4.7 – Embedded Devices for Neuromorphic Time-Series Assessment

Arnab Neelim Mazumder, University of Maryland, Baltimore County
Morteza Hosseini, University of Maryland, Baltimore County
Tinoosh Mohsenin, University of Maryland, Baltimore County

Status

Neuromorphic computing aims to mimic the brain to create energy-efficient devices capable of handling complicated tasks. In this regard, analysis of multivariate time-series signals has led to advancements in different application areas ranging from speech recognition and human activity classification to electronic health evaluation. Exploration of this domain has led to unique bio-inspired commercial off-the-shelf device implementations in the form of fitness monitoring devices, sleep tracking gadgets, and EEG-based brain trauma marker identifying devices. Even with this deluge of work over the years, the necessity of evolving the research direction with day-to-day needs relating to this sphere is still pivotal. The key idea behind the wealth of research in these domains comes from the fact that it is very difficult to generalize human abilities and activities, and it is even more difficult to create devices that can operate at a level as accurate as human-level perception. This is where contemporary machine learning and the more modern deep learning frameworks shine. The current scenario of using automated devices for a variety of health-related applications requires that these devices become more sensitive, specific, user-friendly, and lastly accurate for their intended tasks. This relates to further advancements in the region of algorithm construction and constraint-based design of implementable hardware architectures. The current crop of research in this area investigates deep neural network (DNNs) architectures for the purpose of feature extraction, object detection, classification, etc. DNN models utilize the capacity of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and even to some extent fully connected layers to extract spatial features for time-series assessment which was previously exhaustively calculated via different hand-engineered feature extraction techniques coupled with simple classification algorithms. Along with this, RNNs and their advanced equivalents in the form of long short term memory networks (LSTMs) and gated recurrent unit (GRU) has also been integrated into the deep learning architectures to handle timeseries signals. The idea behind this integration stems from the fact that RNNs and LSTMs are modeled in such a way that they can keep track of previous instances of the input data in order to make a prediction, which makes these architectures very effective for pattern and dependency detection within the time-series data. The other aspect of developing these diverse DNN models is to make them readily implementable in terms of hardware accelerators and therein lies the issue of hardware constrained efficient designs. As a consequence, the computation and model size specifications of different hardware-oriented approaches will result in the advancement of application-oriented software designs which will, in turn, increase the reliability and efficiency of these embedded devices.

Current and Future Challenges

There are several challenges associated with managing time-series signals for classification or recognition tasks. One of the foremost issues of time-series classification is to make these signals interpretable by the DNNs as these signals contain multiple variables relaying information about concurrent actions and it is difficult to process these signals in their raw form. Authors in [1] proposed a solution to this problem by transforming these multimodal signals into windowed images based on their sampling frequencies. Another obstacle that is related to time-series analysis pertains to skewed or imbalanced information belonging to multimodal variables as the data collection procedure with different sensors might not always be the same. As a way around, a common practice is to use weighted sampling of the input features during the training of the DNN models so as to balance the

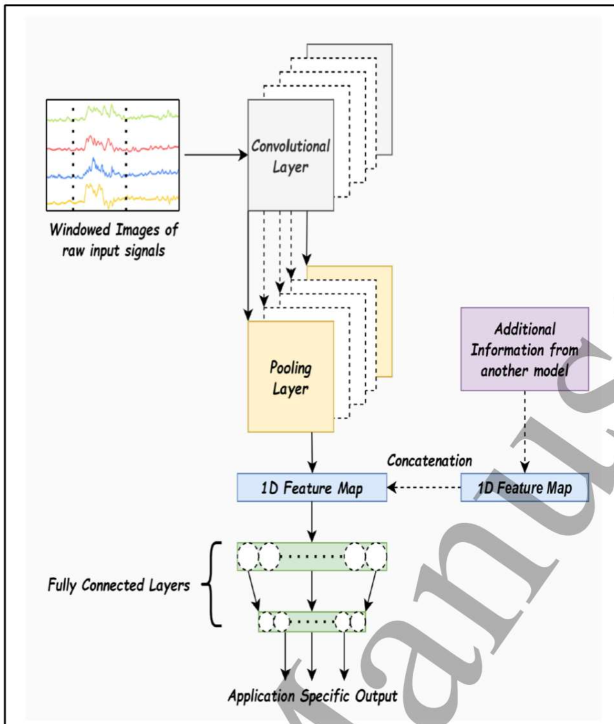


Figure 1. The deep learning framework takes in windowed images of the raw multimodal time-series signals as input to the convolutional layers. Correspondingly, feature extraction is achieved in convolutional layers which results in a two-dimensional feature map. The pooling layers contribute to reducing the feature map size while keeping the spatial features intact. This two-dimensional pooled feature map is reshaped to have one-dimensional form so that it can be forwarded to the next fully connected layers. Finally, the last fully connected layer will have neurons equal to the number of outputs as desired by the application. Furthermore, with regards to multi-input model, supplementary information coming from a separate model can be concatenated with the one-dimensional feature map to bolster the inference accuracy.

impact of all features. Pruning outliers in the dataset by eliminating unnecessary sensor data can alleviate this problem as demonstrated by the authors in [2], however, it is not always feasible to delete multimodal information as the sensor data for multiple variables might be correlated. In addition to this, many of the software frameworks dedicated to time-series classification do not consider the large computation overhead of the DNNs. This has a significant impact when these frameworks are replicated on to resource-limited and low-power embedded platforms where the use of off-the-chip-memories becomes essential. As a result, the performance remains limited by the memory bandwidth while the power consumption stays high due to the rapid accessing of off-thechip memories. The extent of these complications has introduced shallow networks [3], approaches to quantizing model parameters [4] along with ternary [5] and binary [6] models that focus on reducing the memory overhead for efficient resource-constrained hardware accelerator implementation. Authors in [7] provide an example of a fixed-point CNN classifier involving 4-bit fixed point arithmetic that suggests negligible accuracy degradation and authors in [8] present fast BNN inference accelerators to meet the FPGA on-chip memory requirements. Reducing memory footprints in hardware accelerators is also tied up to the cost-effective designing of memory units. On top of this, managing and limiting frequent accesses of these memory units also contribute to latency, power, and energy efficiency as a whole. Thus, a critical challenge in terms of hardware design is to maintain high frequency and energy efficiency with low energy consumption.

