

Editorial

From Silicon to Society: Efficient, Secure, and Responsible Computing in the Age of Artificial Intelligence

Guest Editor: Ivan Ratković

From Silicon to Society: Efficient, Secure, and Responsible Computing in the Age of Artificial Intelligence

Guest Editor: Ivan Ratković

The progress of artificial intelligence (AI) is no longer determined by algorithms and models alone. It depends equally on the efficiency of the hardware and algorithms that execute computational workloads, on the security of models once they are deployed, and on the ability of organizations to integrate AI responsibly into everyday practice. This special issue brings together contributions that address these challenges at different levels - from post-Moore computing substrates and the confidentiality of trained neural networks to AI-supported clinical workflows, the use of generative AI in higher education, the data-driven operation of electric power grids on the path to decarbonization, and lightweight algorithms for large-scale genome analysis.

Although the papers belong to distinct research fields, they share a common perspective: rather than evaluating an isolated component, they examine how technology behaves as part of a larger system. They show how total energy and data movement determine the efficiency of AI hardware, how the outputs of a deployed model can reveal the model itself, how physics-based models and machine learning can be combined to manage uncertainty in the power grid, how an algorithm that relies only on intrinsic sequence signals can make genome analysis fast and independent of external databases, and how data, analytics, institutional policies, and human judgment must be coordinated for AI to deliver real value in hospitals and universities. By combining formal, experimental, and empirical approaches, the issue aims to stimulate dialogue between the hardware, security, and application communities.

1. The paper **“Workflow Systems and Intelligent Data Integration for Personalized Therapeutics in Brain Disorders”** by Daniel A. Ardila-Herrera, Laura Dipietro, Ciro Ramos-Estebanez, Paulo E. P. Teixeira, and Tim Wagner addresses personalized therapeutics in brain disorders from an information-systems perspective. The authors observe that sensing technologies and AI have advanced faster than the clinical information systems and care workflows expected to use them, so that sensing devices, analytic tools, and decision-support systems often remain only partially connected. In response, they propose a workflow-integrated framework that organizes therapeutic care into four phases - data capture, analytic processing, decision integration, and feedback with continuous adaptation - implemented through a layered architecture in which a workflow orchestration layer coordinates data acquisition, analytics, and the clinical interface. The framework is illustrated through the Integrated Motion Analysis Suite (IMAS), with published case studies in stroke rehabilitation and Parkinson’s disease; in a cohort of 28 stroke patients, for example, the system predicted Fugl-Meyer scores and motor recovery potential with R^2 values of 0.75 and 0.83, respectively. Opioid use disorder and chronic pain are discussed as ongoing extension cases. By also examining cloud and federated deployment, interoperability, governance, and barriers to clinical adoption, the paper makes a convincing case that workflow orchestration is a key information-systems function for translating multimodal data and AI into clinical use.

2. The paper **“CopyCNN: Reverse Engineering Based Extraction of Convolutional Neural Networks”** by Yehonatan Lusky, Avi Mendelson, and Yaniv Nemcovsky addresses model extraction, a growing threat to the security and intellectual property of AI systems. The authors present CopyCNN, an attack that recovers both the architecture and the precise weights of black-box, ReLU-based convolutional neural networks using only query access - to their knowledge, the first method capable of recovering exact weights from such networks. Building on cryptanalytic extraction techniques previously demonstrated for multilayer perceptrons, which exploit the critical points of piecewise-linear ReLU networks, the method introduces dedicated strategies for convolutional, residual, batch-normalization, and pooling layers, and

combines graph-based weight correlation and unification with mixed-integer linear programming for global sign recovery. The architecture itself is inferred by testing candidate layer configurations and checking whether the recovered weights form coherent clusters. Experiments demonstrate the successful extraction of two LeNet-5 variants and a ResNet block, as well as higher extraction precision than prior methods on multilayer perceptrons, often with fewer queries. The authors openly discuss the current limitations - accumulated numerical error in deeper networks, the need for access to logits, and the large number of adaptive queries - and outline countermeasures such as query monitoring, label-only outputs, input randomization, and moving-target defenses. For designers and operators of AI systems, the paper is a timely reminder that the weights of a deployed model may be far less private than is commonly assumed.

3. The paper **“The MKA Law and the Milutinović–Katulevskiy Architecture: A Post-Moore Graph-Fabric Roadmap for AI-Scale, Photonic, and Quantum-Adjacent Computing”** by Daniil Katulevskiy, Elizaveta Katulevskaya, Aleksey Katulevskiy, and Veljko Milutinović argues that the next major efficiency jump in computing will not come from device density alone, but from jointly reducing the total task energy and the effective graph depth of workloads executed at fixed quality. The authors formalize this idea through a fixed-workload graph acceleration density metric and the MKA Law, and identify a Compound Gain Regime in which reductions in energy and depth reinforce each other multiplicatively rather than being traded against one another. Building on the Honeycomb architecture lineage originated by Veljko Milutinović, the paper defines a substrate-independent graph fabric spanning six regimes: reconfigurable FPGA/CGRA overlays, analog fabrics, processing-in-memory and memristive fabrics, silicon-photonics fabrics, quantum-dot substrates, and entanglement-assisted quantum/hybrid graph edges. A recently reported analog circuit that solves symmetric positive-definite linear systems in time independent of matrix size anchors the analog regime, while large language model (LLM) training and inference, modeled as token-flow graphs, serve as the first industrial test case, including a conditional projection of billion-dollar-class effects for AI data centers. Particularly valuable from a computer-architecture perspective is the insistence on full-task energy accounting - covering data movement, control, conversion, calibration, and cooling rather than isolated energy per operation - together with explicit falsification criteria, a benchmark suite, and a staged research agenda that make the roadmap testable.

4. The paper **“Students’ Awareness, Perceptions, and Use of Generative AI Tools in Higher Education”** by Davorka Radaković, William Steingartner, and Dora Seleši presents an international empirical study of how students use and perceive generative AI in their studies. A structured questionnaire was administered in two phases to students at five universities in Poland, Slovakia, Slovenia, and Serbia, yielding 227 consenting respondents, and the data were analyzed with considerable statistical care, including confidence intervals, nonparametric comparisons, effect sizes, correction for multiple comparisons, and sensitivity analyses adjusted for recruitment phase and institution. The results show near-universal adoption: 94.3% of respondents had used AI tools, most of them several times a week, primarily for problem solving and programming. Positive perceptions of AI as a learning aid coexist with widespread experience of its errors - almost all users had encountered incorrect output, yet most still reported at least some trust in AI tools, pointing to conditional reliance rather than uncritical confidence. The study also reveals a clear governance gap: only about a quarter of AI users knew that their institution had AI-use guidelines, whereas most supported formal policies and explicit instruction in ethical use, and nearly half reported using AI to generate complete homework tasks. Gender differences proved limited once multiple comparisons were accounted for. The authors conclude that universities should respond with clear guidance, AI-literacy training, verification practices, and responsible-use policies rather than with simple promotion or prohibition of AI tools.

5. The paper **“ForiC - a Fast and Effective Method for Predicting the Origin of Bacterial Chromosome Replication”** by Marko Stojićević addresses a fundamental task in computational genomics: locating the origin of replication (oriC) in bacterial chromosomes. Classical GC-skew and cumulative GC-skew (CGC) analyses are computationally simple but often imprecise, whereas more accurate approaches based on DnaA-box detection and comparative genomics are resource-intensive and depend on external biological information. ForiC bridges this gap with a deterministic, linear-time rule that combines three signals derived directly from the genome sequence - the CGC global maximum, the CGC inflection point associated with it, and the alignment of that point with the GC global minimum - without motif detection, homology searches, or external databases. Calibrated and evaluated on 36 bacterial genomes with experimentally validated origins, using five phylum-stratified training/test splits, ForiC reached a hit accuracy of 90.91% on a representative test subset with a two-window reporting interval, clearly outperforming the GC and CGC skew baselines. Compared with the reference tool OriFinder, which achieves higher positional

precision but often reports multiple candidate regions and relies on BLAST searches, ForiC always returns a single, directly interpretable prediction, remained comparatively stable when oriC-associated regions were removed, and ran in 0.75 seconds versus about 28 minutes in the reported example. Positioned by the author as a screening-oriented tool that complements rather than replaces reference-assisted methods, ForiC is particularly attractive for rapid, large-scale, database-independent genome analysis.

Even though the papers in this issue address different domains and levels of abstraction, they converge on a common conclusion: the impact of AI and other computational methods is determined not only by the capabilities of models and algorithms, but also by the efficiency with which they are executed, the protection of the knowledge they embody, their grounding in the physics of the infrastructure they help to operate, and the workflows, policies, and competencies that govern their use. We hope that readers will find these contributions useful and that they will inspire further research across disciplinary boundaries. We sincerely thank all the authors for their contributions and the reviewers for their careful and constructive evaluations.

===== About guest editor =====

Ivan Ratković, PhD, is a computer architect with twenty years of experience in designing, architecting, developing, and taping out energy-efficient multicore processors and hardware accelerators for artificial intelligence (AI) and high-performance computing (HPC). He is a Chief Architect and CTO of a company developing area- and power-efficient HPC/AI processors, where he leads architecture, design, and verification, with a special focus on the dot-product unit, the RISC-V IOMMU, and the RISC-V vector processing unit. Previously, at Esperanto Technologies, he served as Power Manager and Architect, leading all power- and energy-efficiency efforts - from applications and the operating system down to silicon - for a RISC-V system-on-chip for HPC and AI. Earlier, as a member of the company's senior VLSI technical staff, he worked on vector/tensor unit design, power estimation and optimization, low-power and full-chip verification, gate-level sign-off, and silicon validation, contributing to the successful tape-out of a 24-billion-transistor chip with more than a thousand RISC-V vector cores.

Dr. Ratković received his PhD in Computer Architecture, cum laude, from the Universitat Politècnica de Catalunya (UPC-BarcelonaTech), within a doctoral program run jointly with the Barcelona Supercomputing Center (BSC) and completed partly at UC Berkeley (International Doctorate). His thesis addressed the design of power- and energy-efficient functional units for vector processors. He has published in venues such as IEEE Transactions on VLSI Systems, ACM Transactions on Architecture and Code Optimization, ISLPED, and the Journal of Big Data, serves as a reviewer for IEEE Transactions on VLSI Systems, and has served on a PhD defense committee at UPC-BarcelonaTech. His interests include computer architecture, vector processors, floating-point arithmetic, low-power and energy-efficient design, VLSI, RISC-V, and machine learning. Email: ivan.ratkovic.bg@gmail.com, ORCID ID: 0000-0002-0524-7227.