Frontiers in Neurorobotics

TYPE Review
PUBLISHED 03 February 2023
DOI 10.3389/fnbot.2023.1112839

When neuro-robots go wrong: A review

Muhammad Salar Khan and James L. Olds*

Interdisciplinary Neuroscience Program, Center for Biomedical Science and Policy, Schar School of Policy and Government, George Mason University, Arlington, VA, United States

OPEN ACCESS

EDITED BY
Florian Röhrbein,
Technische Universität Chemnitz, Germany

REVIEWED BY
Jiahao Chen,
Chinese Academy of Sciences (CAS), China
Tiffany Hwu,
University of California, Irvine, United States
Kenneth Stewart,
University of California, Irvine, United States

*CORRESPONDENCE
James L. Olds
✉ jolds@gmu.edu

RECEIVED 30 November 2022
ACCEPTED 19 January 2023
PUBLISHED 03 February 2023

CITATION
Khan MS and Olds JL (2023) When neuro-robots go wrong: A review.
Front. Neurorobot. 17:1112839.
doi: 10.3389/fnbot.2023.1112839

Neuro-robots are a class of autonomous machines that, in their architecture, mimic aspects of the human brain and cognition. As such, they represent unique artifacts created by humans based on human understanding of healthy human brains. European Union's Convention on Roboethics 2025 states that the design of all robots (including neuro-robots) must include provisions for the complete traceability of the robots' actions, analogous to an aircraft's flight data recorder. At the same time, one can anticipate rising instances of neuro-robotic failure, as they operate on imperfect data in real environments, and the underlying AI behind such neuro-robots has yet to achieve explainability. This paper reviews the trajectory of the technology used in neuro-robots and accompanying failures. The failures demand an explanation. While drawing on existing explainable AI research, we argue explainability in AI limits the same in neuro-robots. In order to make robots more explainable, we suggest potential pathways for future research.

KEYWORDS

neuro-robotic systems, explainability, explainable AI (X-AI), neuro-robotic failures, explainable neuro-robots, neuro-robotic models, responsible neuro-robots

1. Introduction

Japan's Henna Hotel (literally translates in English to "strange hotel")—opened in 2015 with a staff of 243 robots—has cut its robotic workforce, stating that the robots started annoying the guests frequently (Hertzfeld, 2019). The firing of the robots comes after many objections from both staff and customers. Many of these robots created more work for the hotel staff instead of reducing it. There are numerous other instances of robot failures. In the famous DARPA Robotics Challenge, many robots fell over (including the IHMC's Atlas robot shown in **Figure 1**), and some fell over multiple times (Guizzoevan and Ackerman, 2015).

Beyond traditional robotic failures, early-stage neuro-robots employing embodied intelligence—notably humanoid (developmental) robots inspired by the human nervous system and most visually resembling humans—also failed in several instances, particularly in unpredictable situations. For example, Pepper, a humanoid robot from the Japanese firm SoftBank, could not deliver at jobs it was designed for, from entertaining residents at nursing homes to welcoming customers in banks (Ryan, 2021). Similarly, the Atlas humanoid robot developed by Boston Dynamics failed while doing parkour (Pescovitz, 2021). Many other humanoids that closely approximate neuro-robots include: the H-7 of the University of Tokyo (Nishiwaki et al., 2007), Asimo of Honda Motor Corporation (ASIMO, 2011), Qrio of Sony Corporation (Carnegie Mellon Today, 2005), DB from Utah-based company Sarcos (Cheng, 2015), or HRP-2 from the Japanese Humanoid Robotics Program (Hirukawa et al., 2004), among others. Although representing high technological achievements, these robots have yet to

FIGURE 1
IHMC's Atlas was one of the robots that fell during the DARPA robotics challenge finals. Source: DARPA (Guizzoevan and Ackerman, 2015).

achieve high success and complete precision, for example, in terms of behavioral diversity and dexterity, as seen in humans. This is because they are not ready to exploit their system-environment interaction to a higher degree and are yet to fully develop in handling unplanned real-life situations (Pfeifer and Bongard, 2007).

Aside from lab competitions, experiments, and toy models, robots are currently deployed in several fields ranging from education, health, and agriculture to transport, hospitality, tourism, and the food industry. Technological capabilities, including robots, are expected to stimulate speedy process automation, improved service delivery, enhanced productivity and efficiency, advanced technological innovation and leadership, and higher economic growth (Romer, 1990; Khan, 2022). Several ethical and legal issues, such as control, autonomy, and regulation, concern the development and design of neuro-robots (Nyholm, 2022), but here we focus on neuro-robotic failures. As neuro-robots get deployed, neuro-robotic failures would cause severe repercussions to human counterparts in all consequential domains. Therefore, understanding how such robotic decisions are implemented is crucial in avoiding current and future neuro-robotic failures. What makes them go wrong? Why do neuro-robots decide in a particular manner? Why were they "too annoying" for the guests in the Hotel? What makes humanoid robots fail? These and many other questions merit our attention toward designing and exploring the issue of explainable robots.

Neural-inspired robots that deploy artificial intelligence (AI) and embodied intelligence (i.e., neuro-robots) need to be explained further. Situated in a natural environment, sensing their environment and acting on it, neuro-robots have control systems based on the principles of nervous systems (Krichmar, 2008). We would assume that robots (and neuro-robots) have become more capable than ever before because the cost of sensors has reduced significantly, and AI algorithms have matured exponentially (Silver et al., 2016). Yet, most AI algorithms, particularly the Blackbox deep neural networks, are still largely unexplainable (de Bruijn et al., 2022; Khan et al., 2022). Similarly, many brain processes are yet to be unknotted, not only in humans but also in animals with simple brains (Krichmar et al., 2019). Hence what follows is that the neuro-robots designed, taking into account sophisticated AI algorithms and human brain mimicry, are, to a large extent, unexplainable.

Explaining and understanding the human brain and how neural activity gives rise to behavior is a permanent subject of discussion and learning in the realm of neuroscience (Ungerleider and Haxby, 1994; Eickhoff et al., 2018; Vijayakumar et al., 2018; Klein et al., 2019; Chen K. et al., 2020). But in the case of AI, European Union (EU) law warrants AI to provide an immediate explanation as required by the "Right to Explanation" enacted in 2016, specifically in the context of adverse decisions affecting Europeans (ITIF, 2018). A neuro-robotic explanation is even more desirable in light of the EU Convention on Roboethics 2025 (Enlightenment of an Anchorwoman, 2010), the Institute of Electrical and Electronics Engineers (IEEE) Guideline for ethical designs 2019 (Bulan, 2019), and Japan's robotics principles (Pepito et al., 2019) inspired by Isaac Asimov's Three Laws of Robotics aiming to protect humans interacting with robots (Anderson, 2007). Furthermore, the latest push toward designing explainable and "responsible" automated systems comes from the White House Office of Science and Technology Policy (OSTP) in the US (The White House OSTP, 2022).

To explain a bit more, the EU's Convention requires that the design of all robots (including neuro-robots) must include provisions for the complete traceability of the robots' actions. Furthermore, the IEEE Guideline for Ethically Aligned Design asks that robot decisions must be transparent and supported by clear reasoning. Similarly, the need for explainability indirectly flows from Japan's "Ten Principles of Robot Law," which requires robots to serve humankind and robotic manufacturers to be held responsible for their creation. Finally, the recent passage of "Blueprint for an AI Bill of Rights" by the White House OSTP in the US also outlines crucial principles that should lead the design, application, and deployment of automated systems (including neuro-robots) to protect the US citizens in the age of AI (The White House OSTP, 2022).

While most of these policies (and guidelines) talk about general AI and robotics, the extent that neuro-robots are automated systems, they fall under the realm of these policies. However, neuro-robots are unique robots embodied in the environment. Also, as they mimic neural processing and most are morphologically similar to humans, they can be used to test brain theories in the laboratory and real-life settings while serving alongside human team members (Krichmar and Hwu, 2022). Policies around neuro-robots would thus need to be more nuanced and detailed. For example, how can we deploy a neuro-robot safely in a dynamic environment to enhance our understanding of neuroscience? Relatedly, can we deploy neuro-robots in a natural setting without any ethical approval in the first instance? Furthermore, as neuro-robots are complex and rely on enhanced human-robot interaction, how do we ensure that neuro-robots present a meaningful and trustworthy explanation in a human-friendly manner interpretable to both engineers and the public alike? Who will be held responsible if the neuro-robots go wrong, the neuro-robots or the developers? How do we gauge their performance in functions such as walking, singing, cooking, playing cards, or typing on keyboards? Do we benchmark their performance against human counterparts? Lastly, how much autonomy do we give them, and do we control them?

Modeled after biological brains, neuro-robots use a combination of neurally-inspired computing and AI. Neuro-robots operate on noisy and often uncertain data (Khan et al., 2022), to make decisions to help reduce the workload of fellow humans. When these robots work, they are of huge utility, allowing for, among other things, prosthetics (McMullen et al., 2014; Johannes et al., 2020; Handelman et al., 2022; Rejcek, 2022) and wearable systems supporting locomotion and learning processes (Colachis et al., 2018; Billard and Kragic, 2019). These systems can also engage in self-teaching modes allowing them to work like humans as waiters, deliverers, receptionists, troubleshooters, and digital assistants (Financial Times, 2022; Meta Fundamental AI Research Diplomacy Team (FAIR) et al., 2022).