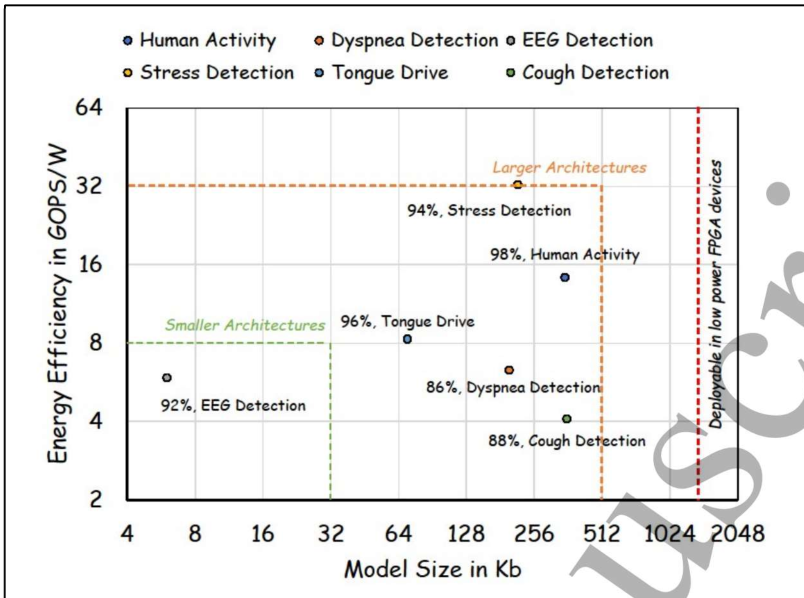


Figure 2. This figure illustrates the trend of energy efficiency against model size of different deep learning architectures deployed on the low power Artix–7 100t FPGA platform which has a memory of 1.65 Mb. The applications focused here are EEG detection [10], human activity recognition [1], stress detection [1], tongue drive systems [1] along with cough and dyspnea detection as part of respiratory symptoms recognition [9]. Depending on the model size, the frameworks can be tiny or large whereas the energy efficiency is dictated by the performance of the design. In the same vein, the plot also shows the device inference accuracy for the different models ranging from 86% up to 98% which further justifies that these architectures are specific enough for low power embedded deployment.

Advances in Science and Technology to Meet Challenges

Deep learning frameworks have been widely successful for classifying time-series signals. However, the challenges mentioned in the previous section make this task ever more difficult. To further boost the performance of deep learning methods for time-series data, some form of digital signal processing is commonly required. To this extent, a common practice is to convert these raw waveforms into windowed overlapping time-series frames. A sliding window of some specific size along with a stepping size is passed through all variables, creating a set of images of shape as desired by the user. Since most time-series signals contain label information at precise time intervals, it is fairly easy to determine the label of the images. Another facet of dealing with time-series signals requires feature extraction relevant to the application that is being targeted. With classical machine learning algorithms, this was achieved using several mathematical and analytical processes to determine the correlation between variables. In contrast, one of the strengths of using CNNs or RNNs in deep learning ensures that the relevant features are being extracted in image or time-space. A general practice in time-series classification is to deploy CNN or RNN layers in conjunction with pooling layers as illustrated in Fig.1. The pooling layers reduce the feature map size so that the cost of computation for the following fully connected layers is minimized. Additionally, the feasibility of hardware deployment of these deep learning algorithms depends on the computational complexity and size of these architectures. It is imperative that these frameworks are reduced in size via quantization, pruning, or by making the networks shallow in the first place so that they fit on embedded devices with small memories. Hence, there comes a point where the designer has to find the sweet spot between the accuracy of the model and the practicality of its size being suitable for low power embedded platforms while also ensuring that the energy efficiency of the target device is

also satisfactory. Fig. 2 shows a comparison among different models with a variety of applications for their model size, classification/detection accuracy, and energy efficiency which establishes that depending on the application, deep learning models can fit on low-power embedded devices with standard performance. Also, a modification to these frameworks can take in additional information in the form of vectors from a separate model to enhance the overall accuracy of the model as demonstrated in
Fig.1.

Concluding Remarks

Human-related time-series data analysis encompasses a wide range of tasks including speech recognition, keyword spotting, health monitoring, and human activity recognition to name a few. This also allows the dedicated development of embedded devices suited for accelerating such tasks. Challenges in processing such time-variant data for device implementation range from pre-processing the raw signals and removing noise and outliers to interpreting long and short dependencies that exist within the nature of the data. Windowing the continuous stream of data into overlapping frames to be processed using a simple DNN or CNN is a common practice for real-world applications in which the long dependencies in data are negligible. On the other hand, novel approaches such as RNNs and LSTMs can improve the overall confidence of analysis for time-series data with long dependencies. When implementing all these methods on resource-bound hardware in which power, energy, memory footprint, and application latency are all limited, it is of utmost importance to design deep learning algorithms with small model sizes and low computation that meet all the application requirements and hardware limitations. In conclusion, there must be a trade-off between performance and implementation feasibility to justify the use of low-power embedded devices to replicate deep learning applications of time-series assessment.

Acknowledgements

[Please include any acknowledgements and funding information as appropriate.]

References

[1] A. Jafari, A. Ganesan, C. S. K. Thalisetty, V. Sivasubramanian, T. Oates and T. Mohsenin, "SensorNet: A Scalable and Low-Power Deep Convolutional Neural Network for Multimodal Data Classification," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 66, no. 1, pp. 274-287, 2019.

[2] M. Hosseini, H. Ren, H.-A. Rashid, A. N. Mazumder, B. Prakash and T. Mohsenin, "Neural Networks for Pulmonary Disease Diagnosis using Auditory and Demographic Information," in *epiDAMIK 2020: 3rd epiDAMIK ACM SIGKDD International Workshop on Epidemiology meets Data Mining and Knowledge Discovery*, 2020.

[3] Y. N. Dauphin and Y. Bengio, "Big Neural Networks Waste Capacity. arXiv:1301.3583," 2013.

[4] Y. Gong, L. Liu , . M. Yang and . L. Bourdev, "Compressing Deep Convolutional Networks using Vector Quantization. arXiv:1412.6115," 2014.

[5] H. Alemdar, . V. Leroy, . A. B. Prost and F. Pétrot, "Ternary neural networks for resource-efficient AI applications," in *International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, 2017.

[6] M. Courbariaux, Y. Bengio and J. P. David, "Binaryconnect: Training deep neural networks with binary weights during propagations," *Advances in neural information processing systems,* vol. 28, pp. 3123-3131, 2015.

[7] C. Y. Lo, F. C. M. Lau and C. W. Sham, "Fixed-Point Implementation of Convolutional Neural Networks for Image Classification," in *International Conference on Advanced Technologies for Communications (ATC)*, Ho Chi Minh City, Vietnam, 2018.

[8] Y. Umuroglu, . N. J. Fraser, G. Gambardella, M. Blott, P. Leong, M. Jahre and K. Vissers, "Finn: A framework for fast, scalable binarized neural network inference," in *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, ACM, 2017, pp. 65-74.

[9] H. Ren , A. N. Mazumder , H.-A. Rashid, V. Chandrareddy, A. Shiri and T. Mohsenin, "End-to-end Scalable and Low Power Multi-modal {CNN} for Respiratory-related Symptoms Detection," in *IEEE 33rd International System-on-Chip Conference (SOCC)*, 2020.

