions.

Typically, theories of change are developed by starting with the goal of the project and working backward through a causal chain to identify the outcomes that are expected to yield the expected impact. These outcomes map to project outputs (e.g., deliverables and products). Project activities that create the outputs require inputs, such as people, money, and equipment. Thus, we obtain a causal chain that shows the requirements to achieve the goals. Using this chain, one can clarify the assumptions underlying the theory of change and justify these assumptions by reference to prior knowledge, experience, or intuition. It helps to arrange the causal chain along a timeline and specify what resources are needed when, as in conventional process management.

Making explicit the assumptions, contexts, and mechanisms that underlie a project plan reveals misunderstandings and conflicts as well as potential pathways to resolve them. A theory of change can help clarify the assumed scope of a project, including which factors and outcomes are integral to the desired change and which are irrelevant or unchanging. Finally, a theory of change is a useful starting point for computational models of the project domain that incorporate the mechanisms posited in the theory.

Participatory System Mapping

Despite its strengths in laying out the hypothesized causes of change, the theory of change approach may oversimplify complex systems as linear sequences of inputs, activities, outputs, outcomes, and impacts and ignore feedback loops between outputs and inputs or activities. Moreover, the approach emphasizes direct causes and causes that are within a narrow project boundary, risking ignoring exogenous disruptive causes. One way of overcoming these limitations is to combine the theory of change with participatory system mapping.

Participatory system mapping is a modeling methodology in which a group of stakeholders collaboratively develop a causal map of an issue. This map includes factors and links between them. A factor is anything expressed as a variable (i.e., can increase or decrease); a link is a causal relationship between factors. The map represents what stakeholders believe to be the causal structure of their system.

Building a map is a valuable exercise in clarifying participants' understanding. The map is a useful resource, not only for documentation but also for further analysis. Participatory system maps, such as the example in Figure 1, provide the thinking tools that can be used for the discussion and exploration of complex issues as well as sense-checking the implications of suggested causal links.

Figure 1 was created using the Participatory System Mapper tool for participatory mapping.[8] The displayed map shows the theory of change geared to our example. The cocreation of such maps by stakeholders does take effort in that they must reflect on each other's perspectives, but the exercise is facilitated by the structure of the map.

Participatory mapping is helpful in developing an understanding of a domain and identifying the project scope. The map may then be formalized as the basis for a theory of change, or by quantifying the links between factors to yield a system dynamics model, or by building an agent-based model that represents the causal analysis embodied in the map.

However, methods for formalizing system maps into more quantified models are still in their infancy. Further research is needed on methods for refining a system map into a theory of change and then into a computational model.

Computational Sociotechnical Systems

We adopt the notion of a *sociotechnical system*, whose social tier includes stakeholders and whose technical tier includes computational artifacts or resources, such as databases and sensors. The stakeholders have (preferences over) goals and values. They interact with each other to advance their goals and to promote their values; they form expectations of one another and track them. Being autonomous, the stakeholders may violate each other's expectations but concomitantly can hold others—and be held by others—to account.[2] The technical artifacts make their interactions possible

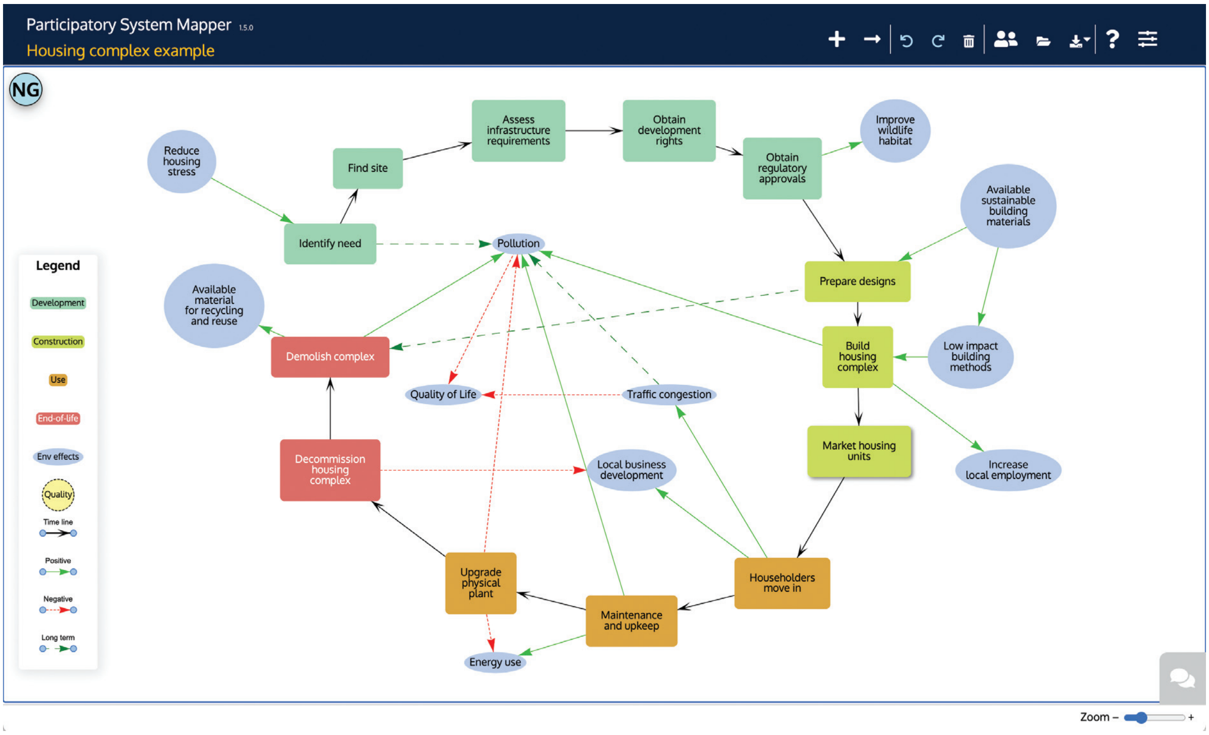


FIGURE 1. An example participatory system map. The rectangles mark the course of the project, color-coded by stage. The ovals represent sustainability factors, and the arrows show causal links. The green arrows are positive relationships, red dotted arrows are negative relationships, and dashed arrows are long-term effects.

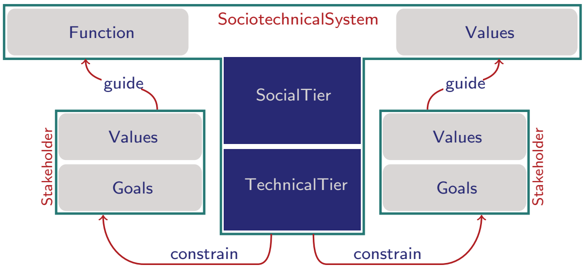


FIGURE 2. A sociotechnical system and its stakeholders schematically comprising a social and a technical tier. It exists for the benefit of its stakeholders, provides some function, and promotes some values. The stakeholders act under the umbrella of the system, constrained by the social tier (e.g., norms) and the technical tier (e.g., affordances). The system itself is guided by and adapts to satisfy the stakeholders it serves, aligning its function with the stakeholders' goals and its values with their values.

and provide affordances that make some interactions easier and some harder.[1]

This vision of a sociotechnical system is computational in that the specific interactions of the stakeholders in the social tier are characterized computationally and are realized by the more detailed computations by the artifacts in the technical tier. Figure 2 illustrates our conception in schematic terms.

We adopt sociotechnical systems as a basis for modeling sustainability. To be sustainable, a system must be open in that the stakeholders can come and go—the system must outlast its stakeholders' tenures in it. Moreover, even for the same stakeholders, their goals and values change over time because of their experiences, needs, and changing social mores. A hallmark of sustainability is the coherence over time of the values realized by the system.

Thinking about sustainability naturally leads to a social tier of norms that guide the behaviors of its members (the stakeholders) yet leaves them with the autonomy to contravene those norms, should that be appropriate. Furthermore, a technical tier remains essential because we need to design our devices and data to be able to respect the dictates of sustainability. As environmental conditions or stakeholder values change (in the extreme case, because of a generational shift, but even otherwise), the currently established norms may no longer be appropriate for many of the stakeholders. When that happens, their behavior in alignment with their values would repeatedly lead to the norms being deviated from.[9] When the norms are deviated from sufficiently often by sufficiently many stakeholders, the social tier has changed in either interpretation of norms: 1) for nonlegislated norms, the deviations are evidence that new norms have emerged, and 2) for legislated norms, there would be ample grounds for revising them.

VISION AND DIRECTIONS

Putting these together, our envisioned conceptual framework for responsible Internet computing involves 1) expanding our conception of a system for an Internet application to include the social structure in which it is (to be) deployed; 2) engaging the stakeholders in capturing their preferences over the functionality desired, their values, and what tradeoffs are acceptable; and 3) developing causal models of possible interventions.

To realize this vision requires advances along each of these three themes. We outline some representative (i.e., not comprehensive, but promising) challenges for each theme.

Causal Modeling

We need ways to achieve better causal modeling, which underpins the theory of change. One way to do so is through the extensive use of agent-based modeling and simulation. In addition, these models should be used as inputs to further participatory mapping to provide empirical grounding to the deliberations being carried out and, thus, help identify misalignment in their assumptions. This process would be recursive in that producing a causal model of one aspect of the project may require additional participatory mapping and uncover additional misalignments.

Value Alignment

Ethics is fundamentally based on values. Responsibility requires that we respect stakeholders' values, which they may have figured out a priori. If the stakeholders support sustainability, they would find that at least some of their values align, suggesting the possibility of them being able to collaborate, though they may be at variance on other values. Even if they do not fully agree, they would need to understand where each other stands. In addition, being able to sufficiently reconcile the tradeoffs they are willing to make would lead to the creation of a social tier that prohibits or disincentivizes certain behaviors and outcomes.

Renewal

Sustainability, by definition, is not a one-shot problem and calls for constant care. A sustainable project must include the ability to monitor its functioning, e.g., to discover if the right goals are being (adequately) met and if the goals being met are the right ones. Based on such observations, if any misalignments arise, we would need a way to make course corrections by changing the process being enacted. Not all such misalignments need to be resolved in a project, so a concomitant challenge is to resolve the project scope such that the participants can focus their efforts on the most germane aspects.

Such improvements can be seen in terms of optimization. Given the stakeholders' preferences, we expect the system to maximize an associated objective that reflects those preferences. When the outcomes are no longer optimal—either because new knowledge indicates better solutions or because the stakeholders' preferences have shifted—we would need to revise the operations in the technical tier as well as the applicable norms in the social tier. As explained, mostly, those revisions may be incremental, but, sometimes, they may require extensive changes.

Putting it together, we can see that achieving responsible Internet computing requires more than an exhortation to be responsible or do the right thing. Computer scientists must engage stakeholders in ways that go beyond current approaches focused on surveys or interviews to 1) help stakeholders understand the causal models of the technical aspects, which they may not understand well; 2) obtain their help in jointly developing models of social interactions, where they would have experience; and 3) jointly elicit and formalize the values and tradeoffs to be embodied in the solution.

ACKNOWLEDGMENTS

This work was partially supported by our universities via a collaboration network called the University Global Partnership Network. Singh was additionally supported by the U.S. National Science Foundation (Grant IIS-2116751).

We dedicate this article to the memory of our friend and collaborator, Prof. Aditya Ghose, who passed away unexpectedly in 2023.

REFERENCES

1. P. K. Murukannaiah and M. P. Singh, "From machine ethics to internet ethics: Broadening the horizon," *IEEE Internet Comput.*, vol. 24, no. 3, pp. 51–57, May/Jun. 2020, doi: 10.1109/MIC.2020.2989935.
2. M. P. Singh, "Norms as a basis for governing sociotechnical systems," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 1, pp. 21:1–21:23, Jan. 2014, doi: 10.1145/2542182.2542203.
3. A. J. G. Silvius and R. P. J. Schipper, "Sustainability in project management: A literature review and impact analysis," *Social Bus.*, vol. 4, no. 1, pp. 63–96, 2014, doi: 10.1362/204440814X13948909253866.
4. L. Sabini, D. Muzio, and N. Alderman, "25 years of 'sustainable projects'. What we know and what the literature says," *Int. J. Project Manage.*, vol. 37, no. 6, pp. 820–838, Aug. 2019, doi: 10.1016/j.ijproman.2019.05.002.
5. F. T. Edum-Fotwe and A. D. F. Price, "A social ontology for appraising sustainability of construction projects and developments," *Int. J. Project Manage.*, vol. 27, no. 4, pp. 313–322, May 2009, doi: 10.1016/j.ijproman.2008.04.003.
6. C. H. Weiss, "Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families," in *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, J. P. Connell, A. C. Kubisch, L. B. Schorr, and C. H. Weiss, Eds. Washington, DC, USA: Aspen Institute, 1995, pp. 65–92.
7. D. H. Taplin and H. Clark, "Theory of change basics: A primer on theory of change," Actknowledge, New York, NY, USA, Tech. Rep., Mar. 2012. [Online]. Available: https://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf
8. N. Gilbert. *PRSM: A Participatory System Mapper*. (2020). [Online]. Available: https://prsm.uk/
9. A. M. Singh and M. P. Singh, "Norm deviation in multiagent systems: A foundation for responsible autonomy," in *Proc. 32nd Int. Joint Conf. Artif. Intell. (IJCAI)*, Macau, Aug. 2023, pp. 289–297, doi: 10.24963/ijcai.2023/33.

HOA KHANH DAM is a professor of software engineering at the University of Wollongong, NSW 2522, Australia. Contact them at hoa@uow.edu.au.

ADITYA GHOSE, deceased, was a professor of computer science at the University of Wollongong, NSW 2522, Australia.