However, as with human intelligence, sometimes robots fail to deliver, as mentioned earlier. In the instance of Henn na Hotel, the robots mistakenly took human snoring at night as a legit guest's request, thus prompting them to disturb the guests at night. In other words, the robots could not perform feature extraction and intelligent comprehension of human snoring. The failure of robots is not startling. Intelligence is the act of decision-making based on prevailing uncertainty (Khan et al., 2022). This fact differentiates robots deploying AI from non-intelligent decision systems based on the flow-chart design, as is the case in most electronics (Friedman and Zeckhauser, 2012). For human beings, such failures are vital for learning during childhood and adulthood. Neuro-robots using machine learning (ML) AI algorithms also require a "training phase" whereby the system is first trained on a human-labeled dataset (MIT Technology Review, 2018). The system then learns from its failures before being permitted to operate in the "wild" (MIT Technology Review, 2018). Therefore, it is understandable that, despite training, humans and neuro-robots might mis-categorize a new data episode that had never been seen or used.

In the case of human intelligence, only recently has neuroscience offered a mechanistic picture of the cellular basis of learning and memory (Liu et al., 2012). However, for neuro-robots, explaining why failures occur is not readily explainable. This is in spite of the EU regulations and other guidelines requiring that robotic actions' traceability and explainability be available to EU citizens to protect them from potential adverse effects.

For this paper, explainable neuro-robots are neuro-robots whose function and decision-making processes must be explained so that the average person can comprehend the basis of a robot behavior. Here we summarize the existing trends in neuro-robots and their underlying circuitry. Then we recap some literature surrounding failures neuro-robots may encounter. Since neuro-robots deploy AI and ML algorithms in their design, a summary of explainable AI methods and associated problems is provided in the following section. Finally, we will offer themes of how the explainability of neuro-robots may be researched as the field advances further.

2. Advances in neuro-robotics and robotic failures go hand in hand

While much of the basis for human higher cognition remains unknown, the existing understanding of biological brains deployed in biomimetic fashion in embodied intelligence, developmental robotics, and AI inform ways to design intelligent robot systems. Here we term these systems neuro-robots. Neuro-robots imitate aspects of the human brain in their design and function (Krichmar, 2018). Such robots can also interact with the nervous system of humans or other animals (Iosa et al., 2016). Neuro-robots differ from other electromechanical devices based on the ability to adapt their behavior on the basis of their experience (Iosa et al., 2016), a characteristic that has been termed "adaptability." Adaptability, in turn, is based on its multiple sensors, whose signals are processed by AI to change the robot's behavior (Iosa et al., 2016).

A neuro-robot can be designed for clinical uses, for instance, neurorehabilitation or neurosurgery (Nordin et al., 2017). They can also be developed for studying the nervous system by mimicking its properties (Krichmar, 2018; Chen K. et al., 2020), as happens in many walking robots based on central pattern generators (Ijspeert et al., 2007).

Immersed in the real world, a neuro-robot takes sensory information from the environment before integrating it into actions (Chen K. et al., 2020). This information intake and resulting computation inform how the artificial brain gives rise to new behaviors based on its experience. Researchers have used neuro-robotic approaches to study the neural correlates of visual perception (Priamikov et al., 2016), tactile perception (Pearson et al., 2007), auditory perception (Rucci et al., 1999), spatial navigation (Lambrinos et al., 2000), schema formation and consolidation (Hwu et al., 2020), neuromodulation (Sporns and Alexander, 2002), attention (Gigliotta et al., 2017), locomotion (Lock et al., 2013), language development (Oudeyer, 2006), and social interaction (Boucenna et al., 2014). Readers interested in how the behavior of neuro-robots help explain their neural control and the analysis of how neural activity leads to behavior may further refer to seminal research (Chen K. et al., 2020).

Neuro-robotic systems have advanced substantially in functions and properties because of significant progress in brain-inspired computing algorithms and hardware (Krichmar, 2018). On the functionality front, most neuro-robots perform single tasks in simple and static situations for now (Zou et al., 2020). But, then, a few multitasking robots perform in dynamic environments (Zou et al., 2020). As for properties, current neuro-robots exhibit features ranging from intelligent perception and flexible movement to interactions with environments (Sanders and Oberst, 2016).

The critical element of the software stack for neuro-robots is brain-inspired computing algorithms, which witnessed tremendous advancement in the past decade. Of those algorithms, two main categories are Artificial Neural Networks (ANNs) and Spiking Neural Networks (SNNs). While the human brain's hierarchical topologies and parallel-processing networks inspire ANNs (Yang and Wang, 2020), SNNs take inspiration from the patterns of neuronal action potentials subserving human brain function (Ghosh-Dastidar and Adeli, 2009). ANNs, specifically deep neural networks, are lauded for their phenomenal success in various machine-learning tasks (Zou et al., 2020). For instance, deep neural networks have already achieved human-level performance in image recognition (van Dyck et al., 2021). While most traditional robots use backpropagation to train ANNs (Hecht-Nielsen, 1989), neuro-robots usually model the ANNs around neuro-anatomically grounded Hebbian learning rules and algorithms (Garagnani et al., 2008). Similarly, SNNs are also very powerful computing and highly energy-efficient paradigms for processing dynamic sequential information (Cao et al., 2015; Zou et al., 2020). These paradigms possess desirable features such as high bio-fidelity, rich coding with complex data, and event-driven idiosyncrasy (Bichler et al., 2012; Zou et al., 2020).

Other important brain-inspired algorithms for robots include attractor neural networks (Solovyeva et al., 2016; Khona and Fiete, 2022). Such networks are recurrent dynamic networks, evolving toward a stable pattern (either single state, cyclic state, chaotic state, or random state) over time (Khona and Fiete, 2022). Attractor networks typically model neuronal processes such as memory, motor behavior, classification, and other biologically inspired processes in machine learning (Li et al., 2015).

The advancement of algorithms goes hand in hand with the development of neural computing hardware, also called neuromorphic architectures. On the one hand, we have neural network accelerators that optimize operations in ANNs and usually leverage parallel processing and efficient data compression (Zou et al., 2020). Examples of such accelerators include ShiDianNao (Du et al., 2015) and TPU (Jouppi et al., 2017). On the other hand, neuromorphic chips are designed to support rich spatiotemporal bio-functionality (Zou et al., 2020). Such chips provide high energy efficiency and event-driven representations. Examples include Neurogrid (Benjamin et al., 2014), SpiNNaker (Furber et al., 2013; Krichmar et al., 2019), IBM's TrueNorth under the SyNAPSE project (Merolla et al., 2011; Modha et al., 2011; DeBole et al., 2019), and the energy-aware computing hardware developed by HRL laboratories (Srinivasa and Cruz-Albrecht, 2012). Alongside neural accelerators and neuromorphic chips, we have many human-inspired robotic hardware (in humanoid platforms) that provides a maximum degree of anatomical fidelity to the human structure and is capable of whole-body motions (ECCE Robots, 2010; Ackerman, 2016b; WIRED, 2017). Examples of these platforms include ECCE (ECCE Robots, 2010; Bonsignorio, 2013), Kengoro (Ackerman, 2016b; WIRED, 2017), and HRP-2 (Hirukawa et al., 2004), among others.

These breakthroughs in algorithms and hardware led to the development of advanced neuro-robots that exhibit intelligent perception and flexible movement (Krichmar, 2018). Regarding functionality, we have two developmental designs: single-task robots and multitask robots (Zou et al., 2020). Single-task robots operate in simple scenarios with limited capability to perform multiple functions. On the other hand, multitask robots navigate dynamic systems and can perform multiple tasks simultaneously. Both these types of robots and systems are prevalent in different real-world applications, such as medical robots (Davies et al., 2000; Chen A. et al., 2020), prosthetic arms (Fazeli et al., 2019), humanoid platforms (Johansson et al., 2020), and automated driving (Spielberg et al., 2019). These applications offer critical opportunities to advance the design of robot systems further.

Neuro-robots deploying non-von Neumann architectures, specifically neuromorphic engineering, can provide low-power processing (on the order of milliwatts or watts, compared to kilowatts for a GPU) and sensing for autonomous systems. For example, the TrueNorth neuromorphic chip of IBM has deployed convolutional neural networks (CNNs) on autonomous robots and other embedded applications with minimal power consumption (Esser et al., 2016; Hwu et al., 2017). Similarly, neuromorphic architectures enable next-generation processing (Indiveri et al., 2011; James et al., 2017). Furthermore, by incorporating a processor-in-memory and event-based design, neuromorphic processors can provide three orders-of-magnitude strategic advantages in performance-per-watt while being robust to radiation effects.