[10] M. Khatwani, H.-A. Rashid, H. Paneliya, M. Horton, N. Waytowich, W. D. Hairston and T. Mohsenin, "A Flexible Multichannel EEG Artifact Identification Processor using DepthwiseSeparable Convolutional Neural Networks," *ACM Journal on Emerging Technologies in Computing Systems (JETC),* 2020.

4.8 Electromyography processing using Wearable Neuromorphic Technologies

Elisa Donati Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland

Status

Electromyography (EMG) is a neurophysiological technique for recording muscle movements. It is based on the principle that whenever a muscle contracts, a burst of electric activity is propagated through the close tissue. The source of the electrical signal in EMG is the summation of action potentials of motor units (MUs) [1]. A MU is composed of muscle fibers innervated by axonal branches of a motorneuron, that is intermingled with fibers of other MUs. The recorded electric activity is linearly correlated to the strength of the contraction and the number of recruited MUs. EMG signals can be acquired both invasively, using needle electrodes, and superficially, by placing electrodes on the skin - called surface EMG (sEMG).

EMG signals have been and are relevant in several clinical and biomedical applications. In particular, they are extensively employed in myoelectric prosthetics control for classifying muscle movements. Wearable solutions for this application already exist, but they have a large margin for improvement, from increasing the granularity of movement classification to reducing computational resources needed and consequently power consumption.

Like any other signal, EMG is susceptible to various types of noises and interferences, such as signal acquisition noise, and electrode displacement. Hence, a pre-processing phase is the first step to perform proper signal analysis, which involves filtering, amplification, compression, and feature extraction both in time and frequency domains [2]. The mainstream approach for movement classification is machine learning (ML), which delivers algorithms with very high accuracy [3], although the high variability in test conditions and their high computational load limit their deployment to controlled environments. These drawbacks can be partially solved by using deep learning techniques that allow for better generalization to unseen conditions but remain computationally expensive, requiring bulky power-hungry hardware, that hinder wearable solutions [4].

Neuromorphic technologies offer a solution to this problem by processing data with low latency and lowpower consumption mimicking the key computational principles of the brain [5]. Compared to state-of-the-art ML approaches, neuromorphic EMG processing shows a reduction of up to three orders of magnitude in terms of power consumption and latency [6–8], with limited loss in accuracy(5−7%) [9, 10].

New approaches have been proposed that directly extract the motorneurons activity from EMG signals as spike trains [12]. They represent a more natural and intuitive interface with muscles but currently limit themselves by processing spikes with traditional ML techniques and do not consider the possibility of using more appropriate frameworks such as spiking neural networks (SNNs).

Current and Future Challenges

Although the performance of myoelectric prosthetics increased conspicuously in the last decade [13], they still can not be used in daily life. The fine-grained control is in fact limited by the number of electrodes. This issue can be overcome by using High-Density EMG (HD-EMG), which typically uses hundreds of electrodes, allowing to monitor larger areas and effectively increasing the precision of the measurements [14]. However, HD-EMG uses more computational resources, in terms of power and time required to classify movements and to generate motor commands. Current technologies are not able to process such an amount of data in-situ and with low latency simultaneously. For this reason, the EMG signals are transmitted, for example via Bluetooth, to a remote system that is quite bulky and heavy, making a wearable solution impractical.

Neuromorphic technologies represent a solution to all the described limitations by processing data in parallel, with low latency, and taking advantage of the low-power nature of analog computing and spiking communication, as the biological system they are inspired from. Although recent results show promising advances, the current challenge of neuromorphic technology is to fill the gap with state-of-the-art ML approaches, in terms of accuracy. One of the main reasons behind this gap is the different amount of resources invested in the respective research fields. In addition, current research that focuses on adopting ML methods and implementing them in neuromorphic hardware faces challenges governed by the unsuitability of such substrates which are primarily targeted for SNNs [11].

To get the most from neuromorphic computing we need a change of paradigm, where the neuromorphic technology can directly interface with motorneurons' spiking activity, instead of continuous sEMG signals. This represents a matching condition between inputs and outputs that optimize the information transfer between the muscle activity and the processing and control unit. The spike trains of motorneurons can be extracted from sEMG signals by means of decomposition algorithms. In particular, the spatial distribution of MUs action potentials can be assessed with activation maps obtained from HD-EMG signals [12]. Nevertheless, current implementations are still computationally expensive, and only recently it was possible for their deployment in real-time. After the decomposition, the spike trains are translated and processed using ML methods instead of better-suited SNNs [15].

Designing neuromorphic systems able to extract and process motorneurons activity from EMG signals will pave the way to a new class of wearable devices that can be miniaturized and directly interface with the electrodes.

Advances in Science and Technology to Meet Challenges

A concrete roadmap towards wearable neuromorphic EMG processing, see Figure 1, could be constructed with short and long-term objectives. In the short term, we should advance neuromorphic computation to bridge the gap with ML methods for EMG classification, and optimize decomposition algorithms to make them run real-time on embedded systems. In the long-term, the decomposition algorithm should be ported into a neuromorphic chip to implement a fully spiking pipeline while the technological breakthroughs in surface smart electrodes could potentially be able to record directly motorneurons' spike trains.

Bridge-the-Gap. The first step is to understand the requirements to improve the accuracy of EMG movements classification. The front-end, which includes pre-processing and spike conversion, has the largest margin for improvement. Signal-to-spike conversion produces spike trains required by neuromorphic devices. The most common signal-to-spike converter is the delta-sigma [7] which is widely applied in biomedical applications, thanks to its lower circuit complexity compared to multi-bit ADCs. However, the delta-modulator generates a high sampling rate and larger data size that can easily push the neurons' firing rate into saturation, making them insensitive to further input variations. Furthermore, SNNs for EMG classification should be optimized and learning algorithms could make them adaptable to different patients.

Embedded Decomposition sEMG decomposition into spike trains is generally based on shape-based algorithms, also called template matching [16] or blind source separation algorithms [17]. The decomposition of the complex sEMG is a computationally expensive procedure in a multidimensional constraint space. To run these algorithms on embedded platforms and in real-time it is imperative to i. reduce the complexity and ii. optimize it for the selected digital embedded architecture (e.g. PULP platform [18]) and exploit its hardware capabilities. The extracted spike trains are then sent to a neuromorphic chip, creating a hybrid digital-analog framework for spike encoding low-power computation.

Spike-based EMG decomposition To build a fully spiking pipeline that can be integrated into a single neuromorphic chip, the MUs identification algorithm needs to be translated into a spiking version. Embedding the entire process into a single chip that can be miniaturized and connected directly to the electrodes will allow online processing, which is optimal for real-time closed-loop applications and less vulnerable to interferences either caused by humans or the environment.