NIGEL GILBERT is a professor of sociology at the University of Surrey, Guilford, GU2 7XH, U.K. Contact them at N.Gilbert@surrey.ac.uk.

MUNINDAR P. SINGH is a professor of computer science at North Carolina State University, Raleigh, NC, 27695, USA. Contact them at m.singh@ieee.org.

EDITOR: Schahram Dustdar, dustdar@dsg.tuwien.ac.at

DEPARTMENT: INTERNET OF THINGS, PEOPLE, AND PROCESSES

This article originally appeared in IEEE Internet Computing vol. 29, no. 2, 2025

Human-Based Distributed Intelligence in Computing Continuum Systems

Praveen Kumar Donta, *Stockholm University, 10691, Stockholm, Sweden*

Boris Sedlak and Ilir Murturi, *TU Wien, 1040, Vienna, Austria*

Victor Casamayor Pujol, *Universitat Pompeu Fabra, 08018, Barcelona, Spain*

Schahram Dustdar, *TU Wien, 1040, Vienna, Austria, and UPF ICREA, 08018, Barcelona, Spain*

Distributed computing continuum systems (DCCSs) are integrated systems that combine cloud, edge, and Internet of Things devices to deliver scalable and low-latency computing resources across diverse applications and environments. Composed of a heterogeneous mix of computational units, storage systems, and communication networks, DCCSs facilitate real-time data processing and analysis by distributing tasks dynamically based on resource availability and demand. The complex structure of DCCSs reflect the intricate organization of the human body, where different systems work together to maintain overall functionality. This article draws parallels between the human body's intelligence mechanisms and the operational strategies needed for DCCSs. We especially explore several human body analogies or principles that can be incorporated into DCCSs to mitigate interpretable and noninterpretable challenges while enhancing overall performance.

Distributed computing continuum systems (DCCSs) are an emerging computing paradigm that integrates cloud, edge, mobile edge, the Internet of Things, and other computing environments into a unified system.[1,2] In DCCSs, computing tasks are dynamically distributed across geographically dispersed nodes, thus effectively reducing latency by processing data closer to their source. This uses resource hierarchy upon demand and thereby enhances overall system efficiency. With DCCSs, tasks are allocated to the most appropriate computing nodes when needed, resulting in improved performance and resilience. This capability meets the growing demand for complex, data-intensive, real-time data processing, big data analysis, and emerging generative artificial intelligence (AI) applications. DCCSs can be used in a wide variety of fields,[3] including smart cities, health care, autonomous vehicles, and industrial automation, where multiple objectives are required, including low latency, efficient resource utilization, high accuracy, and energy efficiency. DCCSs improve operational efficiency by ensuring the optimal use of computing resources and reducing the need for centralized data processing and storage.

Recently, DCCSs have also integrated smart devices with embedded intelligence,[4] allowing them to autonomously perform simple sensing and control tasks, thereby reducing the need for manual intervention. However, intelligence is needed not only within individual devices but across the entire system. In particular, DCCSs require intelligent and efficient mechanisms to ensure reliable integration, optimal performance, and robust management of complex operations. AI and machine learning (ML) are increasingly utilized in DCCSs for advanced control strategies, enabling predictive

1089-7801 © 2025 IEEE
Digital Object Identifier 10.1109/MIC.2024.3460908
Date of current version 30 June 2025.

maintenance, orchestration, elasticity (e.g., autoscaling), efficient resource utilization, latency minimization, and optimization of processes through real-time data analysis.[5] For instance, in Zhang et al.,[6] multilevel edge management and control modes form hierarchical structures to manage varying levels of control across the system, ranging from local controllers to central supervisory systems. Extended with causality, DCCSs can be easier understood and managed, allowing deeper insights into system dynamics and interactions through analysis of their behavior and impact.[7]

Traditional AI and ML techniques have limitations, particularly in DCCSs; one major issue is that they require vast amounts of input and training data, which can be impractical in resource-constrained environments. These approaches are also computationally intensive, requiring significant processing power and energy, which can be costly and inefficient. In some cases, the benefits of deploying traditional AI/ML solutions can outweigh their costs. However, adversarial attacks can undermine systems' security and reliability, especially faulty training models or mismatched input for training.[8] Moreover, it is difficult to verify real-time accuracy, which complicates their deployment in dynamic and critical applications.

In general, AI/ML approaches rely on historical data to generate or suggest outcomes or solve issues. However, addressing noninterpretable future challenges or needs can be difficult when they are based solely on past information. DCCSs require efficient solutions that anticipate and adapt to future demands, enabling full autonomy and adaptability; however, current AI/ML strategies are not sufficient for that. Alternatively, the human ecosystem serves as a prime example of an intelligent living system. Recent studies have highlighted how the human body efficiently adapts and adjusts to changing environmental conditions through its complex and well-coordinated internal systems.[9] Similarly, the intricate structure of the human body closely resembles DCCS architectures, suggesting that insights from biological systems could inspire more effective and adaptive solutions for managing and optimizing DCCSs.[10] This approach ensures that systems remain effective and relevant as they evolve and face future challenges. With this motivation, we present this article with the following contributions:

- We draw parallels between the complex mechanisms within the human body and DCCSs.
- We explore and study various intelligent mechanisms of the human body that could be effectively incorporated into DCCSs to enhance their adaptability and efficiency.

HUMAN BODY ANALOGY WITH DCCSs

The human body is an intricate and highly sophisticated system, composed of cells organized into tissues (approximately 37.2 trillion cells, each specialized for different functions), and organ systems that work in harmony to sustain life. There are several systems in the body that are deeply interconnected, including the skeletal, muscular, nervous, cardiovascular, and respiratory systems. All of these systems depend on one another to function properly. Similarly, DCCSs mirror this complexity and interconnectedness, as shown in Figure 1. We have classified the human body's anatomy into two main categories, infrastructure systems and regulatory systems, which are further discussed in the next sections in relation to DCCSs.

Infrastructure Systems

The human body infrastructure system contains three parts: the skeletal, cardiovascular, and nervous systems. These systems are vital for maintaining the structure of the body, efficiently moving resources, and coordinating functions across various parts. These systems are closely parallel to key elements within DCCSs, such as devices, communication or coordination, data transmission or flow, and learning or knowledge extraction capabilities. Next we illustrate how biological principles can inform and enhance our understanding of complex computing systems.

In the human body, the skeletal system provides a network of bones and joints that facilitate movement while protecting vital organs. Similarly, a DCCS's structural framework is analogous to its physical infrastructure, which includes devices such as servers, routers, and data centers. Just as the skeleton supports the body, this physical infrastructure supports various computing tasks so that the system functions effectively. The cardiovascular system circulates blood, delivering oxygen and nutrients to cells while removing waste, ensuring that every body part receives the resources it needs. In DCCSs, it is data that move through networks and communication channels, thus ensuring that computational resources and information are efficiently distributed across nodes. Hence, data can be seen as the system's "blood." Next, the cardiovascular system's role in maintaining steady blood flow parallels how DCCSs manage energy supply, dataflow, and bandwidth to maintain optimal performance and prevent bottlenecks. Finally, the nervous system transmits signals, processes information, and coordinates actions, enabling the body to respond quickly to stimuli and adapt to new situations. Similarly, in

DCCSs, communication networks and learning capabilities facilitate rapid information exchange between distributed nodes, ensuring task coordination and timely responses.

Medical research shows that neurons are not confined to the brain, they are present in various organs throughout the body, each serving specialized functions: the brain, with its approximately 86 billion neurons, serves as the central hub for processing information and decision making. The spinal cord, which contains roughly 13.5 million neurons, functions as a crucial communication pathway between the brain and the rest of the body. The enteric nervous system, often referred to as the *second brain*, manages the digestive system with approximately 500 million neurons, overseeing gastrointestinal processes independently. The heart contains roughly 40,000 neurons that regulate its rhythm and function, ensuring efficient cardiac operations. The eyes have approximately 100 million neurons in their retinas, which process visual information.[11] Additionally, the skin is equipped with millions of sensory neurons that detect touch, pressure, temperature, and pain, allowing the body to respond to environmental stimuli. This complex distribution of neurons across various organs highlights the body's sophisticated system for managing diverse and essential functions.

Neurons are crucial for both physiological functions and cognitive processes. They enable the brain to acquire, process, and store information and adapt responses based on experience—an aspect that is also central to learning in AI and ML. The widespread distribution of neurons across different organs is an instance of distributed intelligence, similar to how DCCSs operate. Edge and fog devices can perform localized computations, like the heart and gut, while the cloud handles more extensive computations. For instance, DCCSs can directly compute visual data analytics at intelligent edge devices, hence mimicking the high amount of neurons integrated into the system's "eyes." Just as the human body's various parts collaborate to manage complex tasks efficiently, organs with specialized neurons work together to form an integrated and responsive system. This parallel suggests that DCCSs can emulate the decentralized processing and coordination seen in the human body, enhancing its efficiency, intelligence, and autonomy.

Regulatory Systems

The human body's regulatory system contains two parts: the lymphatic and endocrine systems; together, they regulate the body's internal environment, preserving balance and coordinating responses to internal and external changes. The lymphatic system is essential for maintaining fluid balance and protecting against infection. It includes lymph nodes, vascular system vessels, and lymphoid organs like the spleen

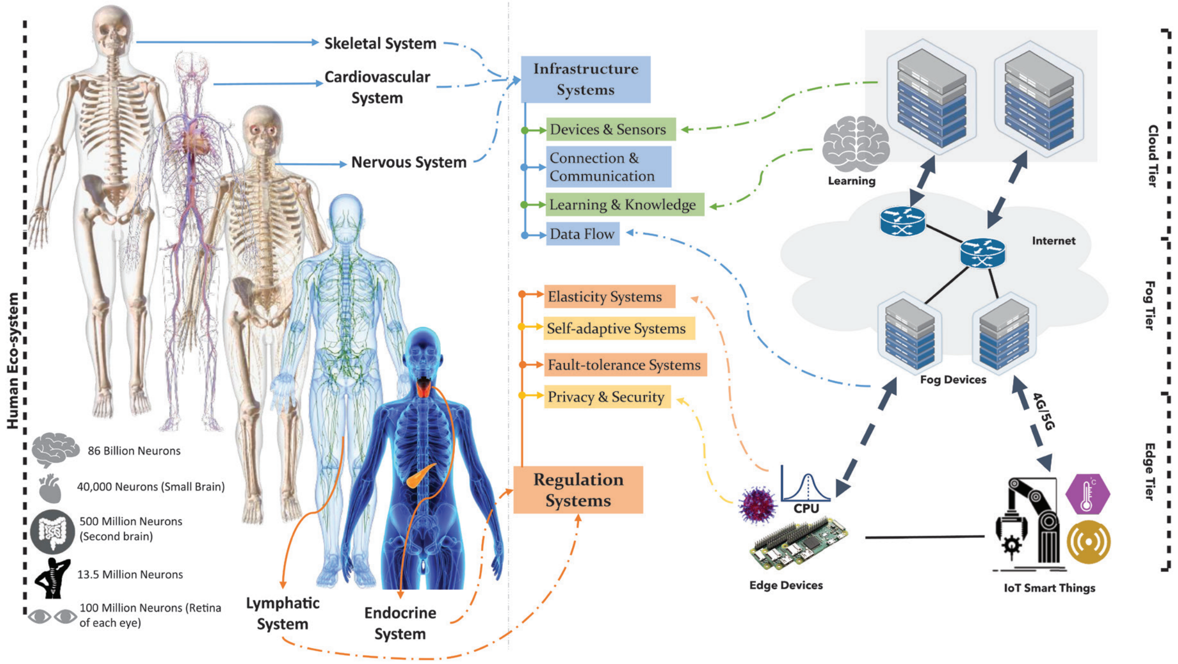


FIGURE 1. Parallels between the human body and DCCSs. IoT: Internet of Things.

and tonsils, which collect excess tissue fluid, filter out pathogens, and support the immune system by transporting lymphocytes. The endocrine system, on the other hand, regulates various physiological processes through hormone secretion. For this, glands release hormones into the bloodstream to control growth, metabolism, mood, and reproduction, thus maintaining overall bodily balance and homeostasis.

Similarly, DCCSs incorporate elastic features to dynamically allocate resources based on demand, e.g., during peak usage times. Thus, it is possible to scale up a service according to external requirements. To adapt to changes and recover from disruptions, DCCSs can also ensure fault tolerance through redundancy and recovery protocols, thus mirroring abilities such as the immune system's response to infection. Similar to the lymphatic system, DCCSs regulate data transfer to ensure efficient communication and security. The endocrine system's role in hormonal signaling parallels distributed control mechanisms in DCCSs, where various components and nodes use protocols and algorithms to coordinate tasks, optimize performance, and adapt to changing conditions.

HUMANLIKE INTELLIGENCE IN DCCSs

The previous section demonstrated that the human body exemplifies distributed intelligence throughout its intricate and interconnected systems, each of which is essential for maintaining overall function and adaptability. In this section, we examine how DCCSs can benefit from incorporating humanlike systems. Specifically, we examine how concepts such as feedback loops and adaptability, self-healing capabilities, efficient communication and coordination, and advanced decision-making processes can significantly enhance DCCS functionality. Figure 2 summarizes which intelligent mechanisms of the human body can be replicated as operational strategies in a DCCS. We assume a healthy human body for the following analogies.