While substantial effort has been invested in making robots more reliable in terms of power processing and dynamic sensing, experience demonstrates that frequent failures often challenge robots. According to researchers, the mean time between failure (MTBF) for robots in field environments is usually within a few hours (Tsarouhas and Fourlas, 2016). Regardless, robots have yet to reach a design that can better cater to fault management (Honig and Oron-Gilad, 2018). Even trained roboticists are not entirely aware of what causes the failure (Steinbauer, 2013). There is a plethora of literature on robotic failures (Laprie, 1995; Carlson and Murphy, 2005; Steinbauer, 2013; Barakova et al., 2015; Lemaignan et al., 2015; Brooks, 2017; Honig and Oron-Gilad, 2018). Failures refer to a degraded ability that causes the system's behavior or service to deviate from the standard or correct functionality (Honig and Oron-Gilad, 2018). Various errors and faults can cause failures in systems. For instance, the Henn na robots experienced failure due to speech-recognition errors. Similarly, in the autonomous car accident, the car crashed into a white truck due to intelligent feature extraction problems (Shepardson, 2017). Finally, Pepper and other humanoid robots failed due to weaker system-environment interaction and unreadiness to handle unplanned situations, among other related issues (Pfeifer and Bongard, 2007).

While failure detection and fault prediction techniques and algorithms in neuro-robots are still emerging, literature discusses failure detection and prediction in traditional robotic controls and manipulators. Several algorithms and methods can accomplish this: second-order sliding mode algorithm (Ferrara and Capisani, 2012), robust non-linear analytic redundancy technique (Halder and Sarkar, 2007), partial least square approach (Muradore and Fiorini, 2012), torque filtering and sensing technique (Fu and Cai, 2021), multiple model adaptive estimation method (Akca and Efe, 2019), multiple hybrid particle swarm optimization algorithm to realize multiple predictions failures (Ayari and Bouamama, 2017), and neural network for prediction of robot execution failures (Diryag et al., 2014). Identifying and understanding failures through the means mentioned are crucial in designing reliable robots that return meaningful explanations to users when and if needed. At the same time, we agree it may be impossible to identify (or predict) all sorts of robotic failures as robots operate in dynamic and unorganized environments interacting in numerous possible ways. This becomes even more challenging for neuro-robots situated in natural settings with many unplanned events. However, several researchers have advanced insightful failure classifications that may also be well relevant for neuro-robots.

Some researchers classified robotic failures into technical failures and social norm violations (Giuliani et al., 2015). Technical issues inside the robots cause technical failures. In contrast, social norm violations refer to inappropriate social cues, for instance, robots looking away from a person while talking to them. Other researchers categorized failures according to the source of the failure (Carlson and Murphy, 2005). Such classification considers physical failures (physical errors in the system's sensors, control system, or communications cause such failures) and human failures (human-made errors in design, for instance, cause these failures). Using information such as relevance (if the fault is relevant to various robotic systems), condition (context of failure), symptoms (indicators to identify the failure), impact (repairable or terminal, for instance), and frequency (how often it occurs), Steinbauer classified failures in two four categories following the RoboCup competitions: Interaction, Algorithms, Software, and Hardware (Steinbauer, 2013). Interaction failures arise from uncertainties in interacting with the environment and humans, whereas algorithmic failures are problems in methods and algorithms. Similarly, software failures are due to the design and implementation faults of software systems, whereas hardware failures are physical faults of robotic equipment.

Furthermore, Brooks classifies failures into communications and processing failures (Brooks, 2017). Communication failures are related to data being processed, including missing data (incomplete data), incorrect data (data distorted during transmission, for example), bad timing (data received too early or late), or extra data (for instance, data sent many times). Processing failures can happen due to poor logic (based on incorrect assumptions), ordering (when events occur in a different order), or abnormal termination due to unhandled exceptions or segmentation faults.

Other researchers echo some of these classifications and devise more inclusive human-robot failure categorizations (Honig and Oron-Gilad, 2018). Per their classification, there are two types of failures: technical (that includes both software and hardware) and interactive failures that arise because of uncertainties while interacting with the environment or agents in the environment, including humans. While all types of failures are important, interactive and algorithmic failures seem even more pertinent to neuro-robots serving and working alongside humans.

To use Laprie's words (Laprie, 1995), some of these failures are "catastrophic"—with a higher cost than the service—and thus, they need to be avoided. To avoid failures, the failures need to be understood and explained. The explanation for most of these failures, particularly the ones relating to interaction and algorithms, is not readily available. An explainable neuro-robot will be expected to unravel some of this explanation to end users and engineers alike.

3. Explainability in AI limits explainability in neuro-robots

Neuro-robots are useful but still not explainable as they employ a combination of AI algorithms and neural-inspired hardware while interacting with environments and agents in evolving settings. In addition, as an interdisciplinary field, neuro-robotics involves control and mechatronics, among other areas. Although in the previous section, we noted some failures and failure detection techniques regarding the control, design, and engineering, we limit ourselves to algorithms for the purpose of this discussion. The challenges inherent in neural-inspired hardware are attributed to the fact that neuroscience itself lacks a complete theory of brain function, which is further compounded by the sheer physical complexity of biological brains (10^{15} computational synaptic elements in the human brain). For AI, the explanation of why failures happen is not easily available (Gunning and Aha, 2019). Explainability in neuro-robots is thus largely limited by the explainability in AI. However, robotic (neuro-robotic) failures identified earlier warrant explanation as they impact lives and livelihoods in many consequential ways. We propose as AI explainability increases, neuro-robots will become more explainable. Thus, this section draws an overview of current explainable AI methods, which provide a foundation for explainable neuro-robots.

Machine learning (ML) explainable techniques (or X-AI methods) offer an understanding and explanation of ML models' decisions in human terms to establish trust with stakeholders, including engineers, users, and policymakers. In the past decade, with the application of AI in several autonomous systems and robots, we have seen a tremendous amount of research interest in X-AI methods. Currently, we can choose from a suite of X-AI methods to untangle deep learning opaque models (Lipton, 2017; Došilović et al., 2018; Xu et al., 2019; Holzinger et al., 2022; Khan et al., 2022). There are various categorizations of X-AI methods based on several criteria, including structure, design transparency, agnosticness, scope, supervision, explanation type, and data type, as listed in **Table 1** (Khan et al., 2022).

Most popular methods include intrinsic vs. *post hoc*, Blackbox vs. Whitebox vs. Graybox approaches, local vs. global approaches, model agnostic vs. model specific approaches, supervised vs. unsupervised methods, and those methods that differ in explanation type or the data type they can handle.

Overall, these excellent foundational methods [summarized in **Table 1**; for more details, readers may consult (Khan et al., 2022)] help produce some model understanding and present bits of human interpretable understanding. However, there is still no comprehensive understanding of how an AI implements a decision while explaining the model decision (Khan et al., 2022). These methods are far from perfect. High-stake ML deployment failures from these Blackbox models underpin the idea that these models fail to offer a satisfactory level of explanation (Ackerman, 2016a; Strickland, 2019; Su et al., 2019). The failures further underscore that these models are uncontestable, opaque, display unpredictable behavior, and in some situations, boost undesirable racial, gender, and demographic biases (Newman, 2021). Consequential settings such as healthcare, criminal justice, and banking have already witnessed substantial harm because of the issues found in Blackbox methods (Newman, 2021). Whitebox models, on the other hand, are also not the best: while these models are more interpretable, they are less accurate (Loyola-González, 2019).

Beyond an incomplete explanation, the explanation is unstable (Bansal et al., 2020; Kozyrkov, 2021). For instance, four X-AI methods were deployed to determine what makes a matchstick a matchstick (Bansal et al., 2020). By changing only a single parameter, the methods returned twelve explanations, suggesting the methods are unstable (Choi, 2021). Even state-of-the-art techniques such as Local Interpretable Model-agnostic Explanation (LIME) and Shapley Additive exPlanations (SHAP) deploying Local Linear Explanations (LLE) also suffer from defects, including unstable explanations after changing a single parameter or even different explanations for the same data point (Amparore et al., 2021). As these methods use some randomized algorithm, for example, the Monte Carlo algorithm,

TABLE 1 Summary of existing explainable AI (X-AI) methods.