Smart electrodes Another long-term game-changer would be the technological breakthroughs that will allow the single electrode to be able to record directly the activity of a single MU, removing the need for decomposition algorithms.

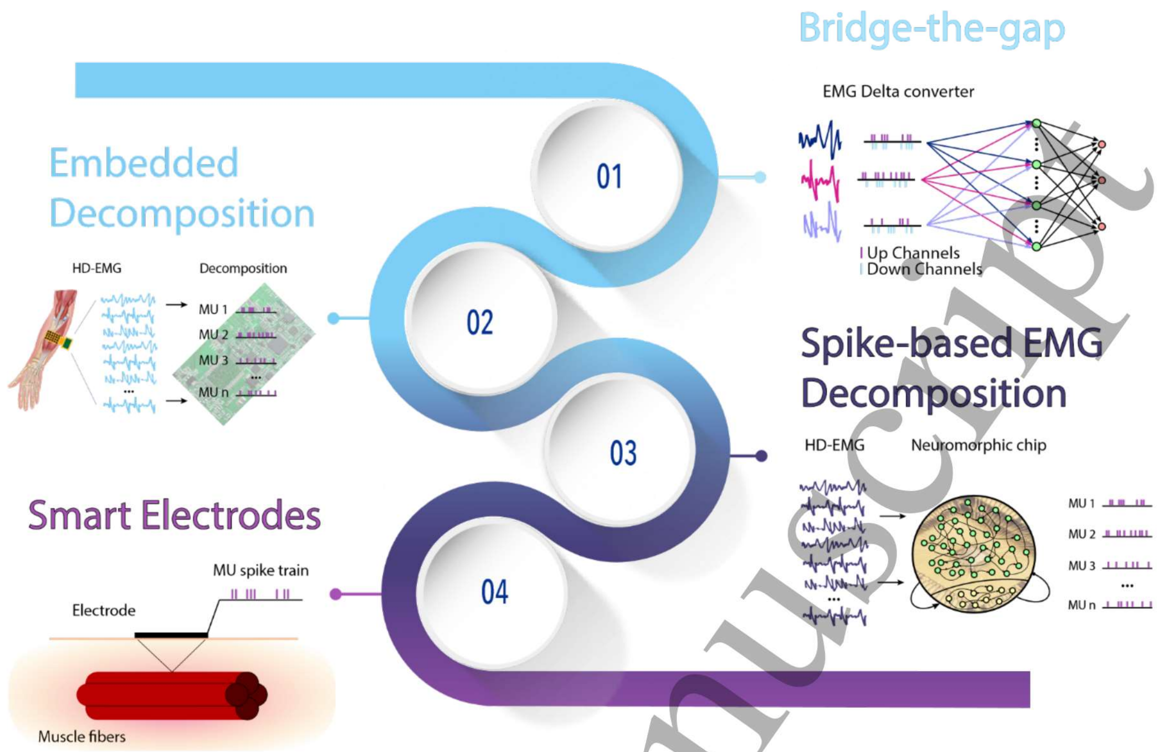


Figure 1: A concrete Roadmap towards neuromorphic Wearable Devices

Concluding Remarks

The need of improving myoelectric prosthetic control to increase the life quality of the patient poses new challenges for implementing real-time, compact, and low-power EMG processing systems. A wearable device based on neuromorphic technology can enable in-situ EMG signal processing and decomposition, without information transfer and external computation. In particular, mixed-signal SNNs implemented on neuromorphic processors can be integrated directly with the sensors to extract temporal data streams in real-time with lowpower consumption.

This roadmap presents the specific case of prosthetic control, nevertheless, the development of this technology could reveal useful to more applications where continuous monitoring is required. In clinical settings, continuous monitoring of EMG signals can be utilized to detect degenerative diseases of motorneurons [19] even for very large time spans such as weeks or months. In rehabilitation, EMG can be used as feedback to adapt the patient training accordingly to its muscular status, after a stroke or neurological impairments [20].

With the current rate of technological and computational improvements the proposed objectives could be realistically achieved within a decade. If successfully executed, this roadmap will bring technology that will improve the quality of life for amputees and patients with motorneuron diseases.

References

[1] Enoka, R. M. (1995). Morphological features and activation patterns of motor units. Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society, 12(6), 538-559.

[2] Rodriguez-Tapia, B., Soto, I., Martinez, D. M., & Arballo, N. C. (2020). Myoelectric interfaces and related applications: Current state of EMG signal processing–A systematic review. IEEE Access, 8, 7792-7805.

[3] Arteaga, M. V., Castiblanco, J. C., Mondragon, I. F., Colorado, J. D., & Alvarado-Rojas, C. (2020). EMGdriven hand model based on the classification of individual finger movements. Biomedical Signal Processing and Control, 58, 101834.

[4] Park, K. H., & Lee, S. W. (2016). Movement intention decoding based on deep learning for multiuser myoelectric interfaces. In 2016 4th international winter conference on brain-computer Interface (BCI) (pp. 1-2). IEEE.

[5] Chicca, E., Stefanini, F., Bartolozzi, C., & Indiveri, G. (2014). Neuromorphic electronic circuits for building autonomous cognitive systems. Proceedings of the IEEE, 102(9), 1367-1388.

[6] Donati, E., Payvand, M., Risi, N., Krause, R., Burelo, K., Indiveri, G., Dalgaty, T. & Vianello, E., (2018). Processing EMG signals using reservoir computing on an event-based neuromorphic system. In 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS) (pp. 1-4). IEEE.

[7] Donati, E., Payvand, M., Risi, N., Krause, R., & Indiveri, G. (2019). Discrimination of EMG signals using a neuromorphic implementation of a spiking neural network. IEEE transactions on biomedical circuits and systems, 13(5), 795-803.

[8] Ceolini, E., Frenkel, C., Shrestha, S. B., Taverni, G., Khacef, L., Payvand, M., & Donati, E. (2020). Handgesture recognition based on EMG and event-based camera sensor fusion: A benchmark in neuromorphic computing. Frontiers in Neuroscience, 14.

[9] Behrenbeck, J., Tayeb, Z., Bhiri, C., Richter, C., Rhodes, O., Kasabov, N., Espinosa-Ramos, J.I., Furber, S., Cheng, G. and Conradt, J., 2019. Classification and regression of spatio-temporal signals using NeuCube and its realization on SpiNNaker neuromorphic hardware. Journal of neural engineering, 16(2), p.026014.

[10] Ma, Y., Chen, B., Ren, P., Zheng, N., Indiveri, G., & Donati, E. (2020). EMG-Based Gestures Classification Using a Mixed-Signal Neuromorphic Processing System. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 10(4), 578-587.

[11] Azghadi, M. R., Lammie, C., Eshraghian, J. K., Payvand, M., Donati, E., Linares-Barranco, B., & Indiveri, G. (2020). Hardware implementation of deep network accelerators towards healthcare and biomedical applications. IEEE Transactions on Biomedical Circuits and Systems, 14(6), 1138-1159.