Feedback Loops and Adaptability

In the human body, feedback loops play a crucial role in regulating processes such as cell growth, division, self-repair, and responses to environmental changes. These loops are essential for maintaining homeostasis through long-range extracellular feedback between cells. Feedback loops involve a process in which a system's output influences its own activity, either by amplifying it (positive feedback) or reducing it (negative feedback).[12] A negative feedback loop is a control mechanism in which the output of a system counteracts the initial stimulus, helping the system to maintain stability and equilibrium. For example, when blood sugar levels rise after eating, the pancreas releases insulin, which helps cells absorb glucose and lower blood sugar levels. As blood sugar drops to normal levels, insulin secretion decreases. If blood sugar levels fall too low, the pancreas releases glucagon to raise it back up, maintaining balanced blood sugar levels. On the other hand, a positive feedback loop is a process in

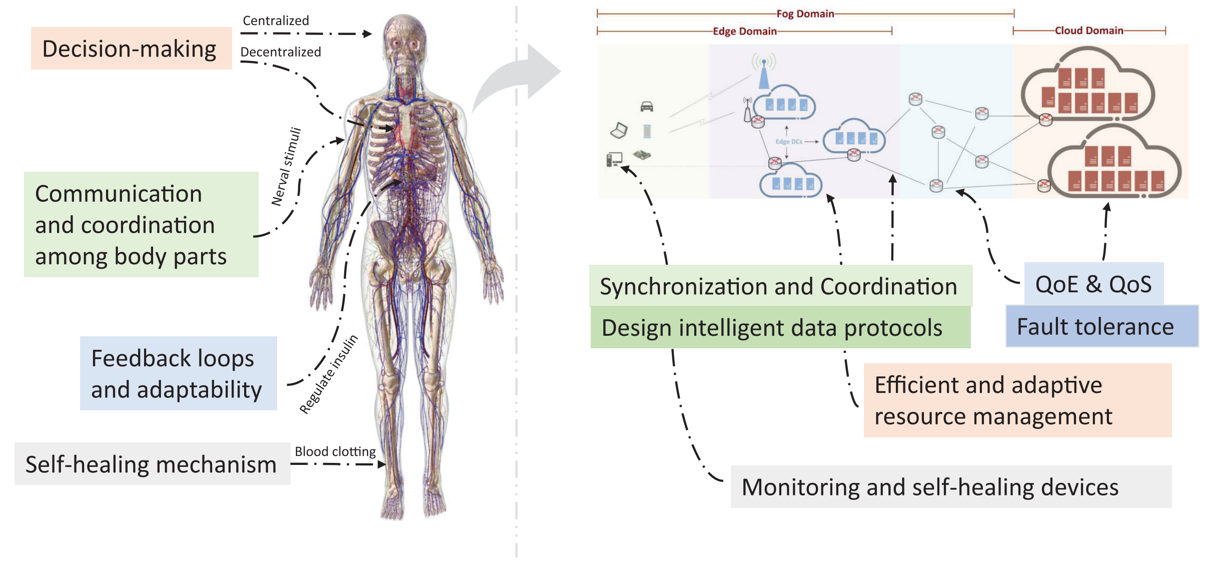

Decision-making
Centralized
Decentralized
Nerval stimuli
Communication and coordination among body parts
Feedback loops and adaptability
Regulate insulin
Self-healing mechanism
Blood clotting
Fog Domain
Edge Domain
Cloud Domain
Edge DCs
Synchronization and Coordination
Design intelligent data protocols
QoE & QoS
Fault tolerance
Efficient and adaptive resource management
Monitoring and self-healing devices

FIGURE 2. Human body's intelligent mechanisms and the operational strategies need for DCCSs. QoE: quality of experience; QoS: quality of service.

which the output of a system amplifies the initial stimulus, pushing the system farther in the same direction. Unlike negative feedback, positive feedback loops often lead to rapid changes and are typically involved in processes that require quick decisions. Blood clotting is a well-known positive feedback loop. When a blood vessel is injured, platelets stick to the site and release chemicals that attract more platelets. As more platelets accumulate, they release even more chemicals, rapidly amplifying the clotting process until the wound is sealed.

To enhance adaptability, resilience, and quality of service (QoS), bioinspired feedback loops can be introduced to DCCSs: negative feedback loops help DCCSs maintain the system's overall stability and adhere to service-level objectives (SLOs) by balancing resource allocation.[13,14] For example, when computational resources such as bandwidth, CPU, or memory meet critical SLO thresholds, the system can automatically scale down certain processes or offload tasks to prevent overload and minimize potential damage. This proactive adjustment helps ensure that performance remains within SLO parameters. Similarly, positive feedback loops are crucial for rapidly scaling resources in response to sudden spikes in demand, thereby supporting the SLOs related to performance and latency. For instance, when a surge in data processing is detected, positive feedback mechanisms can trigger swift allocation of additional computing resources, such as autoscaling virtual machines or distributing tasks across edge devices. Implementing these feedback control strategies in a DCCS allows for coping with unforeseen fluctuations, such as occasional network breakdowns,[15] a sudden rise in resource demand, or on-demand privacy enforcement on streaming data.[16] Thus, in case of failures, dynamic adjustments and resilience mechanisms enhance quality of experience and QoS.

Self-Healing Mechanism

Self-healing is the body's natural ability to restore balance and health without external intervention, often utilizing feedback loops to regulate this process. But can self-healing mechanisms be applied to DCCSs? The answer is, to some extent, yes. In our previous work,[17] we explored this concept by drawing parallels with human wound healing to manage and govern participating devices within a computing continuum with low human intervention. These self-healing agents could be employed to address abnormal device activities by drawing from the human wound-healing process; this process was structured into four distinct stages: hemostasis, inflammation, proliferation, and reshaping. Figure 3 provides a high-level overview of this approach, along with four stages and their operations.

Hemostasis, the body's immediate response to injury, parallels the DCCS approach of isolating and mitigating failures. When a disruption occurs in a DCCS, affected services are swiftly isolated, and tasks are

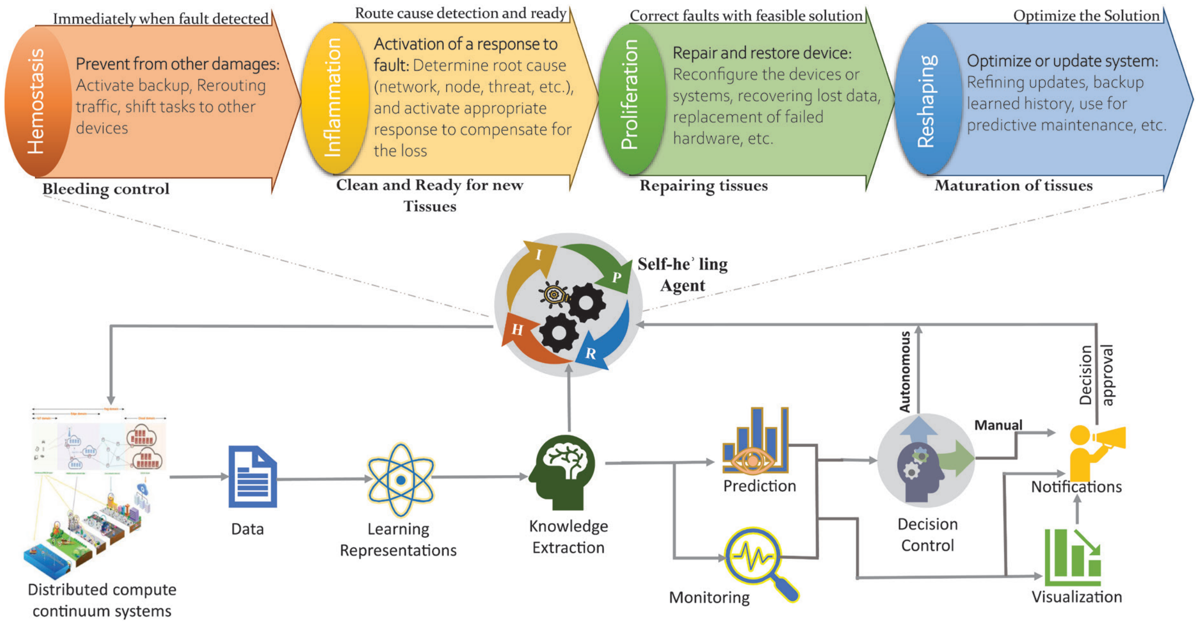

Immediately when fault detected
Hemostasis
Prevent from other damages: Activate backup, Rerouting traffic, shift tasks to other devices
Bleeding control
Route cause detection and ready
Inflammation
Activation of a response to fault: Determine root cause (network, node, threat, etc.), and activate appropriate response to compensate for the loss
Clean and Ready for new Tissues
Correct faults with feasible solution
Proliferation
Repair and restore device: Reconfigure the devices or systems, recovering lost data, replacement of failed hardware, etc.
Repairing tissues
Optimize the Solution
Reshaping
Optimize or update system: Refining updates, backup learned history, use for predictive maintenance, etc.
Maturation of tissues
I
P
H
R
Self-he' ling Agent
Distributed compute continuum systems
Data
Learning Representations
Knowledge Extraction
Prediction
Monitoring
Autonomous
Decision Control
Manual
Decision approval
Notifications
Visualization

FIGURE 3. Wound self-healing mechanism for monitoring, governance, and predictive fault tolerance in DCCSs.

rerouted to stable computing nodes, minimizing damage and preserving overall system stability. *Inflammation* involves immune cells identifying and containing the problem site in the body. Similarly, in a DCCS, this stage entails diagnosing the root cause, whether it is a network issue, malfunctioning node, or security threat. The system gathers relevant data or learned representations, diagnoses the issue, and activates appropriate response measures, such as zero-touch service provisioning, to contain the disruption. This focused action prevents the problem from escalating and prepares the system for recovery. *Proliferation* in wound healing is characterized by the growth of cartilage tissue to replace damaged cells. In a DCCS, this stage is mirrored by reconfiguring devices or systems, restoring services from backups, redistributing tasks across the network, autonomously recovering lost data, and replacing failed hardware. These measures ensure that the system returns to normal operations quickly, maintaining stability and minimizing downtime. Finally, the *reshaping* phase involves strengthening the new tissue and realigning the structure. In a DCCS, this stage is reflected in postrecovery optimization, in which the system refines updates, backs up learned history, and utilizes these data for predictive maintenance. By analyzing the failure and recovery process, the system implements improvements such as enhancing fault tolerance, updating algorithms, and refining monitoring mechanisms, thus bolstering its resilience against future disruptions.

Communication and Coordination

The human body's communication and coordination systems are remarkably complex and interconnected, which is comparable to a DCCS's. The human body employs a sophisticated system of communication and coordination between its various parts, primarily through two main systems: the nervous and endocrine systems.[18] These systems work together to maintain homeostasis, respond to environmental stimuli, and regulate bodily functions. The nervous system acts as a rapid communication network, transmitting electrical signals through neurons to coordinate immediate responses, whereas the endocrine system provides slower but more sustainable communication through hormones. Similarly, a DCCS needs robust communication and coordination systems among devices (ranging from edge, fog, and the cloud). A typical computing system focuses on low-latency and ultrareliable communications, similar to a nervous system. If an edge device detects a critical event, such as a security breach, it immediately transmits this information to other nodes in the network, just like when you touch something hot, the nervous system sends pain signals to your brain. However, sometimes it is not necessary to be low latency, but sustainable and reliable communications are necessary. For example, periodic updates or reconfigurations that maintain system stability and efficiency do not necessarily require faster communications.

Furthermore, circadian rhythms in the human body demonstrate an even broader level of coordination. The suprachiasmatic nucleus in the brain communicates with peripheral clocks in organs such as the liver, muscles, and adipose tissue. The metabolic processes are synchronized with the day–night cycle through a variety of signals, including hormones and neural pathways. For example, the liver clock regulates glucose metabolism in response to feeding patterns, while muscle clocks influence insulin sensitivity and glucose uptake. This coordinated timing system allows the body to anticipate and prepare for daily physiological demands, optimizing energy use and metabolic function. DCCSs employ advanced scheduling and load-balancing mechanisms to maintain resource management equilibrium across the network. DCCSs predict and adapt to fluctuations in network traffic and processing demands in a way that is similar to circadian rhythms in the body. For example, during peak traffic hours, the system dynamically allocates resources by distributing tasks across various nodes, leading to efficient load handling. Similar to how the body conserves energy during periods of rest, the system may downscale operations during periods of low traffic.

Decision Making

Decision-making processes in the human body are highly sophisticated and efficient; in particular, they are balanced between centralized and decentralized components.[19] The brain serves as the central command center for decision making; it processes sensory information, coordinates complex behaviors, and initiates responses. Nevertheless, not all decisions are governed by this centralized system; many organs and systems have a degree of decentralized, autonomous decision making. For example, the heart can regulate its rate based on local conditions without relying on the brain, and the digestive system has its own enteric nervous system that functions independently. The spinal cord triggers immediate responses without involving the brain, while the endocrine system facilitates decentralized communication by releasing hormones, allowing for flexible regulation without central coordination. Similarly, the immune system can detect and respond to threats locally, ensuring rapid and targeted action against infections.

DCCSs apply a similar approach to their decision-making process: the cloud manages high-level orchestration and overall system stability, while decentralized nodes and devices perform autonomous decisions based on local data. However, this process needs decision making not only at the source but also during transmission, which is similar to the endocrine system in the human body. The security systems in DCCSs achieve superficial advantages of adapting features and a working model of the immune system.[20]

CONCLUSION

This article explored the parallels between the human body's intelligence mechanisms and the operational strategies required for DCCSs. We aim to address both interpretable and noninterpretable challenges in DCCSs while improving their overall performance by examining human body analogies and principles. Specifically, we investigate how concepts such as feedback loops and adaptability, self-healing mechanisms, decision-making processes, and communication and coordination can be applied to DCCSs based on the human body analogy. Each of these aspects provides valuable insights into creating a more intelligent and responsive DCCS. In the future, we will implement these techniques and our solutions using existing or new toolsets to bring our insights and gains closer to reality.

ACKNOWLEDGMENTS

This work is supported in part by the European Union grant agreements. 101070186 (TEADAL), and Ayuda CNS2023-144359 financiada por MICIU/AEI/10.13039/501100011033 y por la Unión Europea Next GenerationEU/PRTR.

REFERENCES

1. S. Dustdar, V. C. Pujol, and P. K. Donta, "On distributed computing continuum systems," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 4092–4105, Apr. 2023, doi: 10.1109/TKDE.2022.3142856.
2. P. K. Donta, I. Murturi, V. Casamayor Pujol, B. Sedlak, and S. Dustdar, "Exploring the potential of distributed computing continuum systems," *Computer*, vol. 12, no. 10, 2023, Art. no. 198, doi: 10.3390/computers12100198.
3. M. Nardelli, G. Russo Russo, and V. Cardellini, "Compute continuum: What lies ahead?" in *Proc. Eur. Conf. Parallel Process.*, Cham, Switzerland: Springer Nature Switzerland, Aug. 2021, pp. 5–17, doi: 10.1007/978-3-031-50684-0_1.
4. S. Deng, H. Zhao, W. Fang, J. Yin, S. Dustdar, and A. Y. Zomaya, "Edge intelligence: The confluence of edge computing and artificial intelligence," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7457–7469, Aug. 2020, doi: 10.1109/JIOT.2020.2984887.
5. T. Meuser et al., "Revisiting edge AI: Opportunities and challenges," *IEEE Internet Comput.*, vol. 28, no. 4, pp. 49–59, Jul./ Aug. 2024, doi: 10.1109/MIC.2024.3383758.
6. C. Zhang, W. Zhang, H. Zhong, T. Zhao, and Y. Zhang, "Multi-level edge intelligent management and control mode of safety production based on safety informatics," *Adv. Eng. Informat.*, vol. 62, Oct. 2024, Art. no. 102751, doi: 10.1016/j.aei.2024.102751.
7. V. C. Pujol, B. Sedlak, P. K. Donta, and S. Dustdar, "On causality in distributed continuum systems," *IEEE Internet Comput.*, vol. 28, no. 2, pp. 57–64, Mar./Apr. 2024, doi: 10.1109/MIC.2023.3344248.
8. K. D. Gupta and D. Dasgupta, "Adversarial attacks and defenses for deployed AI models," *IT Prof.*, vol. 24, no. 4, pp. 37–41, Jul./Aug. 2022, doi: 10.1109/MITP.2022.3180330.
9. M. Valentinuzzi, *The Organs of Equilibrium and Orientation as a Control System*, vol. 2. Boca Raton, FL, USA: CRC Press, 1980.
10. V. Casamayor Pujol, P. K. Donta, A. Morichetta, I. Murturi, and S. Dustdar, "Distributed computing continuum systems–Opportunities and research challenges," in *Proc. Int. Conf. Service-Oriented Comput.*, Cham, Switzerland: Springer Nature Switzerland, Nov. 2022, pp. 405–407, doi: 10.1007/978-3-031-26507-5_41.
11. P. A. Muller et al., "Microbiota modulate sympathetic neurons via a gut–brain circuit," *Nature*, vol. 583, no. 7816, pp. 441–446, Jul. 2020, doi: 10.1038/s41586-020-2474-7.
12. H. El-Samad, "Biological feedback control—Respect the loops," *Cell Syst.*, vol. 12, no. 6, pp. 477–487, Jun. 2021, doi: 10.1016/j.cels.2021.05.004.
13. V. Casamayor Pujol, B. Sedlak, Y. Xu, P. K. Donta, and S. Dustdar, "DeepSLOs for the computing continuum," in *Proc. Workshop Adv. Tools, Program. Lang., Platforms Implement. Eval. Algorithms Distrib. Syst.*, Jun. 2024, pp. 1–10, doi: 10.1145/3663338.366368.
14. B. Sedlak, V. C. Pujol, P. K. Donta, and S. Dustdar, "Markov blanket composition of SLOs," in *Proc. IEEE Int. Conf. Edge Comput. Commun. (EDGE)*, Jul. 2024, pp. 128–138, doi: 10.1109/EDGE62653.2024.00025.
15. B. Sedlak, V. C. Pujol, P. K. Donta, and S. Dustdar, "Equilibrium in the computing continuum through active inference," *Future Gener. Comput. Syst.*, vol. 160, Nov. 2024, pp. 92–108, doi: 10.1016/j.future.2024.05.056.
16. B. Sedlak, I. Murturi, P. K. Donta, and S. Dustdar, "A privacy enforcing framework for data streams on the edge," *IEEE Trans. Emerg. Topics Comput. Intell.*,

vol. 12, no. 3, pp. 852–863, Jul./Sep. 2024, doi: 10.1109/TETC.2023.3315131.

17. P. K. Donta, B. Sedlak, V. Casamayor Pujol, and S. Dustdar, "Governance and sustainability of distributed continuum systems: A big data approach," *J. Big Data*, vol. 10, no. 1, Apr. 2023, Art. no. 53, doi: 10.1186/s40537-023-00737-0.
18. J. Holler, "Visual bodily signals as core devices for coordinating minds in interaction," *Philos. Trans. Roy. Soc. B*, vol. 377, no. 1859, Sep. 2022, Art. no. 20210094, doi: 10.1098/rstb.2021.0094.
19. A. D. Scarffe, A. Coates, J. M. Evans, and A. Grudniewicz, "Centralization and innovation: Competing priorities for health systems?" *Int. J. Health Planning Manage.*, vol. 37, no. 5, pp. 2534–2541, Sep. 2022, doi: 10.1002/hpm.3531.
20. I. Murturi, P. K. Donta, V. C. Pujol, A. Morichetta, and S. Dustdar, "Learning-driven zero trust in distributed computing continuum systems," in *Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf. Pervasive Intell. Comput., Int. Conf. Cloud Big Data Comput., Int. Conf. Cyber Sci. Technol. Congr. (DASC/PiCom/CBDCom/CyberSciTech)*, Abu Dhabi, United Arab Emirates, 2023, pp. 0044–0049, doi: 10.1109/DASC/PiCom/CBDCom/Cy59711.2023.10361352.

PRAVEEN KUMAR DONTA is a senior lecturer at the Department of Computer and Systems Sciences, Stockholm University, 10691, Stockholm, Sweden. His research interests include learning in distributed continuum systems. Donta received his Ph.D. degree in computer science and engineering from the Indian Institute of Technology (Indian School of Mines), Dhanbad. He is a Senior Member of IEEE. He is a member of Association for Computing Machinery Contact him at praveen@dsv.su.se.

BORIS SEDLAK is a Ph.D. student in computer science in the Distributed Systems Group, TU Wien, 1040, Wien, Austria. His research interests include edge intelligence, causal methods for the computing continuum, and service-oriented computing. Sedlak received his M.Sc. degree in software engineering and Internet computing from TU Wien. He is a Member of IEEE. Contact him at boris.sedlak@dsg.tuwien.ac.at.

ILIR MURTURI is a postdoctoral researcher in the Distributed Systems Group, TU Wien, 1040, Vienna, Austria. His research interests include the Internet of Things, Distributed Computing Continuum Systems, and EdgeAI. Murturi received his Ph.D. degree in information systems engineering from TU Wien. He is a Member of IEEE. Contact him at imurturi@dsg.tuwien.ac.at.

VICTOR CASAMAYOR PUJOL is a project assistant at Universitat Pompeu Fabra, 08018, Barcelona, Spain. His research interests include self-adaptive methodologies for computing continuum systems, including service-level objective-based definitions, causal and machine learning inference, and robotics. Casamayor Pujol received his Ph.D. degree in information and communication technologies from Universitat Pompeu Fabra. He is a Member of IEEE. Contact him at victor.casamayor@upf.edu.

SCHAHRAM DUSTDAR is a full professor of computer science and heads the Research Division of Distributed Systems at TU Wien, 1040, Vienna, Austria, and UPF ICREA, 08018, Barcelona, Spain. His research interests include the investigation of all aspects related to edge computing, fog computing, and cloud computing. Dustdar received his Ph.D. degree in business informatics from Johannes Kepler Universität Linz. He is a Fellow of IEEE. Contact him at dustdar@dsg.tuwien.ac.at.

EDITOR: Joanna F. DeFranco, The Pennsylvania State University, jfd104@psu.edu

This article originally appeared in Computer vol. 58, no. 8, 2025

DEPARTMENT: INTERNET OF THINGS

Life at Risk: Uncovering the Urgent Security Gaps in Internet of Things-Integrated Cloud Infrastructures

Syed Rizvi and Anthony Demeri, *The Pennsylvania State University*

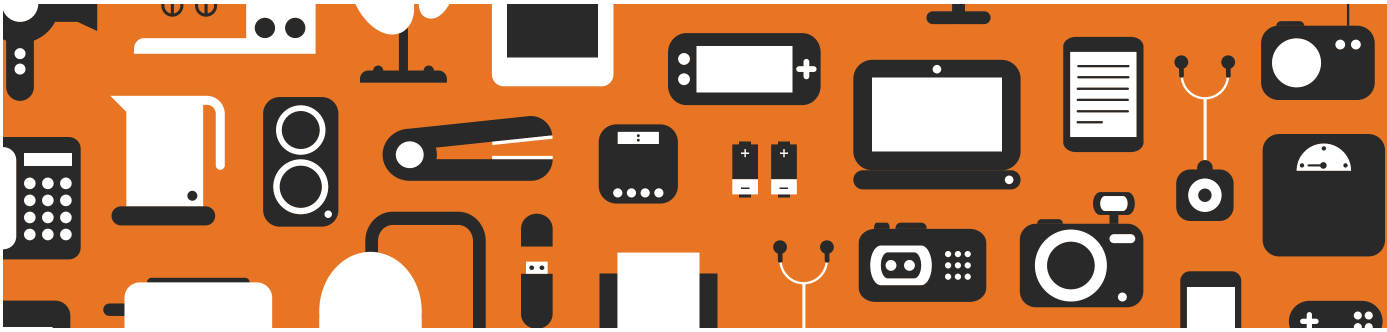
The rapid adoption of cloud-integrated Internet of Things devices has greatly outpaced the development of adequate security practices. In critical domains, this can lead to life-threatening consequences.

Allow yourself the luxury, for a moment, to turn back time. Wind the clocks back a few decades, and stop right around the early days of the Internet's emergence. Let your ears reimagine the beeping, whirring, and honking sounds associated with the multiminute, phone-hogging, dial-up connection. Never a soul could have predicted the tremendous growth of power, size, and quantity for Internet-capable computers. Before long, our watches were streaming video-calls in real time, our vehicles could be started from across the world, and our medications were administered autonomously. Thus, as they say, the era of the Internet of Things (IoT) emerged. Somewhere along the way, as the approximately 75 billion IoT devices[1] trickled across the globe, a parallel industry also arose with prominence: cloud computing. Today, IoT-integrated cloud infrastructures provide undeniable improvements in our quality of life across multiple domains…*but—as* the title gave away—an insecure convergence of these two booming sectors can yield life-threatening consequences.

But before we dive into the nitty-gritty details, let's first consider exactly *why* and *how* IoT devices are so rapidly adopted and integrated into cloud infrastructures. To start, we should note the many advantages of existing IoT devices, which merge the digital and physical worlds, gathering real-time information and making real time decisions—either autonomously or human-directed—at a fraction of the cost of human monitoring. These task-specific devices are used by insurance companies, health-care providers, municipalities, businesses, and, of course, individuals, just like us.[1,2,3,4] Similarly beneficial, cloud computing platforms provide their users plug-and-play access to scalable, high-performance resources at a reduced cost (since users do not maintain their own hardware, cooling platforms, or physical networks). Plainly put, both IoT devices and cloud computing offer their users multiple extremely cost-effective service models, which can significantly improve one's quality of life.

Physically, IoT devices are typically small, with relatively limited processing power and storage capacity, often put in place for an individual's convenience, as might be smart assistants (for example, Amazon Alexa), smart deadbolts, and smart coffee makers.[1] More recently, however, widespread adoption has grown to also include life-saving devices, such as pacemakers,[1] medical infusion pumps,[5] patient vital-monitoring systems,[4,6] traffic control systems, vehicle control systems,[10] water treatment controllers,[7] thermostats,[1] and surveillance systems.[1] From a performance perspective, integrating these devices into cloud infrastructure is both intuitive and resource-efficient, since such integrations expand the capacity for remote management and administration. Of course, although these IoT devices have the potential for enabling an unprecedented increase in quality of life, particularly in the health-care domain,[8] they have continuously been shown to have vast security vulnerabilities.[1,2] Further

Digital Object Identifier 10.1109/MC.2025.3561445
Date of current version: 28 July 2025



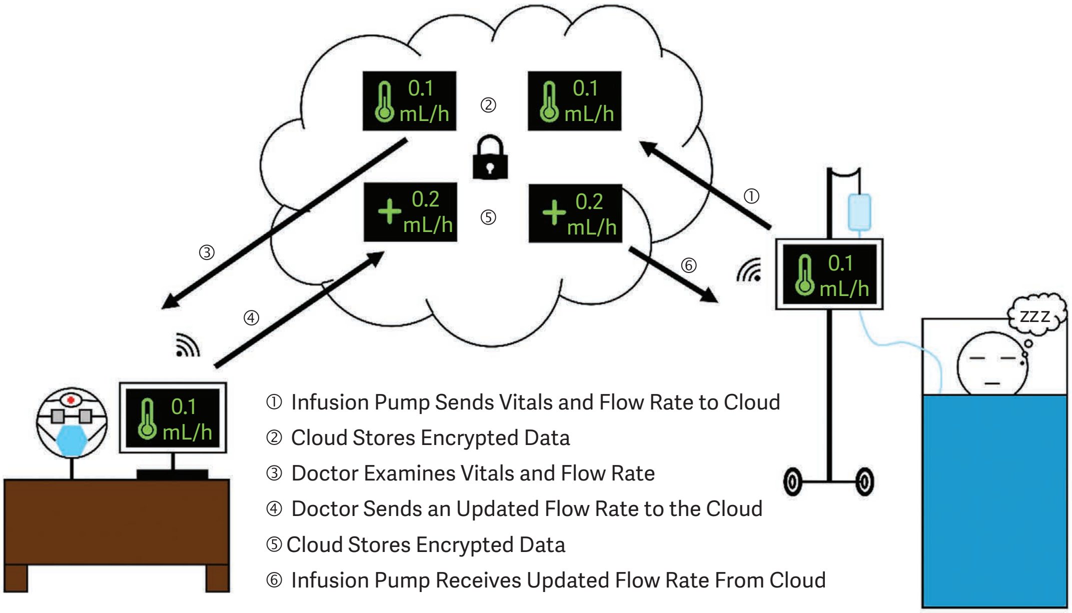

0.1 mL/h
②
0.1 mL/h
0.2 mL/h
⑤
0.2 mL/h
①
③
⑥
④
0.1 mL/h
ZZZ
0.1 mL/h
① Infusion Pump Sends Vitals and Flow Rate to Cloud
② Cloud Stores Encrypted Data
③ Doctor Examines Vitals and Flow Rate
④ Doctor Sends an Updated Flow Rate to the Cloud
⑤ Cloud Stores Encrypted Data
⑥ Infusion Pump Receives Updated Flow Rate From Cloud

FIGURE 1. Remote medication management via a cloud-integrated IoT infusion pump.

exacerbating the problem, the cloud-integration of these already-vulnerable IoT devices results in a significant increase in the surface area of attack vectors, which are both easy to exploit and complex to prevent.[9] Although cloud service providers (CSPs) tout native encryption solutions for the protection of critical data, the primary objectives of this work are to 1) demonstrate existing security practices as wholly insufficient for protecting sensitive cloud service user data and 2) examine how the use of IoT devices in critical domains (for example, health care, smart cities) can lead to life-threatening consequences.

IoT AND CLOUD COMPUTING

To understand the gravity of the risks involved, we should first briefly discuss the interaction between IoT and cloud computing; as with many things, the best way is to look at a diagram. Take a look at Figure 1. In Figure 1, critical data are both sent-to and received-from the cloud infrastructure, where it exists in an encrypted form. Unfortunately, although CSPs suggest such encryption as a sufficient means of protecting data from unauthorized disclosure or manipulation, security researchers note there are no proven countermeasures for entirely protecting data from attackers, especially malicious insiders.[9] As a result of these security gaps, life is at risk for a variety of stakeholders, including manufacturers, industry leaders, municipalities, health-care providers, and, perhaps most importantly, the end user—*you!*

Fortunately, these security gaps have not gone entirely unnoticed by regulatory personnel. In January 2025, the White House launched the "U.S. Cyber Trust Mark," to help consumers better ascertain the security aspects of IoT devices in the home setting. While this is a good start, solving this problem will require all stakeholders (for example, end users, manufacturers, and industry leaders) to find a balance between economic benefits and convenience while maintaining an acceptable level of security risk.

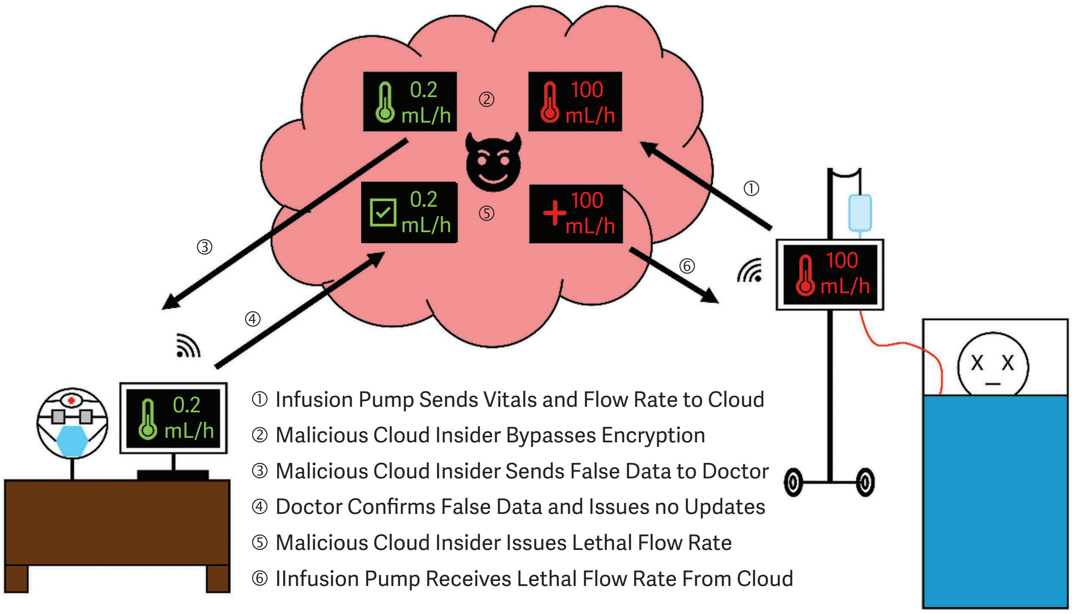

0.2 mL/h
②
100 mL/h
①
0.2 mL/h
⑤
100 mL/h
③
⑥
100 mL/h
④
0.2 mL/h
X X
① Infusion Pump Sends Vitals and Flow Rate to Cloud
② Malicious Cloud Insider Bypasses Encryption
③ Malicious Cloud Insider Sends False Data to Doctor
④ Doctor Confirms False Data and Issues no Updates
⑤ Malicious Cloud Insider Issues Lethal Flow Rate
⑥ IInfusion Pump Receives Lethal Flow Rate From Cloud

FIGURE 2. Cloud-level attack disrupts accurate medical monitoring in IoT-based health care.

SECURITY CHALLENGES

We've already said IoT devices and cloud infrastructures are vulnerable. In fact, in many cases, we may even know *where* they're vulnerable, so, can't we just, you know, fix it? Well, as you might expect, the solution is never quite so simple as the problem. Due to the highly competitive nature of the field, IoT device manufacturers are heavily disincentivized from investing time and resources into security considerations for their products, since such investments increase a product's cost and time-to-market. Furthermore, while some entities have enacted privacy policies, such as the California Consumer Privacy Act and the General Data Protection Regulation, there is no sufficient global security framework or standard for IoT devices. But, even if IoT manufacturers did secure their individual devices from attackers, since cloud computing necessarily involves an inherent sharing of resources with multiple (potentially malicious) entities, integrated IoT devices remain susceptible to a variety of complex cloud-based attacks with cascading, catastrophic consequences.[9] Sadly, even the native encryption solutions offered by cloud service providers are insufficient countermeasures for modern attack vectors, which may originate from within the cloud infrastructure itself. To better illustrate this vulnerability, let us build upon Figure 1; this time, however, in Figure 2, we insert a malicious insider on the cloud infrastructure, which is known to be hard to both detect and prevent.[9] In this case, a malicious insider sends a lethal dosing rate and patient vital signs are critical. Although the IoT pump sends this data to the cloud, since the cloud has been compromised by a malicious insider, false data are instead sent to the authorized medical professional, erroneously indicating the patient is safe, putting life at risk.

Note that although the critical data from Figure 1 was encrypted, a malicious insider is able to bypass such encryption and directly relay false or malicious data to the doctor and the IoT-integrated infusion pump, respectively. While this example might seem too shocking to be plausible, in 2024, Cyble's[5] security researchers noted 75% of infusion pumps have unpatched IoT devices and more than a whopping 50% of hospital IoT devices are vulnerable to attack. With very few medical IoT devices running active antimalware, the risk is more than theoretical; in fact, attackers have already begun to exploit these vulnerabilities in the wild, putting millions of lives at risk.

LIFE AT RISK

Sadly, there are countless examples of real-world exploitation of IoT-integrated cloud infrastructures. Let's take a look at a few examples within the domains of critical infrastructure, control systems, and health care.

Critical infrastructure

Shockingly, in 2021, cyber attackers obtained remote access to supervisor control and data acquisition (SCADA) systems at a drinking water facility in Florida.[7] Initial investigations supposed the attacker obtained SCADA system access through cloud-based remote-management software, such as TeamViewer. After obtaining access, the attacker used the SCADA system to increase the amount of sodium hydroxide (lye) being added to the drinking water via IoT networked controllers. Lye, a caustic chemical, can cause severe damage to human tissues, the consumption of which could be fatal. Fortunately, an employee at the water treatment plant noticed the change and corrected the issue before the attack propagated into the water supply.

Control systems

In 2015, Wired's Andy Greenberg (Security) notoriously broke a story[10] where researchers Chris Valasek and Charlie Miller remotely exploited a stock 2014 Jeep Cherokee while Andy was traveling 70 mph—the five-minute video is worth a watch, if you can spare the time. The researchers began with a few practical jokes, setting the fan speeds to maximum power and displaying a comical picture on the display screen, but, before long, they progressed into more frightening actions. The researchers set the stereo to a disorienting volume, simultaneously spraying the windshield with wiper-fluid, neither of which Andy could override. They killed the engine next, slowing the highway-bound vehicle to a spontaneous and dangerous crawl, without the warning of the brake lights being displayed to other drivers on the highway. Ultimately, Chris and Charlie found thousands of vehicles could be remotely controlled to a severe extent, including disabling the brakes and overriding steering controls.

In 2016, in Lappeenranta, Finland—a particularly cold region, with winter temperatures reaching below 0 °F—hackers shut down an IoT heating controller for multiple apartments, causing a direct threat to life in the sub-freezing temperatures.[1] These heating controllers, like many IoT devices, did not have sufficient means of preventing distributed denial of service (DDoS) attacks. Since these IoT devices were cloud-integrated, with software necessitating connection to a remote management server, the DDoS attacks caused an infinite reboot cycle, ultimately preventing heating operations. Fortunately, this multiday attack yielded no known fatalities.

Health care

Recently, in 2024, Change Healthcare suffered one of the largest cyberattacks in health care to date,[6] causing cash flow shortages of over a billion USD and shutting down services for pharmacies, records, clinics, dentists, and patients. With patients unable to receive care or fill prescriptions, there was an immediate risk posed by an influx of health crises and emergency room visits. In this attack, malign actors allegedly gained remote access to cloud services through the use of a legitimate password. This attack, and others like it, cause direct disruption to patient care, leading to costly lawsuits and, unfortunately, increased fatalities.

> *SADLY, EVEN THE NATIVE ENCRYPTION SOLUTIONS OFFERED BY CLOUD SERVICE PROVIDERS ARE INSUFFICIENT COUNTERMEASURES FOR MODERN ATTACK VECTORS.*

BROADER IMPACT

While the previous case studies demonstrate the acute impact such life-threatening attacks can have on the daily lives of us as individuals, there are also broader impacts to be considered from a societal, economic, and environmental perspective. For society, cyber failures can yield significant erosion in public trust for networked systems. Additionally, impacted stakeholders will face the economic costs associated with data breaches, legal consequences, reputation damage, and shareholder hesitancy. Of course, the environmental risk is also significant, with damage to essential infrastructures causing waste—at a minimum—and potential environmental catastrophe—think energy spills. The bottom line? We're all impacted.

RECOMMENDATIONS

We have demonstrated the bidirectional vulnerabilities inherent in IoT-integrated cloud infrastructures. Naturally, solving the problem similarly mandates a bidirectional approach. Remember, IoT device manufacturers are effectively economically bounded, such that they often cannot adequately address security concerns in their respective products. Furthermore, even if a given manufacturer did implement security measures, there is no globally recognized method for an end user to identify the presence or strength of these measures. Thus, in order to overcome these challenges, we recommend rapid design and publication of global security standards for IoT devices, such that both individual and enterprise-level consumers can easily identify the given strengths (and weaknesses) of a particular device before deciding to integrate it into their respective environment.

Fortunately, for IoT devices, "the IEEE recently published the Standard for Clinical Internet of Things (IoT) Data and Device Interoperability with TIPPSS (trust, identity, privacy, protection, safety, security) principles offering a comprehensive framework for designing secure, interoperable devices that protect individuals and maintain data integrity."[8] These standards are critical for clinical IoT devices, but remain lacking for IoT devices, in general, which can similarly be exploited with catastrophic consequences. While the adoption of such standards by manufacturers will still permit individual consumers to purchase IoT devices which fall short of meeting adequate security guarantees, critical infrastructure can be protected through regulatory requirements for specific security standards.

On the other hand, unlike IoT devices, cloud services are typically provided by large entities, such as Amazon (AWS), Microsoft (Azure), and Google (Cloud). Often times, these providers are simply administrators of third-party hardware and software (as opposed to manufacturers). Thus, in order to mitigate the security challenges at the cloud infrastructure level, it will take a combination of effort from underlying hardware manufacturers—such as AMD and Intel—and software developers. At present, AMD and Intel provide encrypted-memory solutions via their AMD Secure Encrypted Virtualization and Intel Total Memory Encryption platforms. While these options only address a portion of the data security problem, we recommend all cloud service providers adopt such platforms at a minimum. The National Institute of Standards and Technology and the Cybersecurity and Infrastructure Security Agency further advocate for additional mitigations through implementation of rigorous intrusion detection, prevention, and monitoring systems[9]; in practice, this may be accomplished with AI-powered threat detection.[11]

Ultimately, until effective mechanisms exist to guarantee true-isolation of *all* user-data on cloud service resources, data will always be vulnerable. With vulnerable cloud data comes risk for all IoT-integrated devices and, more importantly, their users: us.

The widespread adoption of IoT-integrated cloud infrastructures has the potential to benevolently reshape our world, offering significant advantages to businesses, municipalities, and individual end users. Unfortunately, the vulnerabilities within this ecosystem are ripe for exploitation by malign actors, the consequences of which are life-threatening. We call on regulators, cloud service providers, and manufacturers to immediately prioritize security for all echelons from tiny, wearable IoT devices to enormous cloud-level infrastructures. Moving forward, it is imperative we strike a balance between technological convenience and fiscal responsibility, without compromising our safety.

REFERENCES

1. V. Kumar Jain and J. Gajrani, "IoT security: A survey of issues, attacks and defences," in *Intelligent Learning for Computer Vision* (Lecture Notes on Data Engineering and Communications Technologies), vol. 61, H. Sharma, M. Saraswat, S. Kumar, and J. C. Bansal, Eds., Springer: Singapore, 2021, pp. 219–236.
2. T. Sasi, A. H. Lashkari, R. Lu, P. Xiong, and S. Iqbal, "A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges," *J. Inf. Intell.*, vol. 2, no. 6, pp. 455–513, Nov. 2024, doi: 10.1016/j.jiixd.2023.12.001.
3. S. Ray, K. N. Mishra, and S. Dutta, "Big data security issues from the perspective of IoT and cloud computing: A review," *Recent Adv. Comput. Sci.*

Commun., vol. 14, no. 7, pp. 2057–2078, Oct. 2021, doi: 10.2174/2666255813666200224092717.

4. N. Almolhis, A. M. Alashjaee, S. Duraibi, F. Alqahtani, and A. N. Moussa, "The security issues in IoT - Cloud: A review," in *Proc. 16th IEEE Int. Colloq. Signal Process. Appl. (CSPA)*, Piscataway, NJ, USA: IEEE Press, 2020, pp. 191–196, doi: 10.1109/CSPA48992.2020.9068693.
5. Cyble, "Must-read Cyble research reports of 2024: Trends and key takeaways," Dec. 2024. Accessed: Mar. 21, 2025. [Online]. Available: https://cyble.com/blog/must-read-cyble-reports-2024-trends-key-takeaways/
6. "HealthSec USA Summit 2024 Annual Report," HealthSec Cyber Secur. for Healthcare, Boston, MA, USA, May 2024. Accessed: May 20, 2025. [Online]. Available: https://healthsec.cs4ca.com/wp-content/uploads/HealthSec-2024-Annual-Report.pdf
7. "Compromise of U.S. water treatment facility." CISA (.gov). Accessed: Mar. 20, 2025. [Online]. Available: https://www.cisa.gov/news-events/cybersecurity-advisories/aa21-042a
8. J. F. DeFranco, "Health care 4.0 and clinical Internet of Things," *Computer*, vol. 57, no. 10, pp. 90–92, Oct. 2024, doi: 10.1109/MC.2024.3426568.
9. D. Tank, A. Aggarwal, and N. Chaubey, "Virtualization vulnerabilities, security issues, and solutions: A critical study and comparison," *Int. J. Inf. Tecnol.*, vol. 14, no. 2, pp. 847–862, Mar. 2022, doi: 10.1007/s41870-019-00294-x.
10. A. Greenberg, "Hackers remotely kill a jeep on the highway—With me in it," *WIRED*, Jul. 21, 2015. Accessed: Mar. 24, 2025. [Online]. Available: https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/
11. X. Liang and Y. Xu, "A novel framework to identify cybersecurity challenges and opportunities for organizational digital transformation in the cloud," *Comput. Secur.*, vol. 151, Apr. 2025, Art. no. 104339, doi: 10.1016/j.cose.2025.104339.

SYED RIZVI is a professor of information sciences and technology at The Pennsylvania State University, University Park, PA 16802 USA. Contact him at srizvi@psu.edu.

ANTHONY DEMERI is an active software engineer and doctoral student at The Pennsylvania State University, University Park, PA 16802 USA. Contact him at akd6327@psu.edu.

EDITORS: Norita Ahmad, American University of Sharjah, nahmad@aus.edu
Preeti Chauhan, IEEE Reliability Society, preeti.chauhan@ieee.org

This article originally appeared in Computer vol. 58, no. 4, 2025

DEPARTMENT: DATA

Genomic Gold Rush or Ethical Minefield? Rethinking Data Practices in Health Tech Giants

Aqilah Julaihi, *Warwick Medical School*
Norita Ahmad, *American University of Sharjah*

Direct-to-consumer genomic testing offers unprecedented access to genetic insights but raises significant ethical challenges. Addressing these issues requires transparent data practices, stronger informed consent mechanisms, and ethical governance to ensure equitable and responsible use of genomic innovations.

"Genetics is not just about the genes we inherit but how we use them." This profound statement by Richard Dawkins[1] highlights the transformative potential of genetic information in reshaping human health and disease management. At the heart of this transformation lies genomic data, a comprehensive blueprint housed within every cell of the human body. This data, composed of approximately six billion DNA letters,[2] contains unique variations that can reveal critical insights into an individual's health, ancestry, and predisposition to disease.[3] Advances in technology such as artificial intelligence (AI) have turned this wealth of information into a valuable asset, powering breakthroughs in personalized medicine and precision health care.[4] Yet, the increasing commodification of genomic data by health-care systems and tech giants has brought ethical concerns to the forefront, particularly around privacy, ownership, and the equitable use of this sensitive information.[5]

The rapid growth of direct-to-consumer (DTC) genomic testing has made this once-exclusive knowledge broadly accessible, fostering what some describe as the "democratization" of genetic information. Companies like 23andMe, Ancestry.com, and Nebula Genomics have enabled millions of individuals to explore their genetic profiles, uncovering health risks, ancestry details, and personal traits from the comfort of their homes.[5] However, this ease of access comes with significant ethical challenges. As genomic data becomes increasingly commodified, it is often treated as a lucrative asset by corporations, raising concerns over privacy, data security, and informed consent.[6] Many consumers remain unaware of the risks associated with sharing their DNA, including data breaches, unauthorized usage, and the potential for genetic discrimination. Addressing these challenges requires robust ethical frameworks that prioritize transparency, protect user autonomy, and ensure equitable use of genomic resources.

PRIVACY AND OWNERSHIP OF GENOMIC DATA

Genomic data are deeply personal, offering insights into health, ancestry, and familial connections. However, when individuals submit their genetic information to DTC testing companies, they often unknowingly surrender control over this data. Companies like 23andMe have faced criticism for sharing anonymized genetic data with pharmaceutical firms without explicit consumer consent. This raises significant questions about who truly owns and controls genomic data.[7]

The issue of ownership lies at the heart of ongoing ethical debates. Should genomic data be considered

Digital Object Identifier 10.1109/MC.2025.3533264
Date of current version: 27 March 2025

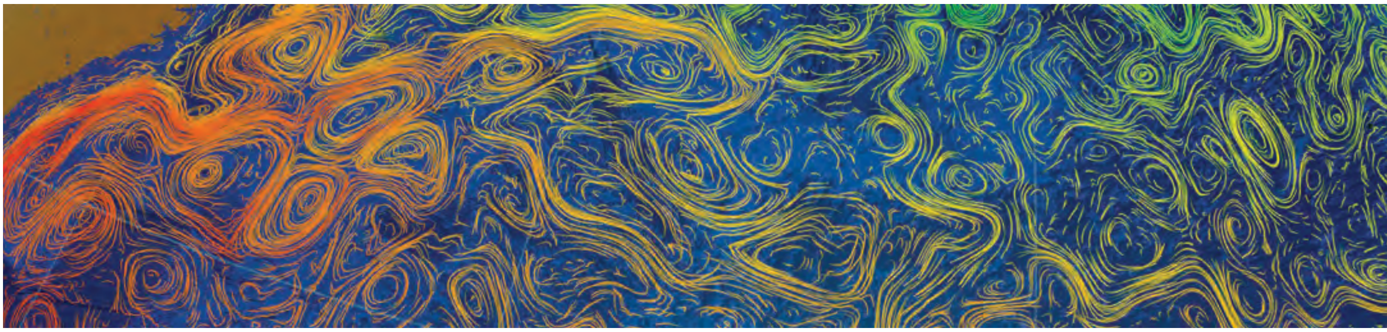


TABLE 1. Comparative overview of genomic data protections across major regions.

Region	Applicable Laws	Protections Offered	Gaps/Challenges
European Union	General Data Protection Regulation (GDPR)	Comprehensive data privacy for personal information, but genomic-specific gaps remain.	Does not explicitly address genomic data as unique; interpretation varies across jurisdictions.
United States	Genetic Information Non-Discrimination Act (GINA)	Protects against genetic discrimination in employment and health insurance.	Excludes life insurance, disability insurance, and other nonhealth-related uses.
Global	Various national and regional frameworks	Patchy protections; often lacks specificity for genomic data.	Inconsistent enforcement; no universal standard for genomic data handling.

personal property, or do companies have the right to commercialize it once submitted? The legal landscape surrounding this question is fragmented and inconsistent. Table 1 shows a comparative analysis of key data protection laws that highlights this disparity.

As illustrated in Table 1, gaps remain even in regions with robust data protection frameworks, such as the European Union's GDPR. For instance, genomic data are not explicitly categorized as unique personal information, leaving its interpretation to varying legal jurisdictions.[8] In the United States, GINA provides protections against genetic discrimination in employment and health insurance, yet it excludes critical areas like life and disability insurance, exposing individuals to potential misuse.[9]

These regulatory gaps have allowed companies to prioritize their commercial interests. For example, 23andMe's partnerships with GlaxoSmithKline raised concerns about transparency, as many users were unaware that their anonymized data could be sold to pharmaceutical companies for drug development.[7] This lack of informed consent underscores the need for stronger regulations and clearer definitions of ownership and control over genetic data.

Ethical concerns are particularly pressing for marginalized populations, such as lower-income or minority groups. These communities may have their genomic data disproportionately used for profit without fair representation or benefits. To address these concerns, experts have suggested treating genomic data as personal property, akin to intellectual property, allowing individuals to retain control over how their data are used, shared, or monetized.[5] This approach aligns with privacy laws like GDPR, which give individuals greater control over their personal data.

INFORMED CONSENT AND TRANSPARENCY

Informed consent is a cornerstone of ethical medical and research practices, yet it remains a significant challenge in the DTC genomic testing industry. Many companies employ complex, jargon-filled consent forms that obscure how genetic data will be collected, used, and shared. As a result, consumers often sign agreements without fully understanding the potential implications, including data sharing with third parties or the monetization of their information.[11]

To better understand this process and its challenges, it is helpful to visualize the lifecycle of genomic data in DTC testing. Figure 1 outlines the key stages, showing critical points where transparency and consent mechanisms are essential:

- *Data collection*: Genetic samples are collected and processed, often with minimal consumer awareness of how their information may be used.
- *Data storage*: Information is stored in databases, which may lack adequate security measures.

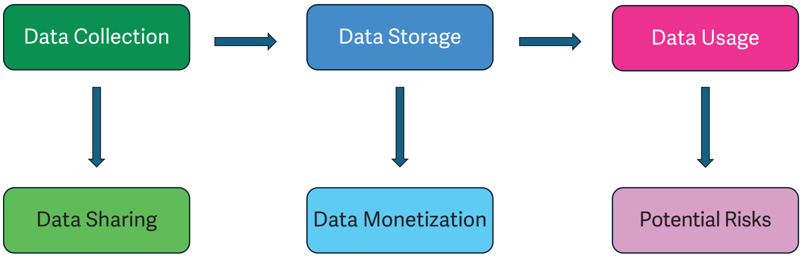


FIGURE 1. Lifecycle of genomic data in DTC testing.

- *Data usage*: Companies analyze data for health insights, ancestry reports, or other consumer-facing services.
- *Data sharing*: Genetic data may be shared with third parties, such as pharmaceutical companies, often without explicit user consent.
- *Data monetization*: Companies may profit from selling anonymized data, raising ethical concerns about ownership and autonomy.
- *Potential risks*: Risks include data breaches, re-identification of anonymized data, and misuse leading to genetic discrimination.

THIS LACK OF INFORMED CONSENT UNDERSCORES THE NEED FOR STRONGER REGULATIONS AND CLEARER DEFINITIONS OF OWNERSHIP AND CONTROL OVER GENETIC DATA.

This lifecycle highlights the importance of robust informed-consent mechanisms at every stage. True informed consent should empower consumers to make decisions based on clear, accessible information about how their genetic data will be handled.[11] Companies should adopt practices that prioritize simplicity and transparency,[12] such as simplifying legal jargon to ensure consumers can easily understand what they are agreeing to, using diagrams or infographics to illustrate data flows and potential uses, allowing users to modify their consent preferences over time, such as opting in or out of specific uses or research projects, and keeping consumers informed of new developments, such as changes to how their data are stored or shared.

Blockchain technology offers promising solutions for enhancing transparency and data security. For example, companies like Nebula Genomics use blockchain protocols to give users more control over data access and sharing. This technology enables an auditable record of who accesses data and for what purpose. However, critics argue that blockchain alone cannot address challenges like ensuring users fully comprehend their rights or the irreversible nature of some data-sharing agreements.[10]

Transparency is also crucial for fostering trust in the DTC genomic testing industry. Consumers must have confidence that their data are being handled ethically and securely. Without clear and accessible consent mechanisms, public trust in these services could erode, limiting participation and undermining the potential benefits of genomic innovation.

MISUSE AND POTENTIAL HARM

The DTC genomic testing industry presents a range of ethical challenges, particularly concerning the misuse of data and its potential harm to individuals. Genetic data are profoundly personal, revealing sensitive information about not only the individual but also their family members.[13] These insights, while invaluable for advancing personalized medicine, can also lead to significant risks if misused.

One of the most pressing concerns is the potential for privacy breaches. Data breaches, such as the 2018 MyHeritage incident that exposed information on over 92 million users, show the vulnerabilities in current genomic data storage practices.[14] Even anonymized data are not immune to reidentification, as advanced algorithms can cross-reference datasets to deduce personal identities. Breaches like these erode consumer trust and expose individuals to unexpected risks.

The commercialization of genomic data raises complex ethical questions. Many DTC companies share or sell anonymized genetic data to pharmaceutical companies and other third parties without obtaining explicit user consent. While this practice supports drug development and other research, it often occurs without adequate transparency, leaving consumers unaware of how their data are monetized.[11] This commodification of genetic information shifts control away from the individual, creating imbalances in data ownership and benefit distribution.

Another significant risk is genetic discrimination. Employers, insurers, or government entities could misuse genetic data to make decisions that disadvantage individuals. For example, an insurer might adjust premiums or deny coverage based on a customer's genetic predisposition to certain diseases.[15] While laws like the GINA offer some protections, they are often limited in scope, excluding areas like life and disability insurance. This regulatory gap leaves individuals vulnerable to exploitation.[9]

The psychological effects of genomic testing can also be profound. Learning about a predisposition to severe or incurable conditions may cause anxiety, depression, or feelings of helplessness. Without proper counseling and clear communication of results, individuals may misinterpret their risk levels, leading to unnecessary health interventions or distress.[16] This issue is especially evident in cases where DTC companies fail to provide adequate support resources for interpreting results. Marginalized populations, including low-income or minority groups, are particularly at risk of exploitation in the genomic testing industry. These communities may lack the resources or education to fully understand the implications of submitting their genetic data.[16] Consequently, they may be disproportionately targeted for data collection without receiving equitable benefits from the resulting advancements.

There are several measures that can be adopted to address these challenges. For example, governments should consider expanding existing legal frameworks to include genomic data protections, ensuring equitable treatment and reducing the risk of misuse. Companies must also clearly communicate how genetic data will be used, stored, and shared, empowering consumers to make informed decisions. Additionally, they should also simplify consent forms and providing ongoing updates about data usage. Finally, both governments and companies should engage marginalized communities in discussions about genomic testing and its implications in order to ensure trust and equitable participation.

IMPLICATIONS FOR PUBLIC TRUST

The rapid expansion of the DTC genomic testing industry has brought ethical concerns into sharp focus, particularly regarding its impact on public trust. Trust is essential for the continued success and growth of the industry, as it ensures consumer participation and the responsible use of genomic data. However, the lack of transparency in data usage, insufficient informed consent mechanisms, and frequent privacy breaches have eroded consumer confidence.[17] When companies fail to clearly disclose how they collect, store, and share genetic data, they increase the risk of misuse and discrimination. For instance, fears about data being sold to third parties or used for purposes beyond what was initially agreed upon are common among consumers.[11] Addressing these concerns requires companies to adopt robust transparency measures, such as detailed consent processes, regular updates on data usage, and clear communication about security protocols.

> *BY ALIGNING TECHNOLOGICAL ADVANCEMENTS WITH THESE STANDARDS, THE DTC GENOMIC TESTING INDUSTRY CAN SAFEGUARD CONSUMER RIGHTS WHILE ACHIEVING ITS POTENTIAL.*

The relationship between informed consent and trust is equally critical. Consumers need to fully understand the implications of sharing their genetic data, including potential emotional and psychological risks. Without accessible and user-friendly consent mechanisms, consumers may feel coerced into agreements they do not fully comprehend. This lack of clarity can lead to harmful consequences and further diminish trust in the industry. Furthermore, public trust hinges on the industry's ability to protect the privacy and security of genetic data. High-profile data breaches, such as the MyHeritage incident, have highlighted vulnerabilities in existing security frameworks.[14] Companies must prioritize strong data protection measures and communicate their efforts transparently to reassure consumers that their information is secure.

Corporate accountability and adherence to ethical standards also play a key role in fostering trust. Exaggerated claims or unsupported predictions in genetic testing can mislead consumers, damaging the industry's credibility. Regulatory oversight can help ensure that companies meet established scientific standards and avoid practices that exploit consumers.[17] Ultimately, building and maintaining public trust requires a collaborative approach. Policymakers, technologists, health-care providers, and ethicists must work together to establish ethical guidelines and regulatory frameworks. By aligning technological advancements with these standards, the DTC genomic testing industry can safeguard consumer rights while achieving its potential.

TOWARDS ETHICAL GENOMIC INNOVATION

Genomic data holds the potential to revolutionize human health and disease management, offering insights that were once inconceivable. DTC genomic testing has democratized access to this information, enabling millions to uncover details about their ancestry, health risks, and personal traits. However, with these advancements come significant ethical challenges, such as privacy risks, informed consent issues, and the erosion of public trust.

As Richard Dawkins famously noted, *"Genetics is not just about the genes we inherit but how we use them."* This sentiment underscores the dual responsibility of using genomic data for progress while safeguarding its ethical use. Genomic data are more than just a scientific resource; it represents an intimate map of human identity.[3] Addressing the ethical concerns associated with its use requires balancing innovation with accountability. The commodification of genetic information underscores the need for transparency, robust governance, and equitable practices to ensure that advancements in genomics benefit society as a whole.

Emerging technologies like blockchain and AI present opportunities to enhance data security and improve personalization. However, these tools must be complemented by strong regulatory frameworks and ethical oversight. Collaboration among stakeholders, including policymakers, researchers, and industry leaders is vital to establishing standards that protect individual rights while fostering innovation.

In conclusion, the future of the DTC genomic testing industry depends on its ability to address ethical concerns proactively. By establishing a foundation of trust, transparency, and accountability, the industry can continue to innovate responsibly, unlocking the transformative potential of genomic data while safeguarding individual and societal well-being.

ACKNOWLEDGMENT

Norita Ahmad is the corresponding author.

REFERENCES

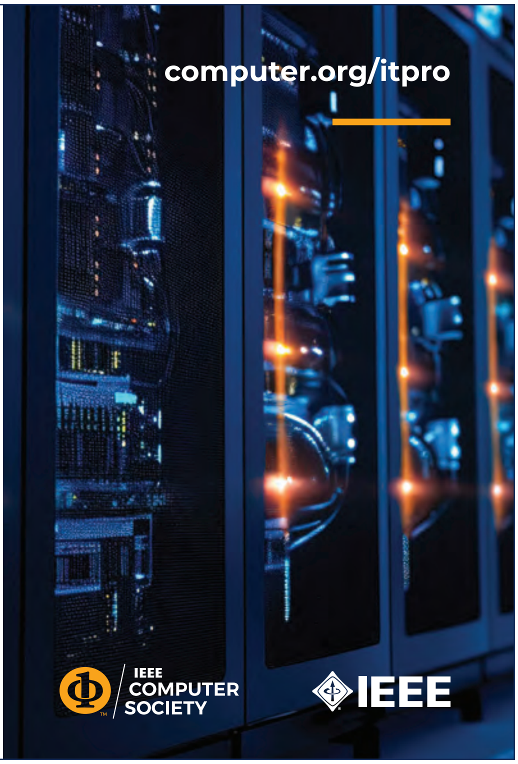
1. R. Dawkins, *The Selfish Gene*. Oxford, U.K.: Oxford Univ. Press, 1976.
2. S. Kim-Hellmuth et al., "Cell type–specific genetic regulation of gene expression across human tissues," *Science*, vol. 369, no. 6509, Sep. 2020, doi: 10.1126/science.aaz8528.
3. M. J. Khoury, "The shift from personalized medicine to precision medicine and precision public health: Words matter!," *CDC Genomics and Precision Health Blog*, Apr. 21, 2016. [Online]. Available: https://blogs.cdc.gov/genomics/2016/04/21/shift/
4. Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of big data – Evolution, challenges and research agenda," *Int. J. Inf. Manage.*, vol. 48, no. 1, pp. 63–71, Oct. 2019, doi: 10.1016/j.ijinfomgt.2019.01.021.
5. J. Moran, "Privacy perspectives on direct-to-consumer genetic testing in the era of big data: Role of blockchain technology in genomics," *Tulane J. Technol. Intellectual Property*, vol. 22, pp. 185–204, Spring 2020.
6. B. Berger and H. Cho, "Emerging technologies towards enhancing privacy in genomic data sharing," *Genome Biol.*, vol. 20, no. 1, Jul. 2019, Art. no. 128, doi: 10.1186/s13059-019-1741-0.
7. J. Ducharme, "A major drug company now has access to 23and-Me's genetic data. Should you be concerned?" *Time*, Jul. 26, 2018. [Online]. Available: https://time.com/5349896/23andme-glaxo-smith-kline/
8. K. Pormeister, "Genetic data and the research exemption: Is the GDPR going too far?" *Int. Data Privacy Law*, vol. 7, no. 2, pp. 137–146, May 2017, doi: 10.1093/idpl/ipx006.
9. Y. Joly, C. Dupras, M. Pinkesz, S. A. Tovino, and M. A. Rothstein, "Looking beyond GINA: Policy approaches to address genetic discrimination," *Annu. Rev.*

Genomics Human Genetics, vol. 21, no. 1, pp. 491–507, Aug. 2020, doi: 10.1146/annurev-genom-111119-011436.

10. D. Hofman and A. Novin, “Blocked and chained: Blockchain and the problems of transparency,” in *Proc. Assoc. Inf. Sci. Technol.*, 2018, vol. 55, no. 1, pp. 171–178, doi: 10.1002/pra2.2018.14505501019.
11. A. E. Raz, E. Niemiec, H. C. Howard, S. Sterckx, J. Cockbain, and B. Prainsack, “Transparency, consent and trust in the use of customers’ data by an online genetic testing company: An exploratory survey among 23andMe users,” *New Genetics Soc.*, vol. 39, no. 4, pp. 1–24, May 2020, doi: 10.1080/14636778.2020.1755636.
12. L. M. Beskow, and K. P. Weinfurt, “Exploring understanding of ‘understanding’: The paradigm case of biobank consent comprehension,” *Amer. J. Bioethics*, vol. 19, no. 5, pp. 6–18, May 2019, doi: 10.1080/15265161.2019.1587031.
13. X. Shi and X. Wu, “An overview of human genetic privacy,” *Ann. N.Y. Acad. Sci.*, vol. 1387, no. 1, pp. 61–72, Jan. 2017, doi: 10.1111/nyas.13211.
14. Reuters Staff, “Security breach at MyHeritage website leaks details of over 92 million users,” *Reuters*, Jun. 5, 2018. [Online]. Available: https://www.reuters.com/article/business/security-breach-at-myheritage-website-leaks-details-of-over-92-million-users-idUSKCN1J1301/
15. C. D. Zick, C. J. Mathews, J. S. Roberts, R. Cook-Deegan, R. J. Pokorski, and R. C. Green, “Genetic testing for Alzheimer’s disease and its impact on insurance purchasing behavior,” *Health Affairs*, vol. 24, no. 2, pp. 483–490, Mar. 2005, doi: 10.1377/hlthaff.24.2.483.
16. J. Collmann and S. A. Matei, *Ethical Reasoning in Big Data an Exploratory Analysis*. Cham, Switzerland: Springer-Verlag, 2016.
17. M. Majumder, C. Guerrini, and A. Mcguire, “Annual review of medicine direct-to-consumer genetic testing: Value and risk,” *Annu. Rev. Med.*, vol. 72, no. 1, pp. 151–166, 2020, doi: 10.1146/annurev-med-070119-114727.

AQILAH JULAIHI is a master of public health candidate at the University of Warwick, CV4 7AL Coventry, U.K. Contact her at aqilaanis@gmail.com.

NORITA AHMAD is a professor of information systems and business analytics at the American University of Sharjah, Sharjah, United Arab Emirates. Contact her at nahmad@aus.edu.



EDITOR: Lakmal Meegahapola, lmeegahapola@ethz.ch

DEPARTMENT: EMERGING ROCKSTAR

This article originally appeared in IEEE Pervasive Computing, vol. 23, no. 3, 2024

Justin Chan: Intelligent Mobile Systems for Equitable Healthcare

Lakmal Meegahapola, *ETH Zurich, 8092, Zurich, Switzerland*

The "Emerging Rockstar" segment in *IEEE Pervasive Computing* highlights rising stars in the field of pervasive computing through captivating interviews. In the series' third article, Dr. Lakmal Meegahapola, a member of the *IEEE Pervasive Computing* editorial board, conducts an interview with Dr. Justin Chan, an Assistant Professor at the Carnegie Mellon University.

From the Editor

IEEE Pervasive: Could you provide an overview of your research, and what inspired you to pursue this research direction?

Justin Chan: In the global context, the country in which you are born has a huge impact on your ability to access basic medical resources. Within the USA, the zip code that you live in can markedly affect healthcare quality and life expectancy. In this setting, my research focuses on building intelligent mobile and embedded systems for equitable healthcare.[1,2,3,4,5,6] This involves designing novel methods that can exploit the sensing capabilities of smart or wearable devices around us in real time for medical diagnostics. By leveraging the ubiquity of commodity devices, mobile systems are able to scale and significantly increase access to healthcare. What drew me toward this line of research is how innovations in computing that advance healthcare delivery for even a single condition can have a profound impact on millions of people. While deploying my research on low-cost newborn hearing screening with smart devices in Kenya, I was struck by how an idea that began in a lab due to my curiosity about how ears worked was now seen by the Nairobi Ministry of Health as having the potential to substantially change the life trajectory of millions of yet-to-be-born children.

Pervasive: Can you please explain a bit more about the main challenges you face in your research?

Chan: Inventing intelligent mobile systems for healthcare is challenging for three key reasons. First, unlike conventional medical devices, which are created for a single piece of calibrated hardware, smart devices are not designed for medical diagnostics. Furthermore, the sensors and computing power differ from one smart device to another. Second, medical diagnostics typically rely on expensive and sensitive sensors. To achieve equitable healthcare, the challenge is to leverage low-cost commodity hardware while still meeting the high standards of clinical performance expected of medical devices. Third, for these systems to scale across different devices and environments, it is often necessary to collect large amounts of diverse data for development and testing. However, dataset collection and curation can be costly and difficult, especially as new devices continue to be introduced to the market. My research toolkit to tackle these challenges spans:

1) wireless sensing and applied machine learning techniques that can generalize across different hardware;
2) hardware–software co-design to achieve both high clinical accuracy and low cost;
3) dataset collection and augmentation methods to scale systems across a large number of devices and environments.

Pervasive: You have also cofounded a company, Wavely Diagnostics. Can you discuss your journey as a startup founder?

Chan: Wavely Diagnostics is a company I cofounded, which is commercializing my work on detecting ear infections with smartphones, with the goal of getting this technology into the hands of millions of people. The technology leverages the speakers

1536-1268 © 2024 IEEE
Digital Object Identifier 10.1109/MPRV.2024.3433949
Date of current version 16 October 2024.

Justin Chan is an Assistant Professor in the School of Computer Science at Carnegie Mellon University. His work on smartphone-based ear infections is now FDA-listed and available to select early-access healthcare systems. His research on newborn hearing screening has led to an international effort called TUNE, which aims to bring universal newborn hearing screening to Kenya. He was also a lead contributor to CovidSafe (now WA Notify), a COVID-19 contact tracing app, which became part of the official efforts by the WA Department of Health to manage the pandemic. He has authored publications in interdisciplinary journals, such as *Nature Biomedical Engineering*, *Science Translational Medicine*, and *Nature Communications*, as well as in computer science and engineering venues, such as MobiSys, MobiCom, SIGCOMM, SIGGRAPH Asia, and UIST. He has received his Ph.D. degree from the University of Washington, and his Bachelor's degree with high honors from Dartmouth. He was also named the runner-up for the ACM SIGMOBILE Doctoral Dissertation Award in 2024.

and microphones on a smartphone to probe and detect ear infections using no additional attachments beyond a paper cone.[1] Initially, the process of working with the FDA seemed daunting due to stories of having to run large and lengthy clinical studies. However, after meeting with the FDA, it turned out that many of their requirements were reasonable and doable. They wanted the device to be tested across a range of demographics, including age, race, gender, and ear conditions, to ensure that the system adapts to differences in ear anatomy and generalizes to a wide population. They also wanted to ensure that the code was well documented and tested in various edge cases, such as the user attaching the paper cone incorrectly or testing it in a loud environment. These experiences were invaluable and deepened my understanding of building robust systems, which I have been able to bring back to my research.

Pervasive: You have conducted several studies in Kenya regarding low-cost newborn hearing screening. Can you tell us more about your efforts in Kenya and what led you to take your research there?

Chan: In high-income countries, such as the USA, every child gets screened for hearing at birth. However, in many low- and middle-income countries, such as Kenya, there is little to no screening for hearing loss. As a result, children are often diagnosed with hearing loss much later, negatively affecting language acquisition and neurodevelopment. A major challenge preventing this screening is the high cost of hearing screening devices that rely on sensitive components to probe the cochlea and listen for the faint sounds caused by the tiny vibrations of the inner ear hair cells. We invented low-cost systems for healthcare workers to perform hearing screening at orders of magnitude lower cost using a $10 smartphone probe[2] or a wireless earbud device,[3] which achieve comparable accuracies to the expensive, conventional screening devices. This has led to a larger international effort, TUNE, which is working across multiple organizations with the goal of bringing universal newborn hearing screening to Kenya. We have been closely engaged with our partner clinics in Nairobi, and we have been running the clinical studies needed to instigate changes in public health policy necessary for large-scale deployment.

Pervasive: Finally, what advice would you offer to upcoming researchers interested in pursuing a career in a similar research field as yours, and what are your future goals?

AS AN OUTSIDER, IT IS OFTEN EASIER TO SPOT GAPS IN CONVENTIONAL WISDOM AND INVENT SOLUTIONS THAT MEANINGFULLY ADVANCE BOTH FIELDS.

Chan: Do not let a lack of experience in a new field hold you back from diving deep and making a unique contribution. As an outsider, it is often easier to spot gaps in conventional wisdom and invent solutions that meaningfully advance both fields. The secret to giving your research the best chance of societal impact is to partner closely with open-minded domain experts in designing a practical technology that has a clear translational pathway from lab prototype to production-ready system. About my future goals, we live in a unique and exciting time for the future of healthcare. The pace of technological innovation is rapid, with new hardware platforms, sensing capabilities, and machine learning models being released seemingly daily, offering the potential to positively impact the healthcare system. At the same time, even modern healthcare systems, such as those in the USA face accessibility challenges, with wait times for a physician appointment reaching 1–2 months in some major cities. My vision for the future is to develop systems that enable every human being on the planet to have access to basic healthcare at their fingertips.

REFERENCES

1. J. Chan, S. Raju, R. Nandakumar, R. Bly, and S. Gollakota, “Detecting middle ear fluid using smartphones,” *Sci. Transl. Med.*, vol. 11, no. 492, 2019, Art. no. eaav1102.
2. J. Chan et al., “An off-the-shelf otoacoustic-emission probe for hearing screening via a smartphone,” *Nature Biomed. Eng.*, vol. 6, no. 11, pp. 1203–1213, 2022.
3. J. Chan et al., “Wireless earbuds for low-cost hearing screening,” in *Proc. 21st Annu. Int. Conf. Mobile Syst., Appl. Serv.*, 2023, pp. 84–95.
4. U. of Washington, “TUNE: Towards universal newborn and early childhood hearing screening in Kenya,” 2022.
5. J. Chan, T. Rea, S. Gollakota, and J. E. Sunshine, “Contactless cardiac arrest detection using smart devices,” *NPJ Digit. Med.*, vol. 2, no. 1, 2019, Art. no. 52.
6. J. Chan, K. Michaelsen, J. K. Estergreen, D. E. Sabath, and S. Gollakota, “Micro-mechanical blood clot testing using smartphones,” *Nature Commun.*, vol. 13, no. 1, 2022, Art. no. 831.

LAKMAL MEEGAHAPOLA is a postdoctoral researcher at ETH Zurich, 8092, Zurich, Switzerland. He earned his Ph.D. in electrical engineering from EPFL, Lausanne, Switzerland. His research interests lie at the intersection of mobile and wearable sensing, digital health, machine learning, deep learning, and human-computer interaction. Contact him at lmeegahapola@ethz.ch.

E-Currency and Fairness ▪ Ransomware Defense ▪ A National Cybersecurity Policy
IEEE SECURITY & PRIVACY
Building Dependability, Reliability, and Trust
Software and Cybersecurity ▪ Big Data: Privacy Versus Accessibility ▪ Resiliency in Cloud Computing
IEEE SECURITY & PRIVACY
DIGITAL FORENSICS PART 2
The NICE Challenge Project ▪ Improving Data Breach Notifications ▪ AI Models
IEEE SECURITY & PRIVACY
Building Dependability, Reliability, and Trust

stay connected.

Conference Calendar

IEEE Computer Society conferences are valuable forums for learning on broad and dynamically shifting topics from within the computing profession. With over 200 conferences featuring leading experts and thought leaders, we have an event that is right for you. Questions? Contact conferences@computer.org.

MARCH

6 March

- WACV (IEEE/CVF Winter Conf. on Applications of Computer Vision), Tucson, USA

16 March

- PerCom (IEEE Int'l Conf. on Pervasive Computing and Communications), Pisa, Italy

17 March

- SANER (IEEE Int'l Conf. on Software Analysis, Evolution and Reengineering), Limassol, Cyprus

20 March

- 3DV (Int'l Conf. on 3D Vision), Vancouver, Canada

21 March

- VR (IEEE Conf. on Virtual Reality and 3D User Interfaces), Daegu, Korea

22 March

- SSIAI (IEEE Southwest Symposium on Image Analysis and Interpretation), Santa Fe, USA

23 March

- SaTML (IEEE Conf. on Secure and Trustworthy Machine Learning), Munich, Germany

APRIL

3 April

- CI2A (Int'l Conf. on Connected Intelligence for Industrial Applications), Punjab, India

12 April

- AST (IEEE/ACM Int'l Conf. on Automation of Software Test), Rio de Janeiro, Brazil
- CAIN (IEEE/ACM Int'l Conf. on AI Eng. – Software Eng. for AI), Rio de Janeiro, Brazil
- FormaliSE (IEEE/ACM Int'l Conf. on Formal Methods in Software Eng.), Rio de Janeiro, Brazil
- ICSE (IEEE/ACM Int'l Conf. on Software Eng.), Rio de Janeiro, Brazil
- MOBILESoft (IEEE/ACM Int'l Conf. on Mobile Software Eng. and Systems), Rio de Janeiro, Brazil
- MSR (IEEE/ACM Int'l Conf. on Mining Software Repositories), Rio de Janeiro, Brazil

15 April

- COOL CHIPS (IEEE Symposium on Low-Power and High-Speed Chips and Systems), Tokyo, Japan

20 April

- DATE (Design, Automation & Test in Europe Conf.), Verona, Italy
- PacificVis (IEEE Pacific Visualization Conf.), Sydney, Australia

26 April

- ISPASS (IEEE Int'l Symposium on Performance Analysis of Systems and Software), Seoul, Korea

27 April

- VTS (IEEE VLSI Test Symposium), Napa, USA

MAY

4 May

- HOST (IEEE Int'l Symposium on Hardware Oriented Security and Trust), Washington, DC, USA
- MOST (IEEE Int'l Conf. on Mobility, Operations, Services and Technologies), Detroit, USA

8 May

- BigDataSecurity (IEEE Conf. on Big Data Security on Cloud), New York City, USA
- CAI (IEEE Int'l Conf. on Artificial Intelligence), Granada, Spain
- HPSC (IEEE Int'l Conf. on High Performance and Smart Computing), New York City, USA
- IDS (IEEE Int'l Conf. on Intelligent Data and Security), New York City, USA
- SmartCloud (IEEE Int'l Conf. on Smart Cloud), New York City, USA

11 May

- SenSys (ACM/IEEE Int'l Conf. on Embedded Artificial

Intelligence and Sensing Systems), Saint Malo, France

12 May

- RTAS (IEEE Real-Time and Embedded Technology and Applications Symposium), Saint Malo, France

13 May

- FCCM (IEEE Annual Int'l Symposium on Field-Programmable Custom Computing Machines), Atlanta, USA

15 May

- ICES (Int'l Conf. on Energy Storage), Shenyang, China

18 May

- CCGrid (IEEE Int'l Symposium on Cluster, Cloud and Internet Computing), Sydney, Australia
- ICFEC (IEEE Int'l Conf. on Fog and Edge Computing), Sydney, Australia
- ICST (IEEE Int'l Conf. on Software Testing, Verification and Validation), Daejeon, Korea
- S&P (IEEE Symposium on Security and Privacy), San Francisco, USA

19 May

- ICDE (IEEE Int'l Conf. on Data Eng.), Hong Kong, China
- ISMVL (IEEE Int'l Symposium on Multiple-Valued Logic), Sendai, Japan

25 May

- FG (IEEE Int'l Conf. on Automatic Face and Gesture Recognition), Kyoto, Japan
- IPDPS (IEEE Int'l Parallel and Distributed Processing Symposium), New Orleans, USA

JUNE

1 June

- ICHI (IEEE Int'l Conf. on Healthcare Informatics), Minneapolis, USA

3 June

- CBMS (IEEE Int'l Symposium on Computer-Based Medical Systems), Limassol, Cyprus
- CVPR (IEEE/CVF Conf. on Computer Vision and Pattern Recognition), Denver, USA

10 June

- SVCC (Silicon Valley Cybersecurity Conf.), San Jose, USA

16 June

- WoWMoM (IEEE Int'l Symposium on a World of Wireless, Mobile and Multimedia Networks), Bologna, Italy

22 June

- DCOSS-IoT (Int'l Conf. on Distributed Computing in Smart Systems and the Internet of Things), Reykjavik, Iceland
- DSN (Annual IEEE/IFIP Int'l Conf. on Dependable Systems and Networks), Charlotte, USA
- ICDCS (IEEE Int'l Conf. on Distributed Computing Systems), Seoul, Korea
- ICSA (IEEE Int'l Conf. on Software Architecture), Amsterdam, Netherlands

26 June

- IEEE Cloud Summit, Washington, DC, USA

27 June

- ISCA (ACM/IEEE Annual Int'l Symposium on Computer Architecture), Raleigh, USA

JULY

6 July

- EuroS&P (IEEE European Symposium on Security and Privacy), Lisbon, Portugal
- ICALT (IEEE Int'l Conf. on Advanced Learning Technologies), Hung Yen, Vietnam

7 July

- COMPSAC (IEEE Annual Computers, Software, and Applications Conf.), Madrid, Spain
- ISVLSI (IEEE Computer Society Annual Symposium on VLSI), Kolkata, India