Criterion	Types	Definitions	Examples
Structure Relates to the complexity of ML models	Intrinsic (Xu et al., 2019; Holzinger et al., 2022)	Easily interpretable model because of their simple structure	Linear regression, logistic regression, decision trees, and k-nearest neighbors
	Post hoc (Madsen et al., 2022; Yera et al., 2022)	Complex structure models that attain interpretability after model training.	Permutation feature importance and neural networks
Transparency in design Relates to the design of a method	Whitebox (Levashenko et al., 2016; Garibaldi, 2019; Loyola-González, 2019; Zaitseva et al., 2020)	By design, Whitebox approach is more transparent and explainable.	Simple decision trees, rule-based models, patterns-based models, linear regression models, bayesian networks, fuzzy cognitive maps, and those following fuzzy logic such as fuzzy decision trees and fuzzy rules-based models
	Blackbox (Robnik-Šikonja and Kononenko, 2008; Loyola-González, 2019)	Blackbox approach contains complex mathematical functions like support-vector machine and neuronal networks. Generally, they are hard to understand and explain.	Deep neural networks and random forests
	Greybox (Pintelas et al., 2020)	Such approaches have features of both Blackbox and Whitebox approaches.	Local Interpretable Model-agnostic Explanations (LIME) and Interpretable Mimic Learning
Scope Relates to the scope of the interpretability	Local (Ribeiro et al., 2016; Lundberg and Lee, 2017)	The scope of interpretability is limited to individual predictions or a small portion of the model prediction space.	Local Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), and Individual Conditional Expectation (ICE)
	Global (Lundberg et al., 2020; Machlev et al., 2022)	Global methods cover the entire model prediction space.	Partial Dependence Plot (PDP) and Accumulated Local Effects
Agnosticity Classification based on the level of agnostic-ness	Model agnostic (Machlev et al., 2022)	Their X-AI algorithm can be applied to any kind of ML model. They do not depend on model internals.	SHAP and LIME
	Model specific (Belle and Papantonis, 2021)	Methods are designed for specific types of ML model.	Neural network methods
Supervision Classification based on the degree of supervision	Supervised (Ancona et al., 2019)	Entail an active manipulation of input data.	LIME, SHAP, integrated gradients, smoothgrad, layer-wise relevance propagation, and perturbation methods
	Unsupervised (Lei et al., 2016; Esmaeili et al., 2019)	Researchers assume no explicit annotations about input data.	Rationale and disentanglement representations
Explanation or data type Additional methods based on type of explanation or data	Explanation type (Belle and Papantonis, 2021; Holzinger et al., 2022)	Explanation output differs in each method. For instance, feature summary return feature statistics.	Feature summary, surrogate models, extract concepts, decision rules, correlation plots, and other visualizations
	Data type (Welling et al., 2016; Hendricks et al., 2018; Pawelczyk et al., 2020)	Classification based on the data type a method can handle.	Graph, image, text/speech, and tabular

to explain decisions and predictions—like deep neural networks mislabeling the image of a lion as a library (Szegedy et al., 2014) or AI failing to predict a husky on the snow (Ribeiro et al., 2016), or even AI taking a horse for a frog (Su et al., 2019)—the resulting explanation may vary (Amparore et al., 2021).

The variability and heterogeneity in explanations stem largely from model dynamics, algorithms, and internal mechanics. Beyond these issues, X-AI methods suffer from external validity issues and cannot handle all sorts of data and environments. The validity issue is even more severe when the methods employed to the same data generate different predictions, as mentioned in the example of matchstick prediction (Bansal et al., 2020). Such external validity issues will be a matter of deep concern with the increasing application of AI to human ML systems in healthcare, education, justice, defense, and security.

A recent article comprehensively outlines the multidimensional challenges faced by X-AI methods (de Bruijn et al., 2022). Some of these issues pertain to data and decision dynamics (varying data and decisions lead to different explanations) and dependency on context (since outcomes may differ for various individuals, general explanations for algorithms may not work). Other challenges speak to the "wicked" nature of the problems (the poorly designed nature of the problems requires multiple answers versus a single answer that current algorithms furnish) and the contested nature of explanations due to biasedness, among other concerns.

As neuro-robots get deployed in social settings and encounter obstacles and humans, they will most likely make mistakes in their activities, such as walking (neuro-robotic navigation). To reiterate, the explainability of neuro-robots' behavior is needed, and even more so from a safety perspective, as we do not want humans to get scared or hurt by neuro-robots. When applied to neuro-robots, some X-AI methods, such as LIME, SHAP, data type-based methods, and neural networks, can offer initial insights into neuro-robotic behavior just like they provided explanations in traditional robotics. Earlier work on robots produced verbal and natural language explanations (data-based explanations) of robotic navigation (Perera et al., 2016; Rosenthal et al., 2016; Stein, 2021). Such works deployed algorithms that translated sensor data into natural language. While these algorithms are interpretable, the entire narration of trajectory may not be appealing. Unlike natural language explanation, Halilovic and Lindner (2022) offer visual explanations using LIME (taking image data as input) to explain deviations from desired navigation path. While we know how LIME explains predictions of a Blackbox model by learning an interpretable model around a deviation (Halilovic and Lindner, 2022), LIME takes more time to generate an explanation (Halilovic and Lindner, 2022) as well as performs better when generating an explanation for a single prediction (Banerjee, 2020). Thus, LIME perhaps cannot be used solely to generate multiple and varying explanations needed from neuro-robots as they operate in fast-evolving complex environments. LIME, however, may be used down the road effectively once its runtime is improved. Also, this improved version of LIME may be used in conjunction with other methods, such as the Partial Dependence Plot (PDP), to generate an explanation of the entire model in neuro-robotics.

Other researchers use neural networks (specifically, deep reinforcement learning) (He et al., 2021) or fuzzy algorithms (Bautista-Montesano et al., 2020) to explain robotic navigation. But, again, while these are great methods, the former explanation is too complex for non-specialists, whereas the latter explanation makes a strong assumption about the underlying algorithm.

Relatedly, many researchers have used model-agnostic X-AI methods to enhance expert understanding of deep learning Blackbox models and their outputs generated in traditional robots (Raman and Kress-Gazit, 2013; Adadi and Berrada, 2018; Rai, 2020); other studies used methods that focused only on classification-based tasks (Ribeiro et al., 2016; Wu et al., 2017), including the application of saliency map by Huber et al. to improve the interpretation of image classification tasks (Huber et al., 2022) and feature extraction to determine the effect of an individual feature on utility function (Elizalde et al., 2008). As opposed to these studies and to account for sequential decision-making that robots perform, researchers employed a framework for generating linguistic explanations from robot's state sequences using an encoder-decoder model (Ehsan et al., 2019; Das et al., 2021). Whereas the former research studies generate primarily expert-centric explanations, the latter user-centric studies seem more promising in the context of neuro-robots as they account for the complexity of sequential decision-making carried out by such robots while interacting with users and objects in the environment.

In real life, neuro-robots will encounter multiple challenges and obstacles. Aside from deviations in navigation and task classification in static settings, there will be other collapses and breakdowns in dynamic environments. Also, the neuro-robots will be challenged to generate meaningful explanations using real-time data and recall a memory from past experiences. In addition, they will face ethical dilemmas. While the present trajectory of X-AI methods holds great promise for the design of transparent robots in labs, they are not yet powerful enough to fully incorporate explanations for unexpected events, instances, and failures.

The issues, from unstable to complex explanations and those around data and decision dynamics mentioned in this section, limit the explanatory power of neuro-robots, particularly those deployed in social settings facing unexpected hiccups and situations. As these embodied neuro-robots employ AI to sense and act on their environment, and AI's outcome explanation is deficient, neuro-robots will lack the explanation required by users, engineers, and policymakers alike. Thus, to make neuro-robots more explainable, one way researchers can accomplish this goal is to develop a robust neural framework that explains the AI outcome to satisfaction, as suggested by the authors in their recent paper (Khan et al., 2022).

In its design, such a neural framework will be inspired by biological mnemonic function to produce an explanation. In biological brains, the members of a cell assembly, by their act of firing action potentials together, are involved in memory formation (i.e., an engram)—this essentially constitutes an explanation. The active cells hence identified are then deactivated optogenetically to reversibly control the recall of the specific memory (Liu et al., 2012). This mnemonic function is considered a particular instance of decision-making since each decision requires a corresponding memory. The neural framework mentioned here, while still needing to be tested, nevertheless provides a way forward for offering a fuller explanation. If successful, neuro-robots may also deploy this framework to offer explanations in the events they are required to give one.

4. Toward explainable neuro-robots

Traditional AI systems are evolving exponentially. Some of these systems engage in self-teaching modes that permit them to surpass human capabilities at games like chess and Go (Silver et al., 2016, 2017) and, most recently, Diplomacy (Buch et al., 2022; Financial Times, 2022; Meta Fundamental AI Research Diplomacy Team (FAIR) et al., 2022). Moreover, neuromorphic architectures are also undergoing further advances (DeBole et al., 2019; Olds, 2019). New chips, such as Intel's Loihi, are being developed to support embedded neuromorphic applications (Davies et al., 2018). In addition to running neural networks on specialized hardware, very low-power neuromorphic vision and auditory sensors are being developed (Liu and Delbruck, 2010; Stewart et al., 2016). Similar to biology, these sensors and processors only respond to change or salient events. When they do respond, it is with a train of precisely timed spikes, like a neurobiological system. The event-driven nature leads to ideal power efficiency for autonomous systems and robots.

With all these new advances and developments, neuro-robots will further mature in design and function. However, increasing instances of robotic failures require neuro-robots to furnish an explanation that uses real-time data and previous experiences. Moreover, since neuro-robots rely massively on human-robot interactions, the explanation would have to be trustworthy and offered in a friendly manner. Finally, as outlined earlier, legislative pressures in Europe and America alongside IEEE guidelines make the design and use of explainable neuro-robots and the emerging explanation even more urgent. Thus, researchers designing neuro-robots must ensure that robotic design and deployment integrate explanatory capability as a core principle.

Accomplishing the next steps will not be easy. The ecosystem of jurisdictions that might adopt appropriate policies to accomplish the above is incredibly diverse, ranging from local to transnational entities and even including Outer Space. Further, there is no current consensus within the neuro-robotics community. Finally, as we have outlined above, the ontology of the field itself is complex. Nevertheless, there are models that might be useful to consider. One of those involves the regulation of pharmaceuticals and medical devices across international boundaries. As with neuro-robotics, the potential to do harm is significant. Additionally, there are important external Incentives for the coordination of policies (e.g., COVID-19 vaccines). Therefore, we recommend creating an international body under the auspices of the United Nations to coordinate a "policy bridge" between researchers and international stakeholders analogous to the World Health Organization (WHO). Such an organization could both catalyze the development of uniform standards and language ontologies relevant to the problem and draft model regulations that jurisdictions might adopt either in whole or in part.