[12] Del Vecchio, A., Germer, C.M., Elias, L.A., Fu, Q., Fine, J., Santello, M. and Farina, D., (2019). The human central nervous system transmits common synaptic inputs to distinct motor neuron pools during non-synergistic digit actions. The Journal of physiology, 597(24), pp.5935-5948.

[13] Valle, G., Mazzoni, A., Iberite, F., D'Anna, E., Strauss, I., Granata, G., Controzzi, M., Clemente, F., Rognini, G., Cipriani, C. and Stieglitz, T., 2018. Biomimetic intraneural sensory feedback enhances sensation naturalness, tactile sensitivity, and manual dexterity in a bidirectional prosthesis. Neuron, 100(1), pp.37-45.

[14] Moin, A., Zhou, A., Rahimi, A., Menon, A., Benatti, S., Alexandrov, G., Tamakloe, S., Ting, J., Yamamoto, N., Khan, Y., & Burghardt, F., (2020). A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. Nature Electronics, pp.1-10.

[15] Urh, F., Strnad, D., Clarke, A., Farina, D., & Holobar, A. (2020, July). On the Selection of Neural Network Architecture for Supervised Motor Unit Identification from High-Density Surface EMG. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 736-739). IEEE.

[16] Kleine, B. U., van Dijk, J. P., Lapatki, B. G., Zwarts, M. J., & Stegeman, D. F. (2007). Using twodimensional spatial information in decomposition of surface EMG signals. Journal of electromyography and kinesiology, 17(5), 535-548.

[17] Holobar, A., & Zazula, D. (2007). Multichannel blind source separation using convolution kernel compensation. IEEE Transactions on Signal Processing, 55(9), 4487-4496.

[18] Rossi, D., Conti, F., Marongiu, A., Pullini, A., Loi, I., Gautschi, M., Tagliavini, G., Capotondi, A., Flatresse, P. & Benini, L., (2015). PULP: A parallel ultra low power platform for next generation IoT applications. In 2015 IEEE Hot Chips 27 Symposium (HCS) (pp. 1-39). IEEE. Vancouver

[19] Chatterjee, S., Roy, S. S., Bose, R., & Pratiher, S. (2020). Feature extraction from multifractal spectrum of electromyograms for diagnosis of neuromuscular disorders. IET Science, Measurement & Technology, 14(7), 817-824.

[20] Trigili, E., Grazi, L., Crea, S., Accogli, A., Carpaneto, J., Micera, S., Vitiello, N., & Panarese, A., (2019). Detection of movement onset using EMG signals for upper-limb exoskeletons in reaching tasks. Journal of neuroengineering and rehabilitation, 16(1), pp.1-16.

4.9 – Collaborative Autonomous Systems

Silvia Tolu, Roberto Galeazzi, Technical University of Denmark (DTU)

Status

Collaborative autonomous systems (CAS) (see Figure 1) are entities that can cooperate among themselves and with humans, with variable level of human intervention (depending on the level of autonomy) in performing complex tasks in unknown environments. Their behaviour is driven by the availability of perception, communication, cognitive and motor skills and improved computational capabilities (on/off-board systems). The high level of autonomy enables the execution of dependable actions under changing internal or external conditions. Therefore, CAS are expected to be able to: 1. perceive and understand their own condition and the environment they operate in; 2. dependably interact with the physical world despite of sudden changes; 3. intelligently evolve through learning and adaptation to unforeseen operational conditions; 4. self-decide their actions based on their understanding of the environment.

Currently, CAS (e.g., collaborative robots - cobots) show limited performances when accomplishing physical interaction tasks in complex scenarios [1]. Recent studies have demonstrated that autonomous robots can outperform the task they are programmed for, but they are limited in the ability to adapt to unexpected situations [2] and to different levels of human-robot cooperation [1]. These limitations are mainly due to the lack of generalization capabilities, i.e., cobots cannot transfer knowledge across multiple situations (environments, tasks, and interactions). One of the most viable pathways to solve this issue is to build intelligent autonomous cobots by incorporating Artificial Intelligence (AI)-based methods into the control systems [3]. These bio-inspired controllers [4] allow taking a different perspective from the classical control approaches, which require a deeper understanding of the mechanics of the interactions and of the intrinsic limitations of the systems beforehand. Main current research directions [5] are focused on the understanding of the biological working principles of the central nervous system (CNS) in order to build innovative neuromorphic computing algorithms and hardware that will bring significant advances in this field; In particular, they will provide computational efficiency and powerful control strategies for robust and adaptive behaviours.

In the next decades, there will be significant developments in CAS related to self-capabilities such as self-inspection, -configuration, -adaptation, -healing, -optimization, -protection, and -assembly. This will be a great enabler of systems acting in real-world unstructured scenarios, such as in remote applications (deep sea or space), in hazard situations (disasters), in healthcare interventions (assistive, rehabilitation, or diagnosis), and in proximity to people.

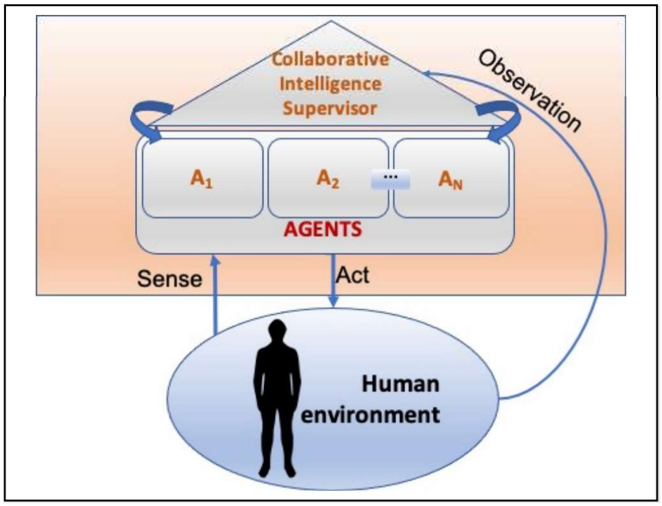


Figure 1. Overall idea of a Collaborative Autonomous Control System. The supervisor manages the entire system, observes and analyses the whole situation and provides information to each agent to improve their autonomous actions and optimize the operations.

Current and Future Challenges

Several fundamental challenges demand to be addressed to enable the deployment of heterogenous autonomous systems able to collaborate towards the achievement of common mission objectives.

These challenges span across different research topics including **online mission planning** and execution for multi-agent systems under uncertainty. Future mission planners [6] should integrate several factors to determine the optimal allocation of agents to the fulfilment of the mission tasks. These factors, among many, include energy availability and depletion rates, physical capabilities of the agents, probability of failures, and amount of collaboration needed. The mission execution demands the development of a revolutionary control paradigm that enables true collaboration among CAS with different functionalities. **Cooperative control** [7] has been so far limited to consensus and synchronization to enable the coordinated dynamic evolution of mostly homogeneous multi-agent systems to perform the same type of actions. The execution of tasks in uncertain environments calls for robust learning/adaptation methods to enable baseline control systems to ensure robust cooperation and coordination of heterogeneous multi-agent robots in various real applications [7]. Another big question is how to endow CAS with a high-level of **fault tolerance** capabilities in order to ensure dependability under a wide variety of operational conditions [8]. Despite the large research effort pursued by the community over the past four decades, condition monitoring and fault tolerant control are lacking efficiency due to the ever-increasing complexity of the systems.

Future work will aim to provide insights about how a CAS will show **robust, compliant, and intelligent physical interactions** with the environment, human beings, or other systems. In this regard, real-time, energy-efficient computing is required to advance the type of primitive collaborations that are achievable so far. With this aim, systems should be equipped with small processors able to ensure low energy consumption, and, at the same time, increase the memory bandwidth. Current alternatives (e.g., multicore central processing units - CPUs [9], new graphical processors - GPUs [10], parallel processing core - SpiNNaker [11]) still suffer from an extremely high energy demand that is not sustainable and they cannot be easily scaled. Additionally, a limited number of processes can run simultaneously, and the speed of the response is still low. Consequently, new neuromorphic architectures are the most promising alternatives to address the increasing demand to create CAS able of a seamless interaction with human beings.

Advances in Science and Technology to Meet Challenges

In this section, we discuss the foremost advances in science and technology that will address the main aforementioned challenges.

Mission Planning - Novel AI-based heuristic methods will be developed to equip mission planners with key functionalities that will increase the value for the human operators. These include: the close-loop decomposition of missions to achieve an adaptive task allocation by leveraging information gathered at mission execution; automated survivability prediction to assess the likelihood of vehicle loss based on faults and failures occurred in past missions; automated reliability assessment to forecast the probability of mission failure based on past missions' information; automated learning from previous missions' performance to tune the future missions' parameters; inclusion of services to extend the mission endurance [12, 13].

Fault-tolerant and Cooperative Control - Paradigms based on cooperation will be created to fulfil the advances in multi-agent systems [8]. Cooperation among agents offers the possibility of achieving fault-tolerance towards sensors and actuators faults through the design of diagnostic solutions that leverage shared proprioceptive and exteroceptive information. Prescribe-time fault tolerant cooperative control solutions for safety critical cyber-physical systems will be achieved; these will provide the basis for efficient fault-tolerant algorithms able to trade-off between fast convergence and acceptable fault-tolerance performance.

Robust, Compliant and Intelligent Physical Interactions - New physical mechanisms will be designed to provide passive properties to the system, to increase the physical interaction performances, and include advanced control aspects for achieving simultaneous robustness and compliance. The advances in Neuro-robotics and Neuromorphic Computing will influence the development of the next generation of intelligent agents [14]. Current neuromorphic computing systems already exploit learning and adaptive skills in systems compared to conventional von-Neumann machines thanks to non-volatile memories and power efficiency performance [15]. However, new types of sensors and actuators will be introduced to enhance the cognitive and learning functionalities of the systems and deal with safety and robustness concerns. Advanced bio-inspired platforms, e.g., brain-on-the-chip devices, will be designed for processing complex brain-inspired computing techniques that will support autonomy, more connectivity, increased decentralization, and high-performance computing. Indeed, neuromorphic technologies will be able to process complex unstructured data and learn to self-respond to external unknown stimuli enabling their use in critical edge applications, for example in autonomous navigation, human-machine interactions and smart healthcare markets.

Finally, innovative applications could be generated through the development of self-reconfigurable modular CAS, systems able to adapt their morphology and functionality to varied environments including unforeseen conditions [16]. This will require self-learning capabilities to develop new knowledge and to decide upon the previous accumulated experience.

Concluding Remarks

This paper has presented the future perspectives of collaborative autonomous systems and the main challenges and research issues that need to be addressed toward their realization. Further to these scientific and technological challenges, there are ethical, social, and legal issues when realising CAS, though these are beyond the scope of this article.

CAS working alongside humans have already been deployed and they support humans' work ensuring high productivity, speed, and accuracy [17]; they also relieve us of many heavy and time-consuming tasks and reduce the overall risk of collisions. CAS provide an economically viable entry-point to automation of processes, i.e., accelerated testing scenarios on products, environmental impacts.

Fusion of fundamental and applied research in both technical and natural sciences will facilitate the development of new theoretical frameworks for the design of intelligent CAS. Multiple disciplines will be merged to pursue a systematic innovation within cyber-physical systems with variable level of autonomy and cooperation; the use of AI and Internet of Everything technologies future proofs the system to address changing market demands and expectations in several technological areas. Applications will be many and varied including, and not limited to, manufacturing, health care, inspection and maintenance, precision farming, autonomous marine operations, and education.

References

[1] Ajoudani, A., Zanchettin, A.M., Ivaldi, S. et al., “Progress and prospects of the human–robot collaboration”, in Auton. Robot. 42, 2018, pp. 957–975. https://doi.org/10.1007/s10514-017-9677-2

[2] Ingrand F., Ghallab M. “Deliberation for autonomous robots: A survey”, in Artificial Intelligence, 247, 2017, pp. 10-44. https://doi.org/10.1016/j.artint.2014.11.003

[3] Kunze L., Hawes N., Duckett T., Hanheide M., and Krajník T., "Artificial Intelligence for Long-Term Robot Autonomy: A Survey", in IEEE Robotics and Automation Letters, 3(4), 2018, pp. 4023-4030. Doi: 10.1109/LRA.2018.2860628.

[4] Capolei M.C., Angelidis E., Falotico E., Hautop Lund H, and Tolu S., “A Biomimetic Control Method Increases the Adaptability of a Humanoid Robot Acting in a Dynamic Environment”, in Frontiers in Neurorobotics, 13, 2019, p. 70. Https://doi.org/10.3389/fnbot.2019.00070

[5] Indiveri G., Sandamirskaya Y., “The Importance of Space and Time for Signal Processing in Neuromorphic Agents: The Challenge of Developing Low-Power, Autonomous Agents That Interact with the Environment”, in IEEE Signal Processing Magazine, 36(6), 2019, pp. 16-28. Doi: 10.1109/MSP.2019.2928376

[6] Thompson F., Galeazzi R., “Robust mission planning for Autonomous Marine Vehicle fleets”, in Robotics and Autonomous Systems, 124, 2020. Doi: 10.1016/j.robot.2019.103404

[7] Zool H. I., Nohaidda S., “A Survey and Analysis of Cooperative Multi-Agent Robot Systems: Challenges and Directions”, in Applications of Mobile Robots, IntechOpen, 1, 2019, Doi: 10.5772/intechopen.79337

[8] Yang H., Han Q.-L., Ge X., Ding L., Xu Y., Jiang B., Zhou D., "Fault-Tolerant Cooperative Control of Multiagent Systems: A Survey of Trends and Methodologies," in IEEE Transactions on Industrial Informatics, 16(1), 2020, pp. 4-17. Doi: 10.1109/TII.2019.2945004.