In conclusion, we have reviewed the current state of play in the field of X-AI within the context of neuro-robotics. Because of the significant consequence of robotic failure on human (and even planetary) welfare, it is imperative to move forward, notwithstanding the challenges and complexity of the field.

Author contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ackerman, E. (2016a). *Fatal tesla self-driving car crash reminds us that robots aren't perfect. IEEE spectrum*. Available online at: https://spectrum.ieee.org/fatal-tesla-autopilot-crash-reminds-us-that-robots-arent-perfect (accessed January 5, 2022).

Ackerman, E. (2016b). *This robot can do more push-ups because it sweats IEEE spectrum*. Available online at: https://spectrum.ieee.org/this-robot-can-do-more-pushups-because-it-sweats (accessed January 5, 2022).

Adadi, A., and Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* 6, 52138–52160. doi: 10.1109/ACCESS.2018.2870052

Akca, A., and Efe, M. (2019). Multiple model kalman and particle filters and applications: A survey. *IFAC PapersOnLine* 52, 73–78. doi: 10.1016/j.ifacol.2019.06.013

Amparore, E., Perotti, A., and Bajardi, P. (2021). To trust or not to trust an explanation: Using LEAF to evaluate local linear XAI methods. *PeerJ Comput. Sci.* 7:e479. doi: 10.7717/peerj-cs.479

Ancona, M., Ceolini, E., Öztireli, C., and Gross, M. (2019). "Gradient-based attribution methods," in *Explainable AI: Interpreting, explaining and visualizing deep learning*, eds W. Samek, G. Montavon, A. Vedaldi, L. Hansen, and K. Müller (Cham: Springer), 169–191. doi: 10.1007/978-3-030-28954-6_9

Anderson, M. (2007). *After 75 years, isaac Asimov's three laws of robotics need updating*. Victoria: The Conversation.

ASIMO, (2011). *Honda unveils all-new ASIMO with significant advancements*. Available online at: https://asimo.honda.com/news/honda-unveils-all-new-asimo-with-significant-advancements/newsarticle_0125/ (accessed January 5, 2022).

Ayari, A., and Bouamama, S. (2017). A new multiple robot path planning algorithm: Dynamic distributed particle swarm optimization. *Robotics Biomim.* 4:8. doi: 10.1186/s40638-017-0062-6

Banerjee, P. (2020). *Explain your model predictions with LIME*. Available online at: https://kaggle.com/code/prashant111/explain-your-model-predictions-with-lime (accessed January 5, 2022).

Bansal, N., Agarwal, C., and Nguyen, A. (2020). SAM: The sensitivity of attribution methods to hyperparameters. *Arxiv [Preprint]* doi: 10.1109/CVPR42600.2020.00870

Barakova, E., Bajracharya, P., Willemsen, M., Lourens, T., and Huskens, B. (2015). Long-term LEGO therapy with humanoid robot for children with ASD. *Expert Syst.* 32, 698–709. doi: 10.1111/exsy.12098

Bautista-Montesano, R., Bustamante-Bello, R., and Ramirez-Mendoza, R. (2020). Explainable navigation system using fuzzy reinforcement learning. *Int. J. Interact. Des. Manuf.* 14, 1411–1428. doi: 10.1007/s12008-020-00717-1

Belle, V., and Papantonis, I. (2021). Principles and practice of explainable machine learning. *Front. Big Data* 4:688969. doi: 10.3389/fdata.2021.688969

Benjamin, B., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A., Bussat, J., et al. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proc. IEEE* 102, 699–716. doi: 10.1109/JPROC.2014.2313565

Bichler, O., Querlioz, D., Thorpe, S., Bourgoin, J., and Gamrat, C. (2012). Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. *Neural Netw.* 32, 339–348. doi: 10.1016/j.neunet.2012.02.022

Billard, A., and Kragic, D. (2019). Trends and challenges in robot manipulation. *Science* 364:eaat8414. doi: 10.1126/science.aat8414

Bonsignorio, F. (2013). Quantifying the evolutionary self-structuring of embodied cognitive networks. *Artif. Life* 19, 267–289. doi: 10.1162/ARTL_a_00109

Boucenna, S., Gaussier, P., Andry, P., and Hafemeister, L. (2014). A robot learns the facial expressions recognition and face/non-face discrimination through an imitation game. *Int. J. Soc. Robot.* 6, 633–652. doi: 10.1007/s12369-014-0245-z

Brooks, D. (2017). *A human-centric approach to autonomous robot failures*. Ph.D. thesis. Lowell, MA: University of Massachusetts Lowell.

Buch, A. M., Eagleman, D. M., and Grosenick, L. (2022). *Engineering diplomacy: How AI and human augmentation could remake the art of foreign relations*. Washington, DC: Science & Diplomacy. doi: 10.1126/scidip.ade6798

Bulan, A. (2019). *IEEE launches ethically aligned design, first edition, delivering "a vision for prioritizing human well-being with autonomous and intelligent systems"*. Piscataway, NJ: IEEE Standards Association.

Cao, Y., Chen, Y., and Khosla, D. (2015). Spiking deep convolutional neural networks for energy-efficient object recognition. *Int. J. Comput. Vis.* 113, 54–66. doi: 10.1007/s11263-014-0788-3

Carlson, J., and Murphy, R. (2005). How UGVs physically fail in the field. *IEEE Trans. Robot.* 21, 423–437. doi: 10.1109/TRO.2004.838027

Carnegie Mellon Today (2005). *Sony's humanoid robot "QRIO" entertains campus community*. Available online at: https://www.cmu.edu/cmnews/extra/050128_qrio.html (accessed January 5, 2022).

Chen, A., Balter, M., Maguire, T., and Yarmush, M. (2020). Deep learning robotic guidance for autonomous vascular access. *Nat. Mach. Intell.* 2, 104–115. doi: 10.1038/s42256-020-0148-7

Chen, K., Hwu, T., Kashyap, H., Krichmar, J., Stewart, K., Xing, J., et al. (2020). Neurorobots as a means toward neuroethology and explainable AI. *Front. Neurorobot.* 14:570308. doi: 10.3389/fnbot.2020.570308

Cheng, G. (2015). *Humanoid robotics and neuroscience: Science, engineering, and society*. Boca Raton, FL: CRC Press. doi: 10.1201/b17949-3

Choi, C. Q. (2021). 7 revealing ways AIs fail. *IEEE Spectr.* 58, 42–47. doi: 10.1109/MSPEC.2021.9563958

Colachis, S., Bockbrader, M., Zhang, M., Friedenberg, D., Annetta, N., Schwemmer, M., et al. (2018). Dexterous control of seven functional hand movements using cortically-controlled transcutaneous muscle stimulation in a person with tetraplegia. *Front. Neurosci.* 12:208. doi: 10.3389/fnins.2018.00208

Das, D., Banerjee, S., and Chernova, S. (2021). "Explainable AI for robot failures: Generating explanations that improve user assistance in fault recovery," in *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*, (New York, NY: Association for Computing Machinery), 351–360. doi: 10.1145/3434073.3444657

Davies, B., Starkie, S., Harris, S., Agterhuis, E., Paul, V., and Auer, L. (2000). "Neurobot: A special-purpose robot for neurosurgery," in *Proceedings of the 2000 ICRA millennium conference IEEE international conference on robotics and automation symposia proceedings (Cat No00CH37065)*, (San Francisco, CA: IEEE), 4103–4108. doi: 10.1109/ROBOT.2000.845371

Davies, M., Srinivasa, N., Lin, T., Chinya, G., Cao, Y., Choday, S., et al. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* 38, 82–99. doi: 10.1109/MM.2018.112130359 de Bruijn, H., Warnier, M., and Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Gov. Inf. Q.* 39:101666. doi: 10.1016/j.giq.2021.101666

DeBole, M., Taba, B., Amir, A., Akopyan, F., Andreopoulos, A., Risk, W., et al. (2019). TrueNorth: Accelerating from zero to 64 million neurons in 10 years. *Computer* 52, 20–29. doi: 10.1109/MC.2019.2903009

Diryag, A., Mitić, M., and Miljković, Z. (2014). Neural networks for prediction of robot failures. *J. Mech. Eng. Sci.* 228, 1444–1458. doi: 10.1177/0954406213507704

Došilović, F., Brčić, M., and Hlupić, N. (2018). "Explainable artificial intelligence: A survey," in *Proceedings of the 2018 41st international convention on information and communication technology, electronics and microelectronics (MIPRO)*, (Chiayi City: MIPRO), 210–215. doi: 10.23919/MIPRO.2018.8400040

Du, Z., Fasthuber, R., Chen, T., Ienne, P., Li, L., Luo, T., et al. (2015). "ShiDianNao: Shifting vision processing closer to the sensor," in *Proceedings of the 2015 ACM/IEEE 42nd annual international symposium on computer architecture (ISCA)*, (Dublin: ISCA), 92–104. doi: 10.1145/2749469.2750389

ECCE Robots (2010). *ECCE ROBOTS: Your guide to the world of robotics*. Available online at: https://robots.ieee.org/robots/ecce/ (accessed January 5, 2022).