[9] Fardet T., and Levina A., “Simple models including energy and spike constraints reproduce complex activity patterns and metabolic disruptions”, in PLOS Computational Biology 16(12), 2020. Doi: https://doi.org/10.1371/journal.pcbi.1008503

[10] Naveros F., Garrido J. A., Carrillo R. R., Ros E., and Luque N.R., “Event- and Time-Driven Techniques Using Parallel CPU-GPU Co-processing for Spiking Neural Networks”, in Frontiers Neuroinformatics, 11, 2017, pp. 7. Doi: https://doi.org/10.3389/fninf.2017.00007

[11] Furber S., “Large-scale neuromorphic computing systems”, in J. Neural Eng. 13(5), 2016. Doi: http://dx.doi.org/10.1088/1741-2560/13/5/051001

[12] Thompson F., and Guihen D., “Review of mission planning for autonomous marine vehicle fleets”, in Journal of Field Robotics, *36*(2), 2019, pp.333-354. Doi: 10.1002/rob.21819

[13] Atyabi A., MahmoudZadeh S., and Nefti-Meziani S., “Current advancements on autonomous mission planning and management systems: An AUV and UAV perspective”, in Annual Reviews in Control, 46, 2018, pp.196-215. Doi:10.1016/j.arcontrol.2018.07.002

[14] Li J., Li Z., Chen F., Bicchi A., Sun Y., Fukuda T., "Combined Sensing, Cognition, Learning, and Control for Developing Future Neuro-Robotics Systems: A Survey," in IEEE Transactions on Cognitive and Developmental Systems, 11(2), 2019, pp. 148-161. Doi: 10.1109/TCDS.2019.2897618.

[15] Woo J., Hun Kim J., Im J-P, Eon Moon S., “Recent Advancements in Emerging Neuromorphic Device Technologies”, 2(10), 2020. Doi: https://doi.org/10.1002/aisy.202000111

[16] Hauser S., Mutlu M., Léziart P.-A., Khodr H., Bernardino A., Ijspeert A.J., “Roombots extended: Challenges in the next generation of self-reconfigurable modular robots and their application in adaptive and assistive furniture,” in Robotics and Autonomous Systems, 127, 2020. Doi:10.1016/j.robot.2020.103467.

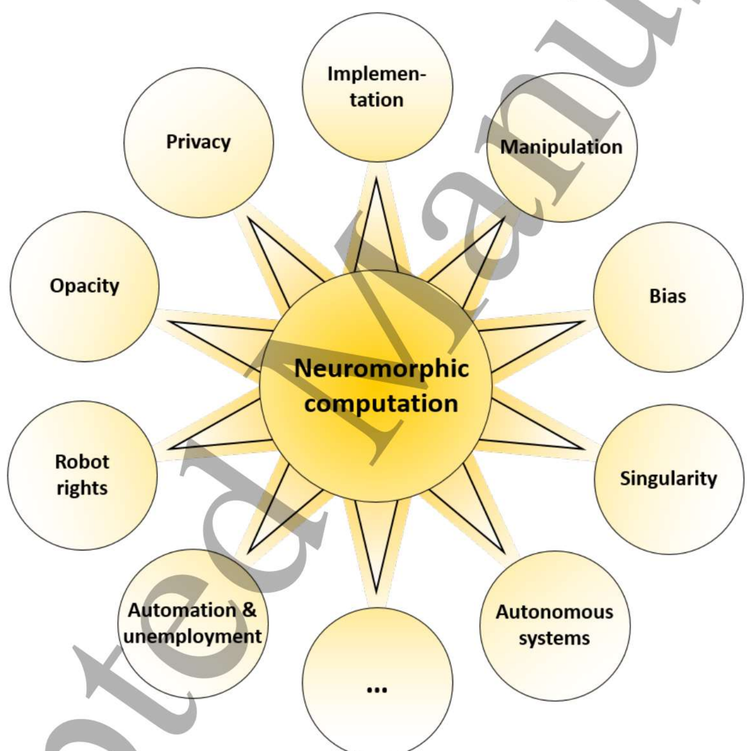
[17] Galin R., Meshcheryakov R., Kamesheva S., Samoshina A., “Cobots and the benefits of their implementation in intelligent manufacturing”, in IIOP Conference Series: Materials Science and Engineering, 862, 2020. Doi:10.1088/1757-899X/862/3/032075

5. 1 The ethics of developing neuromorphic technology

Martin Ejsing Christensen
Research Officer
The Danish Council on Ethics
mech@dketik.dk

Sune Holm
Associate Professor
Department of Food and Resource Economics
University of Copenhagen
suneh@ifro.ku.dk

Like the development of other forms of artificial intelligence, the development of neuromorphic technology may raise a number of ethical questions. [1], [2].



Figur 1 Some of the most salient ethical issues raised by the development of neuromorphic technology

One issue concerns privacy and surveillance. The development of most forms of artificial intelligence depends upon access to data, and as far as these data can be seen as private or personally identifiable, it raises a question about when it is (ethically) defensible to use such data. On the one hand, some argue that persons have a right to be let alone and exercise full control over information about themselves, so that any use of such data presupposes fully informed consent. On the other hand, others recognize the importance of privacy but argue that it may sometimes be outweighed by the fact that reliable applications for the good of everyone presuppose access to high quality representative data [3].

Another issue concerns opacity. Many forms of artificial intelligence support decision making based on complex patterns extracted from huge data sets. Often, however, it will be impossible not only for the person who makes the final decision but also for the developer to know what the system's recommendations are based on and it is in this sense that it is said to be opaque. For some such opacity does not matter as long as there are independent ways of verifying that the system delivers an accurate result, but others argue that it is important that the system is explainable [4]. In this way, a tension is often created between accuracy and transparency, and what the right trade-off is may often depend upon the concrete context.

Opacity is closely connected with the question of bias since opacity may hide certain biases. There are different forms of bias but in general, bias arises when automated AI decision support systems are based on data that is not representative of all the individuals that the system supports decisions in relation to [5]. There are different opinions as to when the existence of bias in automated decision support systems poses a serious problem. Some argue that 'traditional' unsupported human decision-making is biased, too, and that the existence of bias in automated AI decision support systems only pose a serious problem if the bias is more significant than the pre-existing human bias. Others argue that features such as opacity or the lack of suitable institutional checks and balances may tend to make the existence of bias in automated decision support systems more problematic than 'ordinary' human bias [6]. A separate problem is created by the fact that it sometimes will be easier to identify and quantify bias in AI systems than in humans, making a direct comparison more difficult.