Ehsan, U., Tambwekar, P., Chan, L., Harrison, B., and Riedl, M. (2019). Automated rationale generation: A technique for explainable AI and its effects on human perceptions. *Arxiv [Preprint]* doi: 10.1145/3301275.3302316

Eickhoff, S., Yeo, B., and Genon, S. (2018). Imaging-based parcellations of the human brain. *Nat. Rev. Neurosci.* 19, 672–686. doi: 10.1038/s41583-018-0071-7

Elizalde, F., Sucar, L., Luque, M., Díez, F., and Reyes Ballesteros, A. (2008). "Policy explanation in factored Markov decision processes," in *Proceedings of the 4th European workshop on probabilistic graphical models, PGM 2008*, Hirtshals, 97–104.

Enlightenment of an Anchorwoman (2010). *European union's convention on Roboethics 2025*. Available online at: https://akikok012um1.wordpress.com/european-union%E2%80%99s-convention-on-roboethics-2025/ (accessed January 5, 2022).

Esmaeili, B., Wu, H., Jain, S., Bozkurt, A., Siddharth, N., Paige, B., et al. (2019). "Structured disentangled representations," in *Proceedings of the twenty-second international conference on artificial intelligence and statistics*, (Cambridge MA: PMLR), 2525–2534.

Esser, S., Merolla, P., Arthur, J., Cassidy, A., Appuswamy, R., Andreopoulos, A., et al. (2016). Convolutional networks for fast, energy-efficient neuromorphic computing. *Proc. Natl. Acad. Sci. U.S.A.* 113, 11441–11446. doi: 10.1073/pnas.1604850113

Fazeli, N., Oller, M., Wu, J., Wu, Z., Tenenbaum, J., and Rodriguez, A. (2019). See, feel, act: Hierarchical learning for complex manipulation skills with multisensory fusion. *Sci. Robot.* 4:eaav3123. doi: 10.1126/scirobotics.aav3123

Ferrara, A., and Capisani, L. (2012). "Second order sliding modes to control and supervise industrial robot manipulators," in *Sliding modes after the first decade of the 21st century: State of the art (Lecture notes in control and information sciences)*, eds L. Fridman, J. Moreno, and R. Iriarte (Berlin: Springer), 541–567. doi: 10.1007/978-3-642-22164-4_20

Financial Times (2022). *A machiavellian machine raises ethical questions about AI*. London: Financial Times.

Friedman, J., and Zeckhauser, R. (2012). Assessing uncertainty in intelligence. *Intell. Natl. Secur.* 27, 824–847. doi: 10.1080/02684527.2012.708275

Fu, B., and Cai, G. (2021). Design and calibration of a joint torque sensor for robot compliance control. *IEEE Sens. J.* 21, 21378–21389. doi: 10.1109/JSEN.2021.3104351

Furber, S., Lester, D., Plana, L., Garside, J., Painkras, E., Temple, S., et al. (2013). Overview of the SpiNNaker system architecture. *IEEE Trans. Comput.* 62, 2454–2467. doi: 10.1109/TC.2012.142

Garagnani, M., Wennekers, T., and Pulvermüller, F. (2008). A neuroanatomically grounded Hebbian-learning model of attention–language interactions in the human brain. *Eur. J. Neurosci.* 27, 492–513. doi: 10.1111/j.1460-9568.2008.06015.x

Garibaldi, J. (2019). The need for fuzzy AI. *IEEE CAA J. Autom. Sin.* 6, 610–622. doi: 10.1109/JAS.2019.1911465

Ghosh-Dastidar, S., and Adeli, H. (2009). Spiking neural networks. *Int. J. Neural Syst.* 19, 295–308. doi: 10.1142/S0129065709002002

Gigliotta, O., Seidel Malkinson, T., Miglino, O., and Bartolomeo, P. (2017). Pseudoneglect in visual search: Behavioral evidence and connectional constraints in simulated neural circuitry. *eNeuro* 4:ENEURO.0154-17.2017. doi: 10.1523/ENEURO.0154-17.2017

Giuliani, M., Mirnig, N., Stollnberger, G., Stadler, S., Buchner, R., and Tscheligi, M. (2015). Systematic analysis of video data from different human–robot interaction studies: A categorization of social signals during error situations. *Front. Psychol.* 6:931. doi: 10.3389/fpsyg.2015.00931

Guizzoevan, E., and Ackerman, E. (2015). *DARPA robotics challenge: A compilation of robots falling down-IEEE spectrum*. Available online at: https://spectrum.ieee.org/darpa-robotics-challenge-robots-falling (accessed January 5, 2022).

Gunning, D., and Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Mag.* 40, 44–58. doi: 10.1145/3301275.3308446

Halder, B., and Sarkar, N. (2007). Robust nonlinear analytic redundancy for fault detection and isolation in mobile robot. *Int. J. Automat. Comput.* 4, 177–182. doi: 10.1007/s11633-007-0177-2

Halilovic, A., and Lindner, F. (2022). "Explaining local path plans using LIME," in *Advances in service and industrial robotics. RAAD 2022. Mechanisms and machine science*, Vol. 120, eds A. Müller and M. Brandstötter (Cham: Springer), 106–113. doi: 10.1007/978-3-031-04870-8_13

Handelman, D., Osborn, L., Thomas, T., Badger, A., Thompson, M., Nickl, R., et al. (2022). Shared control of bimanual robotic limbs with a brain-machine interface for self-feeding. *Front. Neurorobot.* 16:918001. doi: 10.3389/fnbot.2022.918001

He, L., Aouf, N., and Song, B. (2021). Explainable deep reinforcement learning for UAV autonomous path planning. *Aerosp. Sci. Technol.* 118:107052. doi: 10.1016/j.ast.2021.107052

Hecht-Nielsen, R. (1989). "Theory of the backpropagation neural network," in *Proceedings of the international 1989 joint conference on neural networks (IJCNN)*, Washington, DC, 593–605. doi: 10.1109/IJCNN.1989.118638

Hendricks, L., Hu, R., Darrell, T., and Akata, Z. (2018). Grounding visual explanations. *Arxiv [Preprint]* doi: 10.1007/978-3-030-01216-8_17

Hertzfeld, E. (2019). *Japan's Henn na hotel fires half its robot workforce*. Available online at: https://www.hotelmanagement.net/tech/japan-s-henn-na-hotel-fires-half-its-robot-workforce (accessed January 5, 2022).

Hirukawa, H., Kanehiro, F., Kaneko, K., Kajita, S., Fujiwara, K., Kawai, Y., et al. (2004). Humanoid robotics platforms developed in HRP. *Robot. Auton. Syst.* 48, 165–175. doi: 10.1016/j.robot.2004.07.007

Holzinger, A., Saranti, A., Molnar, C., Biecek, P., and Samek, W. (2022). "Explainable AI methods - A brief overview," in *Beyond explainable AI. XXAI 2020. Lecture notes in computer science()*, Vol. 13200, eds A. Holzinger, R. Goebel, R. Fong, T. Moon, K. Müller, and W. Samek (Cham: Springer), 13–38. doi: 10.1007/978-3-031-04083-2_2

Honig, S., and Oron-Gilad, T. (2018). Understanding and resolving failures in human-robot interaction: Literature review and model development. *Front. Psychol.* 9:861. doi: 10.3389/fpsyg.2018.00861

Huber, T., Limmer, B., and André, E. (2022). Benchmarking perturbation-based saliency maps for explaining atari agents. *Front. Artif. Intell.* 5:903875. doi: 10.3389/frai.2022.903875

Hwu, T., Isbell, J., Oros, N., and Krichmar, J. (2017). "A self-driving robot using deep convolutional neural networks on neuromorphic hardware," in *Proceedings of the 2017 international joint conference on neural networks (IJCNN)*, (Anchorage, AK: IEEE), 635–641. doi: 10.1109/IJCNN.2017.7965912

Hwu, T., Kashyap, H., and Krichmar, J. (2020). "A neurobiological schema model for contextual awareness in robotics," in *Proceedings of the 2020 international joint conference on neural networks (IJCNN)*, Glasgow, 1–8. doi: 10.1109/IJCNN48605.2020.9206858

Ijspeert, A., Crespi, A., Ryczko, D., and Cabelguen, J. (2007). From swimming to walking with a salamander robot driven by a spinal cord model. *Science* 315, 1416–1420. doi: 10.1126/science.1138353

Indiveri, G., Linares-Barranco, B., Hamilton, T., van Schaik, A., Etienne-Cummings, R., Delbruck, T., et al. (2011). Neuromorphic silicon neuron circuits. *Front. Neurosci.* 5:73. doi: 10.3389/fnins.2011.00073

Iosa, M., Morone, G., Cherubini, A., and Paolucci, S. (2016). The three laws of neurorobotics: A review on what neurorehabilitation robots should do for patients and clinicians. *J. Med. Biol. Eng.* 36, 1–11. doi: 10.1007/s40846-016-0115-2

ITIF (2018). *The impact of the EU's new data protection regulation on AI*. Washington, DC: ITIF.

James, C., Aimone, J., Miner, N., Vineyard, C., Rothganger, F., Carlson, K., et al. (2017). A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications. *Biologically Inspired Cogn. Architectures* 19, 49–64. doi: 10.1016/j.bica.2016.11.002

Johannes, M., Faulring, E., Katyal, K., Para, M., Helder, J., Makhlin, A., et al. (2020). "Chapter 21 - The modular prosthetic limb," in *Wearable robotics*, eds J. Rosen and P. Ferguson (Cambridge, MA: Academic Press), 393–444. doi: 10.1016/B978-0-12-814659-0.00021-7

Johansson, B., Tjøstheim, T., and Balkenius, C. (2020). Epi: An open humanoid platform for developmental robotics. *Int. J. Adv. Robot. Syst.* 17:1729881420911498. doi: 10.1177/1729881420911498

Jouppi, N., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., et al. (2017). "In-Datacenter performance analysis of a tensor processing unit," in *Proceedings of the 44th annual international symposium on computer architecture*, (New York, NY: Association for Computing Machinery), 1–12. doi: 10.1145/3079856.3080246

Khan, M. (2022). Absorptive capacities and economic growth in low and middle income economies. *Struct. Chang. Econ. Dyn.* 62, 156–188. doi: 10.1016/j.strueco.2022.03.015

Khan, M., Nayebpour, M., Li, M., El-Amine, H., Koizumi, N., and Olds, J. (2022). Explainable AI: A neurally-inspired decision stack framework. *Biomimetics* 7:127. doi: 10.3390/biomimetics7030127

Khona, M., and Fiete, I. (2022). Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* 23, 744–766. doi: 10.1038/s41583-022-00642-0

Klein, H., McCabe, C., Gjoneska, E., Sullivan, S., Kaskow, B., Tang, A., et al. (2019). Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.* 22, 37–46. doi: 10.1038/s41593-018-0291-1

Kozyrkov, C. (2021). *Explainable AI won't deliver. Here's why*. Available online at: https://kozyrkov.medium.com/explainable-ai-wont-deliver-here-s-why-6738f54216be (accessed January 5, 2022).

Krichmar, J. (2008). Neurorobotics. *Scholarpedia* 3:1365. doi: 10.4249/scholarpedia.1365

Krichmar, J. (2018). Neurorobotics—A thriving community and a promising pathway toward intelligent cognitive robots. *Front. Neurorobot.* 12:42. doi: 10.3389/fnbot.2018.00042

Krichmar, J., and Hwu, T. (2022). Design principles for neurorobotics. *Front. Neurorobot.* 16:882518. doi: 10.3389/fnbot.2022.882518

Krichmar, J., Severa, W., Khan, M., and Olds, J. (2019). Making BREAD: Biomimetic strategies for artificial intelligence now and in the future. *Front. Neurosci.* 13:666. doi: 10.3389/fnins.2019.00666

Lambrinos, D., Möller, R., Labhart, T., Pfeifer, R., and Wehner, R. (2000). A mobile robot employing insect strategies for navigation. *Robot. Auton. Syst.* 30, 39–64. doi: 10.1016/S0921-8890(99)00064-0

Laprie, J. (1995). "Dependable computing and fault tolerance: Concepts and terminology," in *Proceedings of the twenty-fifth international symposium on fault-tolerant computing, 1995, 'Highlights from twenty-five years'*, (Los Alamitos, CA: IEEE Computer Society Press).

Lei, T., Barzilay, R., and Jaakkola, T. (2016). Rationalizing neural predictions. *Arxiv [Preprint]* doi: 10.18653/v1/D16-1011

Lemaignan, S., Fink, J., Mondada, F., and Dillenbourg, P. (2015). "You're doing it wrong! Studying unexpected behaviors in child-robot interaction," in *Social robotics (Lecture notes in computer science)*, eds A. Tapus, E. André, J. Martin, F. Ferland, and M. Ammi (Cham: Springer International Publishing), 390–400. doi: 10.1007/978-3-319-25554-5_39

Levashenko, V., Zaitseva, E., Kvassay, M., and Deserno, T. (2016). "Reliability estimation of healthcare systems using fuzzy decision trees," in *Proceedings of the 2016 federated conference on computer science and information systems (FedCSIS)*, (Gdansk: IEEE), 331–340. doi: 10.15439/2016F150

Li, G., Ramanathan, K., Ning, N., Shi, L., and Wen, C. (2015). Memory dynamics in attractor networks. *Comput. Intell. Neurosci.* 2015:191745. doi: 10.1155/2015/191745

Lipton, Z. (2017). The mythos of model interpretability. *Arxiv [Preprint]*

Liu, S., and Delbruck, T. (2010). Neuromorphic sensory systems. *Curr. Opin. Neurobiol.* 20, 288–295. doi: 10.1016/j.conb.2010.03.007

Liu, X., Ramirez, S., Pang, P., Puryear, C., Govindarajan, A., Deisseroth, K., et al. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* 484, 381–385. doi: 10.1038/nature11028

Lock, R., Burgess, S., and Vaidyanathan, R. (2013). Multi-modal locomotion: From animal to application. *Bioinspir. Biomim.* 9:011001. doi: 10.1088/1748-3182/9/1/011001

Loyola-González, O. (2019). Black-box vs. White-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access* 7, 154096–154113. doi: 10.1109/ACCESS.2019.2949286

Lundberg, S., and Lee, S. (2017). "A unified approach to interpreting model predictions," in *Proceedings of the 31st international conference on neural information processing systems*, (Red Hook, NY: Curran Associates Inc), 4768–4777.

Lundberg, S., Erion, G., Chen, H., DeGrave, A., Prutkin, J., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 56–67. doi: 10.1038/s42256-019-0138-9

Machlev, R., Heistrene, L., Perl, M., Levy, K., Belikov, J., Mannor, S., et al. (2022). Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy AI* 9:100169. doi: 10.1016/j.egyai.2022.100169

Madsen, A., Reddy, S., and Chandar, S. (2022). Post-hoc interpretability for neural NLP: A survey. *ACM Comput. Surv.* 55, 1–42. doi: 10.1145/3546577

McMullen, D., Hotson, G., Katyal, K., Wester, B., Fifer, M., McGee, T., et al. (2014). Demonstration of a semi-autonomous hybrid brain-machine interface using human intracranial EEG, eye tracking, and computer vision to control a robotic upper limb prosthetic. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 784–796. doi: 10.1109/TNSRE.2013.2294685

Merolla, P., Arthur, J., Akopyan, F., Imam, N., Manohar, R., and Modha, D. (2011). "A digital neurosynaptic core using embedded crossbar memory with 45pJ per spike in 45nm," in *Proceedings of the 2011 IEEE custom integrated circuits conference (CICC)*, (Burlington: CICC), 1–4. doi: 10.1109/CICC.2011.6055294

Meta Fundamental AI Research Diplomacy Team (FAIR), Bakhtin, A., Brown, N., Dinan, E., Farina, G., Flaherty, C., et al. (2022). Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science* 378, 1067–1074. doi: 10.1126/science.ade9097

MIT Technology Review (2018). *More efficient machine learning could upend the AI paradigm*. Cambridge MA: MIT Technology Review.

Modha, D., Ananthanarayanan, R., Esser, S., Ndirango, A., Sherbondy, A., and Singh, R. (2011). Cognitive computing. *Commun. ACM* 54, 62–71. doi: 10.1145/1978542.1978559

Muradore, R., and Fiorini, P. (2012). A PLS-based statistical approach for fault detection and isolation of robotic manipulators. *IEEE Trans. Ind. Electron.* 59, 3167–3175. doi: 10.1109/TIE.2011.2167110

Newman, J. (2021). *Explainability won't save AI*. Washington, DC: Brookings.

Nishiwaki, K., Kuffner, J., Kagami, S., Inaba, M., and Inoue, H. (2007). The experimental humanoid robot H7: A research platform for autonomous behaviour. *Philos. Trans. A Math. Phys. Eng. Sci.* 365, 79–107. doi: 10.1098/rsta.2006.1921

Nordin, A., Rymer, W., Biewener, A., Schwartz, A., Chen, D., and Horak, F. (2017). Biomechanics and neural control of movement, 20 years later: What have we learned and what has changed? *J. Neuroeng. Rehabil.* 14:91. doi: 10.1186/s12984-017-0298-y

Nyholm, S. (2022). A new control problem? Humanoid robots, artificial intelligence, and the value of control. *AI Ethics* doi: 10.1007/s43681-022-00231-y

Olds, J. (2019). *Ideas lab for imagining artificial intelligence and augmented cognition in the USAF of 2030*. Fort Belvoir, VA: Defense Technical Information Center.

Oudeyer, P. (2006). *Self-organization in the evolution of speech*. Oxford: Oxford Academic. doi: 10.1093/acprof:oso/9780199289158.001.0001

Pawelczyk, M., Haug, J., Broelemann, K., and Kasneci, G. (2020). "Learning model-agnostic counterfactual explanations for tabular data," in *Proceedings of the web conference 2020*, (New York, NY: Association for Computing Machinery), 3126–3132. doi: 10.1145/3366423.3380087

Pearson, M., Pipe, A., Melhuish, C., Mitchinson, B., and Prescott, T. (2007). Whiskerbot: A robotic active touch system modeled on the rat whisker sensory system. *Adapt. Behav.* 15, 223–240. doi: 10.1177/1059712307082089

Pepito, J., Vasquez, B., and Locsin, R. (2019). Artificial intelligence and autonomous machines: Influences, consequences, and dilemmas in human care. *Health* 11, 932–949. doi: 10.4236/health.2019.117075

Perera, V., Selveraj, S., Rosenthal, S., and Veloso, M. (2016). "Dynamic generation and refinement of robot verbalization," in *Proceedings of the 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, (New York, NY: IEEE), 212–218. doi: 10.1109/ROMAN.2016.7745133

Pescovitz, D. (2021). *Watch these humanoid robots do Parkour (and sometimes fail gloriously!)*. Chicago, IL: Boing Boing.

Pfeifer, R., and Bongard, J. (2007). *How the body shapes the way we think: A new view of intelligence*. Cambridge, MA: Boston Review. doi: 10.7551/mitpress/3585.001.0001

Pintelas, E., Livieris, I., and Pintelas, P. (2020). A grey-box ensemble model exploiting black-box accuracy and white-box intrinsic interpretability. *Algorithms* 13:17. doi: 10.3390/a13010017

Priamikov, A., Fronius, M., Shi, B., and Triesch, J. (2016). OpenEyeSim: A biomechanical model for simulation of closed-loop visual perception. *J. Vis.* 16:25. doi: 10.1167/16.15.25

Rai, A. (2020). Explainable AI: From black box to glass box. *J. Acad. Mark. Sci.* 48, 137–141. doi: 10.1007/s11747-019-00710-5

Raman, V., and Kress-Gazit, H. (2013). Explaining impossible high-level robot behaviors. *IEEE Trans. Robot.* 29, 94–104. doi: 10.1109/TRO.2012.2214558

Rejcek, P. (2022). *Communications FS. Robotic arms connected directly to brain of partially paralyzed man allows him to feed himself*. Available online at: https://blog.frontiersin.org/2022/06/28/robotic-arms-feed-partially-paralyzed-man-bmi/ (accessed January 5, 2022).

Ribeiro, M., Singh, S., and Guestrin, C. (2016). ""Why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, (New York, NY: Association for Computing Machinery), 1135–1144. doi: 10.1145/2939672.2939778

Robnik-Šikonja, M., and Kononenko, I. (2008). Explaining classifications for individual instances. *IEEE Trans. Knowl. Data Eng.* 20, 589–600. doi: 10.1109/TKDE.2007.190734

Romer, P. (1990). Endogenous technological change. *J. Polit. Econ.* 98, S71–S102. doi: 10.1086/261725

Rosenthal, S., Selvaraj, S., and Veloso, M. (2016). "Verbalization: Narration of autonomous robot experience," in *Proceedings of the twenty-fifth international joint conference on artificial intelligence*, (New York, NY: AAAI Press), 862–868.

Rucci, M., Edelman, G., and Wray, J. (1999). Adaptation of orienting behavior: From the barn owl to a robotic system. *IEEE Trans. Robot. Autom.* 15, 96–110. doi: 10.1109/70.744606

Ryan, K. (2021). *Softbank's hyped robot keeps failing at its jobs*. New York, NY: Inc.com.

Sanders, S., and Oberst, J. (2016). Brain-inspired intelligent robotics: The intersection of robotics and neuroscience. *Science* 354:1445. doi: 10.1126/science.2016.354.6318.354_1445b

Shepardson, D. (2017). *Tesla driver in fatal "Autopilot" crash got numerous warnings: U.S. government*. Available online at: https://www.reuters.com/article/us-tesla-crash-idUSKBN19A2XC (accessed January 5, 2022).

Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature* 529, 484–489. doi: 10.1038/nature16961

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of Go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270

Solovyeva, K., Karandashev, I., Zhavoronkov, A., and Dunin-Barkowski, W. (2016). Models of innate neural attractors and their applications for neural information processing. *Front. Syst. Neurosci.* 9:178. doi: 10.3389/fnsys.2015.00178

Spielberg, N., Brown, M., Kapania, N., Kegelman, J., and Gerdes, J. (2019). Neural network vehicle models for high-performance automated driving. *Sci. Robot.* 4:eaaw1975. doi: 10.1126/scirobotics.aaw1975

Sporns, O., and Alexander, W. (2002). Neuromodulation and plasticity in an autonomous robot. *Neural Netw.* 15, 761–774. doi: 10.1016/S0893-6080(02)00062-X

Srinivasa, N., and Cruz-Albrecht, J. (2012). Neuromorphic adaptive plastic scalable electronics: Analog learning systems. *IEEE Pulse* 3, 51–56. doi: 10.1109/MPUL.2011.2175639

Stein, G. (2021). "Generating high-quality explanations for navigation in partially-revealed environments," in *Advances in neural information processing systems*, ed. G. Stein (Red Hook, NY: Curran Associates, Inc), 17493–17506.

Steinbauer, G. (2013). "A survey about faults of robots used in robocup," in *RoboCup 2012: Robot soccer world cup XVI (Lecture notes in computer science)*, eds X. Chen, P. Stone, L. Sucar, and T. van der Zant (Berlin: Springer), 344–355. doi: 10.1007/978-3-642-39250-4_31

Stewart, T., Kleinhans, A., Mundy, A., and Conradt, J. (2016). Serendipitous offline learning in a neuromorphic robot. *Front. Neurorobot.* 10:1. doi: 10.3389/fnbot.2016.00001

Strickland, E. (2019). *Racial bias found in algorithms that determine health care for millions of patients. IEEE spectrum*. Available online at: https://spectrum.ieee.org/racial-bias-found-in-algorithms-that-determine-health-care-for-millions-of-patients (accessed January 5, 2022).

Su, J., Vargas, D., and Kouichi, S. (2019). One pixel attack for fooling deep neural networks. *IEEE Trans. Evol. Computat.* 23, 828–841. doi: 10.1109/TEVC.2019.2890858

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., et al. (2014). Intriguing properties of neural networks. *Arxiv [Preprint]*

The White House OSTP (2022). *Blueprint for an AI bill of rights*. Washington, DC: OSTP.

Tsarouhas, P., and Fourlas, G. (2016). Mission reliability estimation of mobile robot system. *Int. J. Syst. Assur. Eng. Manag.* 7, 220–228. doi: 10.1007/s13198-015-0408-9

Ungerleider, L., and Haxby, J. (1994). 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165. doi: 10.1016/0959-4388(94)90066-3

van Dyck, L., Kwitt, R., Denzler, S., and Gruber, W. (2021). Comparing object recognition in humans and deep convolutional neural networks—An eye tracking study. *Front. Neurosci.* 15:750639. doi: 10.3389/fnins.2021.750639

Vijayakumar, N., Op de Macks, Z., Shirtcliff, E., and Pfeifer, J. (2018). Puberty and the human brain: Insights into adolescent development. *Neurosci. Biobehav. Rev.* 92, 417–436. doi: 10.1016/j.neubiorev.2018.06.004

Welling, S., Refsgaard, H., Brockhoff, P., and Clemmensen, L. (2016). Forest floor visualizations of random forests. *Arxiv [Preprint]*

WIRED (2017). *A freaky humanoid robot that sweats as it does push-ups*. Available online at: https://www.wired.com/story/a-freaky-humanoid-robot-that-sweats-as-it-does-push-ups/ (accessed January 5, 2022).

Wu, M., Hughes, M., Parbhoo, S., Zazzi, M., Roth, V., and Doshi-Velez, F. (2017). Beyond sparsity: Tree regularization of deep models for interpretability. *Arxiv [Preprint]* doi: 10.1609/aaai.v32i1.11501

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., and Zhu, J. (2019). "Explainable AI: A brief survey on history, research areas, approaches and challenges," in *Natural language processing and chinese computing*, eds J. Tang, M. Kan, D. Zhao, S. Li, and H. Zan (Cham: Springer), 563–574. doi: 10.1007/978-3-030-32236-6_51

Yang, G., and Wang, X. (2020). Artificial neural networks for neuroscientists: A primer. *Neuron* 107, 1048–1070. doi: 10.1016/j.neuron.2020.09.005

Yera, R., Alzahrani, A., and Martínez, L. (2022). Exploring post-hoc agnostic models for explainable cooking recipe recommendations. *Knowl. Based Syst.* 251:109216. doi: 10.1016/j.knosys.2022.109216

Zaitseva, E., Levashenko, V., Rabcan, J., and Krsak, E. (2020). Application of the structure function in the evaluation of the human factor in healthcare. *Symmetry* 12:93. doi: 10.3390/sym12010093

Zou, Z., Zhao, R., Wu, Y., Yang, Z., Tian, L., Wu, S., et al. (2020). A hybrid and scalable brain-inspired robotic platform. *Sci. Rep.* 10:18160 doi: 10.1038/s41598-020-73366-9