The development of forms of artificial intelligence based on neuromorphic technology also raises questions about manipulation of human behavior, online as well as offline. One context in which such questions arise is advertising and political campaigning, where AI generated deep knowledge about individuals' preferences and beliefs, which may be used to influence them in a way that escapes the individuals' own awareness. Similar issues may also arise in connection with other forms of artificial intelligence such as chatbots and care or sex robots that simulate certain forms of human behavior without being 'the real deal'. Even if persons develop some form of emotional attachment to such systems, some argue that there is something deeply problematic and deceptive about such systems [7], while others point out that there is nothing intrinsically wrong with such systems as long as they help satisfy human desires [8]. If, as described in section 4.1, neuromorphic technologies will make it possible for robots to move from extremely controlled environments to spaces where they collaborate with humans and exhibit continuous learning and adaptation, it may make such questions more pressing.

A distinct set of issues are raised by the possibility of developing AI systems that do not just support human decision making but operate in a more or less autonomous way such as 'self-driving' cars and autonomous weapons. One question that such systems raise concerns the way in which they should be programmed in order to make sure that they make ethically justifiable decisions (in most foreseeable situations). Another question concerns how responsibility and risk should be distributed in the complex social system they are a part of. If, as described in section 4.2, neuromorphic engineering offers the kind of technological leaps required for achieving truly autonomous vehicles, the development of neuromorphic technologies may make such questions more pressing than at present.

A distinct issue relates to sustainability. As pointed out in the introduction, 5-15% of the world's energy is spent in some form of data manipulation (transmission or processing), and as long as a substantial amount of that energy comes from sources that contribute to climate change through the emission of greenhouse gases, it raises a question as to whether all that data manipulation is really necessary or could be done in a more energy efficient way. And in so far as neuromorphic technologies, as e.g. pointed out in section 4.7., shows a reduction of up to three orders of magnitude in terms of power consumption compared to state-of-the-art ML approaches, it seems to provide robust ethical support for the development of neuromorphic technologies.

As mentioned in the beginning of this section, the ethical questions raised by the development of neuromorphic technology is not unique to this technology but related to the development of artificial intelligence as such. The successful development of neuromorphic technology may make some of the issues more pressing, and a central task for future work on the ethics of neuromorphic technology will, accordingly, be to inquire into the exact way in which the issues are raised by the development of neuromorphic technology. But the existing forms of artificial intelligence already raise many of the questions described so far. Besides these questions, however, the development of neuromorphic technology (as well as other forms of artificial intelligence) may also raise a number of questions that are more speculative either because it is unclear whether the development will take place, when it will happen or what the precise consequences will be.

One such issue has to do with automation and unemployment. Artificial intelligence systems have already replaced humans in certain job functions (e.g., customer service), but it has been suggested that most job functions will be affected by the development of artificial intelligence at one point [9]. Because such a development has the potential to disrupt the social order (e.g., through mass unemployment) it raises an important ethical (and political) question as to how artificial intelligence systems should be introduced into society [10].

Another more speculative issue relates to artificial moral agents and so-called robot rights. If the development of neuromorphic (and other) forms of artificial intelligence leads to the creation of systems that possess some or all the traits that make us ascribe rights and responsibilities to humans, it may thus raise a question about whether such rights and responsibilities should be ascribed to artificially intelligent systems [11], [12].

Thirdly, some have also pointed out that the development of neuromorphic (and other) forms of artificial intelligence may create issues related to the so-called singularity. The idea is that the technological development may lead to the creation of general forms of artificial intelligence that surpass the human level of intelligence and then begin to control the further development of artificial intelligence in ways that may not be in the interests of the human species and perhaps even threaten its very existence. Whether such a scenario is likely has been questioned [13], but some argue that even a slight risk should be taken serious given the potentially devastating consequences [14].

No matter what one thinks is the right answer to the ethical questions raised by the development of neuromorphic technology, it is, finally, worth noticing that it still leaves an important practical question: how best to make sure that the actual development and implementation of neuromorphic technology will take place in an ethically defensible way. For some questions, governmental regulation may be the best means. For others, the best solution may be to trust the community of developers to make the right, value-based decisions when designing systems, while some questions, perhaps, should be left to the enlightened citizenry. In the end, however, it will probably be up to an inquiry into the concrete situation to decide when one or the other approach – or combination of approaches – provides the best means of securing an ethically defensible development of neuromorphic technology.

References:

[1] Müller, Vincent C., "Ethics of Artificial Intelligence and Robotics", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/ethics-ai/>.

[2] Coeckelbergh, Mark*, AI Ethics,* Cambridge: MIT Press, 2020.

[3] Topol, Eric, Deep Medicine: How Artificial Intelligence Can Make Healthcare Human again, New York: Basic Books, 2019.

[4] Cohen, I. (2018). Is There a Duty to Share Healthcare Data? In I. Cohen, H. Lynch, E. Vayena, & U. Gasser (Eds.), Big Data, Health Law, and Bioethics (pp. 209-222). Cambridge: Cambridge University Press. doi:10.1017/9781108147972.020

[5] Binns, Reuben, "Fairness in Machine Learning: Lessons from Political Philosophy", *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, in *Proceedings of Machine Learning Research*, 81: 149–159, 2018.

[6] Ploug, Thomas and Søren Holm, "The right to refuse diagnostics and treatment planning by artificial intelligence*", Medicine, Health Care and Philosophy*, 2020 Mar;23(1):107-114.

[7] Nyholm, Sven, and Lily Frank, "From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible?", 219-243, in *Robot Sex: Social and Ethical Implications*, Cambridge: MIT Press, 2017.

[8] Danaher, John, "The Philosophical Case for Robot Friendship", *Journal of Posthuman Studies*, 3(1): 5–24, 2019.

[9] Baldwin, Richard, *The Globotics Upheaval: Globalisation, Robotics and the Future of Work*, New York: Oxford University Press, 2019.

[10] Goos, Maarten, "The Impact of Technological Progress on Labour Markets: Policy Challenges", *Oxford Review of Economic Policy*, 34(3): 362–375, 2018.

[11] Turner, Jacob, *Robot Rules: Regulating Artificial Intelligence*, Berlin: Springer, 2019.

[12] Coeckelbergh, Mark, "Care Robots and the Future of ICT-Mediated Elderly Care: A Response to Doom Scenarios", *AI & Society*, 31(4): 455–462, 2016.

[13] Bostrom, Nick: *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press, 2014.

[14] Floridi, Luciano: "Should We Be Afraid of AI? Machines Seem to Be Getting Smarter and Smarter and Much Better at Human Jobs, yet True AI Is Utterly Implausible. Why?", *Aeon*, 9 May 2016. https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible