VOLUME 11

ARTIFICIAL INTELLIGENCE, ROBOTICS & DATA SCIENCE

Topic Coordinators
Sara Degli Esposti & Carles Sierra

CSIC SCIENTIFIC CHALLENGES: TOWARDS 2030

Challenges coordinated by:
Jesús Marco de Lucas & M. Victoria Moreno-Arribas

CHALLENGE 6

ETHICAL, LEGAL, ECONOMIC, AND SOCIAL IMPLICATIONS

Coordinators
P. Noriega (IIIA, CSIC)
T. Ausín (IFS, CSIC)

Participant researchers and centers
H. Mueller (IAE, CSIC)
M. Ellman (IAE, CSIC)
A. R. Cardoso (IAE, CSIC)
F. Aguiar González (IFS, CSIC)
D. López (IFS, CSIC)
M. Toboso (IFS, CSIC)
A. Wagner (IFS, CSIC)
A. García Aracil (INGENIO, CSIC – UPV)
S. Degli Esposti (IPP, CSIC)
L. M. Miller Moya (IPP, CSIC)
L. P. Jiménez (IRI, CSIC- UPV)
J. Pareto (IRI, CSIC- UPV)
C. Torras (IRI, CSIC- UPV)

1. EXECUTIVE SUMMARY

In six decades of history, AI has become a mature and strategic discipline, successfully embedded in mainstream ICT and powering innumerable online applications and platforms. Several official documents stating specific AI policies have been produced by international organisations (like the OCDE), regional bodies (EU), several countries (US, China, Spain, Germany, UK, Sweden, Brazil, Mexico...) as well as major AI-powered firms (Google, Facebook, Amazon). These examples demonstrate ***public interest and awareness of the economic and societal value of AI*** and the urgency of discussing the ethical, legal, economic and social implications of deploying AI systems on a massive scale. There is widespread agreement about the relevancy of addressing *ethical aspects of AI*, an urgency to demonstrate ***AI is used for the common good,*** and the need for better ***training, education and regulation*** to foster responsible research and innovation in AI.

This chapter is organised around four main areas: *ethics, law, economics and society (ELES)*. These areas shape the development of AI research and innovation, which in turn, influence these four areas of human activity. This interplay opens questions and demands new methods, objectives and ways to design future technologies. This chapter identifies the main impacts and salient challenges in each of these four areas.

In *ethics* the widespread consensus of the ethical aspects of AI has raised several AI- related questions for practical ethics about a responsible practice of AI, and therefore the challenge of ***embedding ethics in engineering through education and ethics by design.*** In addition, the concern about the unfettered autonomy of AI systems brought about the insight that it is possible to ***deploy artificial entities that behave ethically*** rising at the same time a fundamental debate about the feasibility and desirability of ***moral agency*** of artificially intelligent entities and systems.

The area of *law* faces the impact of new forms of social behaviour induced by AI that eludes current terminology and regulatory frameworks —like explainability of AI decisions and hidden agency in AI-powered applications— and therefore faces the challenge of ***adapting law to the uses of AI,*** including in this case the *legal personality of artificial systems*. Simultaneously, legal practice is adopting AI technologies for *automating compliance* in a rapidly evolving online environment, hence faces the challenge of ***developing new technologies for governance in augmented reality.***

Economics has received a notable *methodological impact* from AI technologies — like *deep learning* (DL) and agent-based modelling— and also has a profound influence in AI notions and methods. This creates the challenge to ***find synergies to fill the gap (epistemic as well as methodological) between economics and AI research agendas.*** Moreover, the significance of AI in the economy, in the role of stakeholders and its relevance for social well-being in general rises the need to ***identify and explain the repercussions of IA in the economy and in political economy policy.***

Finally, regarding *social science and society*, the emergence of new social phenomena linked to the digitally augmented reality where humans interact with intelligent systems has undeniable effects on how *socio-cognitive rationality* is built and *collective action* is taken. In fact, these hybrid space is creating a and new environment where human kind will co-evolve with AI entities. The challenge is to ***anticipate how that environment may be developed*** in order to foster human flourishing. The formation of public opinion, the evolution of social practices and sharing economy via AI-powered social networks are some of the phenomena induced by those interactions and hence justify the ***need for systematic empirical research of their consequences.***

CSIC is in a privileged position for a successful interdisciplinary approach to the challenges of AI in the ELES area. An interdisciplinary framework is enabled by the confluence and synergies of the expertise provided by the following institutes: IAE (Institute for Economic Analysis), IFS (Institute of Philosophy), IIIA (AI Research Institute), IPP (Institute for Public Policies) and IRI (Robotics and Industrial Informatics Institute). Confluence and synergies would improve with the involvement of other CSIC institutes, in areas like anthropology, archaeology and music.

Some organisational actions may have a very positive impact on synergic interactions among ELES relevant CSIC institutions. In particular the creation of informal laboratories on topics related with the challenges described in this document; the fostering off joint participation in projects; the facilitation of joint contracts and research supervision; and the recruitment of personnel with expertise in psychology/cognitive science and law. One handicap that deserves attention is the lack of a strong group on law and technology in the CSIC structure.

2. INTRODUCTION AND GENERAL DESCRIPTION

There is a growing consensus about the strategic value of AI. The consensus is based on the awareness of the role AI plays in numerous and assorted applications in robotics, self-driving vehicles, e-commerce, health, security and advertisement. The acknowledged strategic value of AI comes as the result of the undeniable success of the embedding of AI in mainstream ICT. However, this success of AI can only be explained by two reasons. First, a maturation process of the discipline, the development of general purpose artefacts and the consolidation of a critical mass of experts, professionals and firms. And second, a timely combination of powerful IT infrastructure and the massive adoption of internet provided the fertile substrate for the mature discipline.

Two remarks as a matter of clarification. First, in this chapter we use the term AI as a liberal. Thus, AI includes all of classical AI, the technologies that come from it —including robotics, *machine learning* (ML) and *big data* (BD)— and, in general, AI entities and systems. Moreover since, as mentioned above, AI is embedded in ICT and is pervasive online, we sometimes call AI what, strictly speaking, would be an AI-enabled or AI-powered system or application.

Not unfrequently the press, and even some of the strategic documents, present an ambivalent perspective of the development of AI: sometimes unfounded optimism of the benefits of AI, sometimes dire predictions of its dangers. A likely explanation of the ambivalence is that AI, as most disruptive scientific and technological innovations, faces the Collingridge dilemma; which postulates that when the discipline is emerging it is difficult to foresee its consequences but by the time one understands it, it may be too late to prevent the unwanted outcomes. We presume one may adopt a cautions but proactive attitude, striving to elucidate where AI is influencing society most, in order to anticipate its consequences and act in accordance. We follow that path by choosing the standpoints of the disciplines that are more salient with respect to the dilemma.

In this chapter we discuss the *social impact* of AI. We focus on four aspects: ethics, law, economics and society. In all four cases we explore how the uses of AI have social effects that pertain to these areas; we also explore how the development of AI poses questions and influences the four areas and how these four areas, in turn, influence the development of AI. Specifically, in the next paragraphs we outline, for each area, what that interplay with AI is about.

2.1. Ethics

Because of its object of study, AI research concerns itself with basic questions about the human mind and the purposeful activity of humans and society. Hence, AI applications impinge sensitive human and social activity by automating, enhancing, or taking over some tasks and roles that rely on traits of human intelligence. Thus, the ethical import of AI is two-fold. First, by posing some *classical ethical questions (introspection, responsibility, autonomy, rights and wrongs, justice)* in a new context where moral reasoning is to be formally and empirically modelled into an artefact. Second, in the ethical implications associated with the deployment and use of AI technologies in *sensitive applications (autonomous weapons, massive face recognition systems, medical diagnosis and prognosis, affective and care robotics)* in the framework of responsible research and innovation.

The acknowledgement of the potential benefits of AI together with its undesirable consequences has permeated into the AI research agenda recognising the opportunity, and the need, of research and innovation that puts the well-being of humans at the centre. This focus has been made explicit through terminology that reflects subtle shadings in the understanding of how AI research and innovation should be pursued: *responsible AI; trust-worthy AI; AI for-good; human-aware or human-centric AI.*

This shift of attention has found its way into policy at regional, national and institutional levels. A good example of how this approach is articulated can be found in the European Approach to Artificial Intelligence, where ethics (together with a regulatory framework) plays a key role as postulated in its Ethics Guidelines for Trustworthy Artificial Intelligence[1].

2.2. Law

The intended development of a human-centric AI acknowledges the need for a *proper regulatory framework*. One that fosters responsible research and innovation in AI, facilitates a productive adoption of AI technology and protects the rights and needs of individuals, enterprises and society in general. Rather than a short-sighted approach to legal groundwork, most policy documents, and the EU one in particular, invite an effort to elucidate the rights that need to be protected, the directives, guidelines, standards and regulations that enable the protection of those rights, while fostering the best potential adoption of AI technologies by society.

In addition to this instrumental role of law in the development and adoption of AI, legal notions and legal philosophy provide valuable notions, intuitions and problems to the foundation and practice of AI. For example, the need to organise interactions among artificial (and natural) entities brings about a need for governance that may be articulated in economic terms as a mechanism or an institution, but in the legal tradition AI draws inspiration form the notions of norms, norm enforcement and compliance in general. On the other hand, AI by its own developments and by the artefacts it produces poses questions to legal foundations and practices that may be as mundane as the protection of the industrial property of a convolutional neural network that is trained for a specif type of medical diagnosis, to the subtle questions of moral agency and the dispute of the pertinence of the "legal persona" of artificial entities.

2.3. Economics

The rapid rise of AI poses a challenge both to the field of economics itself and will have a lasting, profound impact on the organisation of the economy and the political economy of society. It is important that research institutions provide solid support for embracing of the huge economic opportunities that the adoption of AI has to offer. This requires research which is informed by debates within economics and other social sciences and at the same time understands the new tools that the AI revolution has brought. There is otherwise a

1 The policy document and guidelines are available in https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625 and https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

very real danger that AI will be seen as aiding the concentration of political and economic power in the hands of a few and destabilising the political economy.

On a methodological side the cooperation of researchers in AI and economics should be especially fruitful as both are used to work with *quantitative data* and think in statistical models. Economics has already delivered inspiration for AI research and stands to benefit considerably from research which is currently ongoing in AI. However, methods are developed very quickly in both fields and new research will need to build *hybrid models* to facilitate communication between the fields. There are currently huge opportunities for an organisation like CSIC if research across centres and interchanges could be better coordinated.

2.4. Society

The adoption of AI technologies has also a profound *impact in society*. However, as what happens with other technologies, its uses may have unexpected consequences. The distinguishing trait of AI technologies, though, is that they mirror or potentiate cognitive and social behaviour of individuals and groups and by so doing modify society. In fact, one of the most significant outcomes of AI is the invention of an *augmented reality* where natural and artificial entities share a digital, and physical, interaction space. The upshot of such interactions is the emergence of new social phenomena and, eventually, some sort of co-evolution of society and natural autonomous entities.

With such prospects in view, a responsible attitude is to dedicate analytic attention to the ethical and societal impacts of AI, educate professionals and the public on those effects and help develop expertise in AI within the government and public institutions. The content of this chapter responds to this spirit.

3. IMPACT IN BASIC SCIENCE PANORAMA AND POTENTIAL APPLICATIONS

3.1. Ethics

A widespread acknowledgement of the ethical signiftcance of AI. Over the last few years there has been an increasing awareness of the ethical aspects of AI. It is grounded, mainly, on the realisation of the important *ethical implications* of the uses of AI (impact on labour, autonomous vehicles, face-recognition software, AI-powered e-commerce, recommender systems, etc.).

This realisation has four main effects: (i) It motivates the design of policy agendas and industry charters with a strong ethical component (e.g. the EU). (ii) It has prompted several initiatives and specific actions in favour of an ethical AI (manifestos, endowments, creation of institutes and research programmes, funding for research and support of coordinated actions, public pronouncements of major AI empowered stakeholders). (iii) It originates a pressing need to articulate directives and good practices for responsible AI research and Innovation. (iv) It induces a sharp recognition of the need for education in ethics and AI at all levels of society.

The debate on the autonomy of AI systems. From its very beginnings, the point of AI was to model rationality and construct artefacts that exhibit rational behaviour. After almost sixty years, the range of behaviour that has been implemented is quite broad and the performance is rather proficient in many cases. Now, if you "encapsulate" those forms of rationality into a system that can take its own decisions based on its encapsulated rationality, you have a certain type of "autonomy" that is specific to AI. In fact, such autonomy is the fundamental feature that differentiates AI from other disruptive technologies. The issue of autonomy becomes more sophisticated with the development of *autonomous agents*, systems that "act on their own" (see Chapter 2 in this book). These are systems that are autonomous in the previous sense but in addition, they are *situated* in a changing context where they are meant to interact with the environment and with other agents (human or not). The key point is that they decide how to interact based on their own internal decision model, that is their encapsulated rationality. These ideas apply not only to software systems, but also to robots that exhibit behaviour that is situated and reactive and is thus also referred to as autonomous. Common examples of these autonomous artificial entities are web information harvesters used for Google searches, automated online "trader bots", online airline ticketing services and conversational personal aides (like Siri and Alexa). But in the same category one can place autonomous agents whose behaviour is more complex and in a fundamental way, like self-driving cars and killer drones.

As agency of this type becomes more frequent and the interactions become also more consequential, the ascription of *responsibility* and *accountability* of those actions arises (which is not only an ethical issue but a practical and legal one). It is the problem of many hands and many things connected. They include engineering ethics of designers, manufacturers, and maintenance systems; ethical aspects of the artefacts themselves (characterisation of moral agency of artificial entities); and ethical attitudes of the users (hybrid socio-technical systems of humans and intelligent autonomous artefacts). Hence, the debate is served: What does it mean when someone claims that an artificial system is autonomous? In what sense one may claim that an artificial entity has moral agency? What are the features of human agency that one may contend are essential to moral agency? Are they implementable in an autonomous agent? While one side of the debate is to determine to what extent an artificial entity may be autonomous, the other side is to decide whether it is even desirable that truly autonomous artificial intelligent entities be deployed.

The insight that it is possible to develop artiftcially intelligent systems that behave ethically. It has been postulated that one can use AI to control the unwanted side-effects of AI. The underlying insight is that ethical behaviour can be imbued in artificial systems as a form of control. The idea is to imbue values in the architecture of the autonomous entity or imbue values in the governance of the environment where such autonomy is present. The ideal situation is when one may prove that a given degree or extent of alignment of the autonomy with the values is achieved (the so called Value Alignment Problem).

3.2. Law

Addressing the pervasiveness, speed of innovation and the opacity of the uses of AI technologies. Legal practice is the result of a long evolving tradition forged on experience and guided by ideals like justice, common good and dignity of human beings. Its institutions are stable and effective for a vast majority of situations, however AI brings two major disruptive elements: the *speed* at which technology (and its pervasive and assorted use) changes, and its *opacity.* The first component creates loopholes and uncharted spaces for activities and actions that circumvent or are beyond the reach of current normative frameworks and by the time these frameworks adapt to the unwanted effects of the uses of that technology, the activities and actions have changed or use technologies that evade the adapted frameworks. Such is the case, for instance in the use of biometric identification technologies that, by the time they get standardised in a context of use industry has already deployed new deceiving devices. The second problem is that AI-powered activities and practices are often based on design components whose complexity makes it hard to determine compliance with a given regulation or include black boxes that limit the disclosure of functionalities or the assessment of the form or extent to which certain compliance is achieved or evaded. Such is the case, for example, of the explainability requirement in legal appeal and redress of automated decisions (Article 22 of the GDPR).

AI enabled forms of agency and autonomy. It is not always clear who the principal of some actions is or should be. For instance, in the case of self-driving vehicles and automated diagnosis where the "many-hands problem" *dilutes responsibility* and the degree of autonomy afforded to artificially intelligent entities by sophisticated use of AI technologies *dilutes accountability*. Cloning of AI processes or agents may produce compounded effects that are difficult to foresee or contend with–like the stock exchange crises produced by fast-speed trading. Automatic micro contracting questions also the notion of collateral and auditable transactions, in as much as by the time a contract is agreed upon, it has already produced its intended effect and ceases to exist. In the wide picture, there is the debate of legal persona of autonomous artificially intelligent entities.

The automation of governance and compliance. While "legal expert sytems" promised to automate legal reasoning, machine readable regulation promises to make compliance a matter of AI processing and thus eliminate the need for legal experts and interpretation.

Reality is far more complex. In both cases. Underneath the challenging ideal of *compliance by design* there is a strong substrate of legal theory and a substantial collection of formal and AI artefacts that afford some automation in the process of governance and compliance within a legal system: legal ontologies, normative logics, norm specification languages, norm-based automated reasoning, online institutions, normative multiagent systems, to name a few. In addition, there are juice debates on enforcement, punishment and reparation, principle and value-based argumentation and, not surprisingly, a large amount of prototypes and actual applications of these ideas. And there are challenging opportunities, as seen in the next section of this chapter.

3.3. Economics

Methodological and theoretical impact. The field of economics has, in the last five decades, developed a formal language to describe society and analyse the data it generates. The combination of formal models to describe individual and aggregate behaviour has been a big advantage to the field when dealing with data. A rich formal toolkit has developed to tackle issues of causality and integrate theoretical ideas regarding the function of society with empirical work. The Nobel prize 2019 was, for example, awarded to a team of three economists for their experimental approach to alleviating global poverty.

However, *this toolkit is challenged by the wide adoption of AI methods* like ML and causal inference and the increased availability of large quantities of data which are often unstructured, i.e. text, sound and images. The use of *artificial agents in multiagent systems* could provide new simulation-based models which take seriously the heterogeneity of individual preferences and information sets to model the role of governance, values, culture, path analysis and other emerging collective phenomena.

But AI has also adopted concepts, methods and paradigmatic examples from economics into its mainstream toolbox. The most significant examples are *game theory, social choice and voting, together with mechanism design*. These disciplines have not only had a profound impact on the foundations for AI models of rationality and social coordination, but have also provided standards of rigour and test-cases. Moreover, these disciplines have had an enormous impact because of their *systematic use in AI-based applications*. Work on causal inference in econometrics could provide a useful input into the debates around this key issue in ML.

The economic impact of AI. AI affects the economy in many ways. The increased use of AI leads to *productivity gains* in economic activity in general (through the automation of processes and a reshaping workforce capabilities). We will see a wave of *innovation in products and services* that foster labour productivity (like robots, software, data services). But these positive gains will also be linked to shifts in economic power. Automation will change labour market dynamics by destroying some routine jobs and creating new jobs in what is known as job polarisation. Administrative, manual, or repetitive jobs disappear as they are replaced by machines. Second, AI raises the demand for well-educated workers with technical degrees, while reducing the demand for low education workers creating a skill gap. Third, the presence of people and firms with *data science* (DS) capabilities has an impact on national competitiveness. The importance of data for supervised ML means that those companies and governments with access to large amounts of data will advance faster in the development of AI. These considerations have important implications in terms of market regulation, anti-trust and trade.

The Impact of AI in Political Economy: The economic changes resulting from the deployment of AI solutions will increase social and economic inequalities and an imbalance of power between those individuals, countries and regions active in that deployment compared to those lagging behind. Society will need to be able to compensate the losers of the possible economic changes implied by the adoption of AI. Political institutions will need to adapt the rising importance of recommender systems for politics in social media and the further concentration of opportunities, wealth and income. Higher education institutions have a critical role to play in equipping students with the personal and professional competencies (kind of knowledge, skills and attitudes) and capabilities (relevant to active students' engagement in society) that will be in demand in the face of the proliferation of AI and DS. The role of social media for political dialogue needs to be understood particularly well in Spain - a country with a long history of political polarisation (dos Españas).

3.4. Social Science and Society

New metaphors and tools for socio-cognitive rationality and collective action. With the evolution of multiagent systems,there was a gradual recognition of the need to differentiate agents and the social environment where agents interact as two different first class entities (see Chapter 2). This distinction feeds theory and intuitions from the social sciences into AI and, in turn, AI research in these topics has enabled several developments that are relevant for social psychology and sociology. For example:

- Agent models and architectures that account for socio-cognitive rationality (self image, mental models, awareness of the environment, attention).
- Governance devices that articulate interactions in the social space including norms and norm enforcement, but also social norms, organisations, team-work.
- Tools and artefacts. Social coordination conceptual frameworks, methodologies and support tools for design and deployment of *socio-cognitive systems* and *hybrid* online social systems (human and artificial agents).
- Methodologies and tools for network activity analysis, sentiment analysis, text mining, semantic searches...
- Crowd-based technologies: for collective epistemology —like knowledge aggregators, recommender systems, opinion markets—, for problem

decomposition-integration —web surveys, crisis mapping—, for collaborative work —education, constitutional reforms—, and so on.
- Formal treatment of collective action and coordination.
- Agent-based simulation of social phenomena.

Research and development domains in massive on-line AI-powered social interactions. For example:

- The emergence of a research space in social life mediated or supported by social networks and online communities (learning, travelling, gaming, gambling, dating).
- A new character of public opinion: modes of polarisation, information silos; new rhetorics (fake news, false authority, search filtering); "influencer" roles and devices; profiling and micro-targeting, cascading messages through and across social networks and platforms (video and post sharing, re-tweets, whatsapp…)
- Evolution of social practices like entertainment, education, shopping, dating; and evolution of individual skills and habits associated with those web-based AI-powered activities (reading skills, summary representation, memory and information search, friendship networks, antisocial behaviour…)
- The sharing economy: peer-to-peer retail, lodging, transportation, time-banks, lending.

Algorithmic bias, risks of manipulation, social control, hidden agency and other unintended consequences. Despite all the advantages that the large scale application of ML is bringing to digital services, concerns have emerged around the unintended or perilous consequences of automation on online platforms and the opacity of AI systems (Pasquale, 2015). Recommendation systems can reinforce established opinions or reinforce polarised views by presenting to the users information consistent or similar to their preferences and preventing them from being exposed to dissonant information. Search engines, for example, have been criticised for enabling the creation of filter bubbles around users. This type of phenomenon—also known as echo chambers—can be exploited by malicious actors as part of disinformation campaigns. Face recognition systems may be quite useful for security purposes but may rise privacy concerns and abusive control from authoritarian governments. The use of traces of personal web activity (searches, purchases, mobility…) is used to draw a digital profile of individuals that may be used for many different purposes, not always beneficial for the individual lawful.

4. KEY CHALLENGING POINTS IN ETHICS, LAW, ECONOMICS AND SOCIETY

As suggested in the introduction to this chapter, from a social perspective, the ***ultimate challenge*** AI brings about is to solve its version of the Collingridge's dilemma:

How to anticipate the repercussions of AI and act timely in order to achieve the best and avoid the unwanted. In fact, this dilemma may be unfolded in two components: one epistemic, the other ethical, as follows.

The ***epistemic*** component is:

C_1: *To determine what are the repercussions —actual and potential— of the developments of AI.*

The ***ethical*** component (C_2) is:

C_2: *To strive for the best outcomes from AI.*

The first component (C_1) is essentially linked with the different technologies involved in AI. Much of what is discussed in the rest of this book provides indication of what the outcomes of AI research and development may be. However, the repercussions of current and future AI, though, is a matter more of the humanities and the social sciences and addressed in the challenges of this section. The ethical component (C_2) involves two aspects. First, one has to postulate a notion of "good" with respect to which repercussions are valued. As mentioned in the introduction of this chapter, "good AI" has been characterised as trust-worthy, human-centred, responsible. They all share the main concern of caring and protecting human beings. Minimisation of harm, protection of vulnerable individuals and groups and the achievement of sustainable development goals and public well-being, for example, fall within the scope of this concern. But there are several other values. The point is that several values —often conflicting— may be involved in assessing whether a repercussion of AI is good or not, and to what degree.

The second aspect is how to go about in order to achieve good repercussions. In this case two complementary perspectives are involved: First, choosing among potential courses of action —a matter of value-driven behaviour— which can either be learned or imbued. Second, promoting some courses of action and preventing others —a matter of normative systems or policy. This decomposition of the ultimate challenge still hides a fundamental question:

What differentiates AI from other disruptive technologies in a fundamental way? We argue that the answer is ***artificial autonomy.***

It is usually agreed that the ostensible objective of AI is to model human rationality and embed it into artefacts. Sometimes this objective is met by automating specific tasks that typically involve human intelligence in some form, like playing chess, trading stock, face recognition, medical diagnosis. Some other times the goal is not just to automate some activities but to mimic the cognitive processes involved, like natural language understanding, generalised pattern recognition, learning, planning or automated inference. And in both approaches, AI has been quite successful in its six decades of existence. However, some AI researchers discern a more elusive goal: to build artefacts endowed with *general intelligence*. That is, artefacts whose competence is not limited to specific tasks, traits or skills but that they possess an intelligence that, like human intelligence, allows them to act successfully within the ever-changing circumstances of a rich environment. But this capabilities entail that the artefacts must be purposeful, self-conscious, creative and, ultimately, *autonomous*.

It is debatable whether general intelligence, and therefore "true autonomy" will or should ever be feasible in artificial systems. Nevertheless, from the skill gap induced by sophisticated robots, the perverse use of micro-targeting in political campaigns, the need for automated self-censoring of twitter, to the advent of self-driving vehicles and killer drones, it is undeniable that artificially intelligent entities exhibit some form of autonomous behaviour. Hence, a third key component of the ultimate challenge for AI is:

C_3: *To develop means for harnessing artificial autonomy*

The next paragraphs translate these three components into concrete challenges for each of the four areas.

4.1. Challenges in Ethics

1. **Embedding ethics in AI systems** (*The value alignment problem*). How to build autonomous artificial entities that are probably aligned with human values. How to embed social values in the design and development of technologies? What kind of new human rights are related to AI development? How could one ensure a safe human-AI interaction? sub-challenges
   a. ***Development of guidelines and standards.*** Soft approach to imbue values in autonomous artificial entities.

b. ***Value-driven behaviour in artificial entities.*** Conceptual frameworks and tools. (i) *Individual behaviour:* agent architectures; reasoning about values, individual values, context-dependent values. (ii) *Social enforcement*: the role of norms and incentives; social norms and social environment; paradigms and framing. (iii) *Value assessment*.

c. ***Provably reliable ethically-aligned behaviour.*** Formal techniques and devices for characterising, testing and ensuring that the behaviour of artificial entities is aligned with given values.

d. ***Paradigmatic cases.*** (i) *Building socially responsive artefacts*. Evolution and adaptation of value-driven systems (see Chapter 2). (ii) *Policy-assessment systems.* Interplay between individual value-based behaviour and value-alignment of emergent social behaviour. (iii) *Value-sensitive applications* like self-driving vehicles, public surveillance, uses of health records, screening and micro-targeting.

2. **Embedding ethics in engineering.** What issues in AI deserve an ethical approach? How to identify realistic expectations about AI achievements and their effects? Who should be aware of ethical issues in AI? What are the key contents and the means for fostering a realistic view AI? How to develop awareness to the ethical issues of AI? What tools are needed for ethics education for AI? How to expand the ethical literacy of engineers and society? What kind of narratives can we use? The point is, then, to design an agenda for developing critical —ethical, objective, realistic— literacy of AI in society and research.

   There is an acute need for education on the ethical aspects of AI. As stated in the IEEE document (IEEE, 2019)"*...the key is to embed ethics into engineering in a way that does not make ethics a servant, but instead a partner in the process. In addition to an ethics-in-practice approach, providing students and engineers with the tools necessary to build a similar orientation into their inventions further entrenches ethical design practices*". The challenge is in fact that engineers and practitioners come together with social scientists and philosophers to develop contents for such literacy, for example, case studies, interactive virtual reality gaming, narratives (science-fiction), and additional course interventions that are relevant to students, professional and society at large.

   Core outcomes of the agenda should be:

   a. ***Guidelines and teaching syllabus for ethically aware AI R&D.*** It is essential to offer ethical background to students and professionals in technological fields. The ethical dimension of AI must be part of the formative curricula of engineers, computer scientists, and other specialists involved in the design and development of AI and intelligent robots. The standard curricula and delivery should be complemented with non-classical materials like quality science-fiction, serious games and the like.

b. Unless this education is effective, the abuse of ethics expertise and research by Big Tech Companies is likely to effectively stifle government regulations. Thus, it is crucial to preserve independent and public ethical research and education on AI.

c. ***Guidelines and curriculum development to foster critical awareness of AI.*** For IT AI experts, IT professionals and schools that includes ethical awareness as well as critical understanding of AI risks and promises.

d. ***Establishing an observatory of ethics in AI.*** A systematic analysis of the ethical impact of AI and the means to foster ethically desirable outcomes.

3. **Characterising "moral agency" in artiftcial entities.** Could one qualify artificial entities as agents with consciousness, intentional mental states, or indeed moral agents, with ability to engage in moral judgements. Is it feasible to provably imbue values in artificial systems and make those values operational?
The goal is to develop the theories, formalisms and behavioural standards that clarify the extent and shapes to which moral agency can be predicated of an artificial entity.
Such development involves:

a. ***General AI.*** What type of autonomy in artificial entities can be achieved without presuming general AI? What are the constitutive assumptions of general AI? Is general AI an achievable quality for an artificial entity?

b. ***Values.*** In what sense one may claim that ethical theories apply to AI systems. How to make values operational —observable and commensurable, as well as an implementable cognitive construct— in artefacts. Value taxonomies and their relevance. Means to imbue values: norms, incentives, codes of conduct, education and internalisation of norms.

c. ***Artificial autonomy.*** Examining —under the light of artificial agency— moral issues that are meant to be made operational. For example, damage, accountability, responsibility and moral judgement.

4.2. Challenges in Law

1. **Developing "robot laws".** Adapting legal conventions in order to make them applicable to artificial entities as well as to systems and situations that involve them.
   a. ***To develop a strategic taxonomy*** of relevant legal issues, subjects and domains for regulating activity involving artificial entities.
   b. ***To explore the "legal personality" of artificial autonomous entities*** and to identify consequent responsibilities, and thus liabilities, associated with that personality.
   c. ***To re-cast crime, tort and misdemeanour in AI.*** Specifically, the key assessment of features—accountability, responsibility, guilt and liability—and the key handling of processes—infraction, detection, charging, judging, blame assignment, reparation—, need to be performed in AI-assisted socio-technical systems.
   d. ***To address AI-enabled misconduct on digital platforms and its policy implications.*** In order to facilitate *legal appeal and redress in automated decision-making* and *third-party evaluation of AI systems performance and accuracy*, an analysis of instances of misrepresentation, unfair micro-targeting, automated and AI-driven forms of social control and nudging, or the improper delegation of entitlements and responsibilities, should be carried out.
2. **Developing governance technologies for social interactions in augmented reality.**
   a. ***Normative frameworks*** that apply to artificial entities, and to hybrid populations.
   b. ***Normative languages*** that allow back-and-forth rewriting processes from natural language, to legally precise expressions, to machine readable representation. And the machine readable representation of a normative framework is amenable to automated compliance, enforcement and evolution.
   c. ***Institutional frameworks*** for hybrid systems involving artificial autonomous behaviour.
   d. ***Anchoring*** hybrid online social-coordination systems on the actual socio-legaltechnological environment.

4.3. Challenges in Economics

1. **To develop a methodology to fill the gap between economics and AI methodologies for research.** There are three clear areas of development:

a. ***Integrating unstructured data into economic models.*** It will be relevant to explore the development of AI technologies for *feature extraction* from unstructured data available through *text, sound and images*. ML has led to a revolution in this regard and has made it possible to explore entirely new sets of data like TV shows, social media content or satellite imagery. However, it is unclear how to integrate these methods for data extraction into economics models and help address socially important questions.

b. ***Agent-based modelling.*** There are three main areas of potential co-development and some obvious research topics of common interest: (i) *Agents* models of rationality; incentives and motivations; agency and delegation; deductive rationality vs social rationality. (ii) *Multiagent systems* including topics like mechanism design, collective decision making, negotiation and contracting, online institutions, organisational theory, governance. (iii) *Agent-based simulation*. Microbehaviour and collective outcomes, population dynamics, governance, values, culture.

c. ***Experimental economics.*** Design methodologies and best practices for experiments involving a mixed population of humans and artificial agents. Develop protocol standards and specification languages for the definition, implementation and analysis of online experiments with hybrid subject populations.

d. ***Developing a common frameworks for causality and policy interventions.*** Economics has a well-developed econometric toolkit for the identification of causal effects. Currently there is a rival system under development outside economics and, while it is clear that the new causal inference has big formal advantages, there is a danger that methods will be re-invented. On the other hand, multiagent systems and online institutions have approached notions of collective action, governance and policy-making with a perspective that has not permeated into economic theory yet.

e. ***Improving forecasting accuracy.*** Forecasting economic outcomes has a long tradition in economics and is mostly model-based in the field. This approach clashes with the use of ML as the latter provides little scope for integrating policy decisions. However, important organisations like central banks and financial institutions rely on forecast models to make decisions. A new kind of hybrid modelling is needed to accommodate these uses.

2. **To develop a theory for the new AI-enabled commons.** Notions like moral hazard, social choice, behavioural economics play an increasingly important role in AI research. The outcomes of such use should influence economics research but have not been taken up by economists yet.
3. **To identify and explain the economic and political economy repercussions of AI.** In three main directions:
   a. ***Shaping the economic impact of AI and compensate losers.*** Research into the labour market repercussions of the adoption of AI is now common. However, there needs to be a better understanding of the economic impact before technologies are broadly adopted to understand where and which adoption is desirable and who will be the losers from it. This requires broad thinking which combines both research on the labour market impact, allocation of capital across firms and the social and cultural shifts that this brings.
   b. ***Exploring the unintended consequences and externalities of the digital information economy.*** The rise of social media systems has given tools like recommender systems and micro-targeting enormous influence on individual behaviour and welfare. Especially the younger generation is increasingly dependent on social media. The welfare effects of this and the scope for policy interventions can only be understood through multidisciplinary research.
   c. ***Understanding the impact of AI technologies and platforms on the social contract and the functioning of political institutions.*** The concentration of political and economic power in the firms controlling AI systems and the rise of social media as a way to gain information provides a challenge to political institutions. News get filtered through automatised training and the algorithms employed by powerful firms have a direct impact on political debates. Social media and AI-enabled online interactions are transforming the notions of *accountability, representation and responsiveness* of public officials and institutions, on one side and, on the other, the role of individuals and organisations in democratic deliberation, democratic agreement, and public opinion.

4.4. Challenges in Society

1. **Drawing a road-map to develop the use of AI for human flourishing and co-evolution with artiftcial systems.** From teleworking to distant learning, social activity takes place in virtual environments. The

coexistence of humans and AI supported systems and agents require the mutual adaptation of humans and machines toward co-evolution. Artificial entities have capabilities, skills and entitlements that may be designed and deployed with human flouring in mind. To ensure that the large scale adoption of artificial agents and robots create positive effects on the economy and on society, machines and humans need to grow together and learn from each other.

2. **Developing a conceptual framework for the implementation of socio-cognitive agents and socio-cognitive technical systems.** On one side, the challenge requires to address the ***psychological, sociological and computational aspects of artificial rationality social behaviour,*** which involves individual rationality (introspection and the interplay between beliefs, desires and intentions) as well as social rationality (that is, expectations about the behaviour of other agents or about the impact of one's behaviour on the behaviour of others and the possibility of affecting and being affected by other agents or by the social environment where interactions take place). On the other side, the challenge demands the ***construction of online systems including socio-cognitive agents,*** which may be both artificial and human.
3. **Identifying criteria and specifying indicators to evaluate emerging effects of AI on society.** There is a growing need for empirical research to better understand users' demands, values and needs, and also to unveil unexpected consequences of AI-powered human and human-machine activity. To promote a positive collaboration between humans and AI machines and their healthy coexistence, we need to further study how humans interpret and react to AI-assisted systems, and assess the extent to which they may be vulnerable to be deceived or misled by autonomous machines and AI-enabled social interactions.
   The topics to study may be organised around three main lines:
   a. ***Evolution of individual cognitive practices:*** attention, search, recall, collaboration, motivation, locus of control, self-perception, dignity.
   b. ***Evolution of AI-enhanced social practices:*** epistemology, collective agency and coordination, social care —for risk and disadvantaged population, like disabled, elderly, minorities, children, migrants, displaced—; education, health, gaming and gambling; sex and group entertainment; political activism; social control.
   c. ***Unintended consequences of AI:*** biases, discrimination, abusive microtargeting, hidden agency, political manipulation, unfair profiling, misappropriation of personal web traces, and so on.

This challenge requires the combination of small scale qualitative research with large scale quantitative studies. Through both large-scale field experiments exploring the use and adoption of AI-based recommendation systems we can better understand users' requirements and pitfalls. Small-scale laboratory experiment, in contrast, may help shed light on perceptions of vulnerable groups, such as kids, elderly or disabled persons interacting with robots or other autonomous machines.

CHALLENGE 7

LOW-POWER SUSTAINABLE HARDWARE FOR AI

Coordinators

T. Serrano (IMSE-CNM, CSIC – US)

A. Oyanguren (IFIC, CSIC - UV)

Participant researchers and centers

J. Villagrá (CAR, CSIC – UPM)

J. J. Hernández Rey (IFIC, CSIC – UV)

L. Fiorini (IFIC, CSIC – UV)

A. Argyris (IFISC, CSIC – UIB)

C. Mirasso (IFISC, CSIC – UIB)

F. Campabadal (IMB-CNM, CSIC)

J. M. Margarit (IMB-CNM, CSIC)

L. Terés (IMB-CNM, CSIC)

M. Delgado-Restituto (IMSE-CNM, CSIC – US)

B. Linares-Barranco (IMSE-CNM, CSIC – US)

1. EXECUTIVE SUMMARY

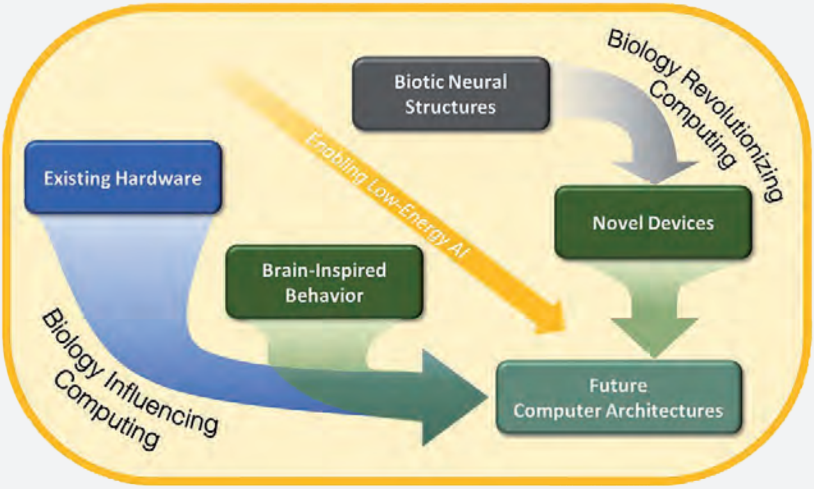
The R&D of high-performance unconventional hardware aims to implement efficient low power high-speed AI systems. This line of work is required to achieve sustainable AI systems and to develop new applications that due to the requirements of high-speed response and power constraints cannot be implemented with the current hardware solutions.

Current AI systems are typically based on running heavy computation algorithms on big powerful remote servers. Both the communication of data to and from the servers and the specific computation consume a considerable amount of energy. To make AI part of our future daily life in a sustainable way, research should be done to move the computation to edge power efficient processors.

Biological brains are a model of natural processors exhibiting an astonishing ability to implement cognitive algorithms with *low energy resources* and *high speed response*.

New bioinspired hardware architectures for AI emulating the computational principles and architectures of biological brains should be developed. The next generation AI hardware should be specifically tailored to run the AI algorithms at the edge in real time and in an energy efficient way.

FIGURE 1.1—Falta



The high-speed sustainable hardware should be based on advanced technology combining existing CMOS technology with new emerging devices as memristors and photonic devices. This beyond CMOS technology will make possible to achieve efficient highly parallel architectures of neural computing units highly interconnected though adaptable synaptic devices with long-term memory. The close integration of memory and computation saves data communication energy and improves the computation speed.

The high-speed sustainable hardware for AI have the potential to make a real breakthrough in a plethora of everyday applications requiring high speed cognitive processing of sensor data and decision taking with an affordable energy budget such as: autonomous cars, autonomous drones, robotics... Other application environments such as wearable medical devices, industrial production, visual inspection of production lines, or surveillance can benefit from this technology. High-speed analysis of *big data* (BD) can also impact and may allow scientific discoveries in basic fields like high-energy physics or astronomy.

CSIC combines experts in designing brain interfacing systems, unconventional neuromorphic hardware architectures, neural processing photonic systems, beyond CMOS technologies, and real-time processing of massive data ***that***

can jointly work in an effort to achieve beyond state-of-the-art AI hardware systems and AI applications specially suited for fast BD analysis.

2. INTRODUCTION AND GENERAL DESCRIPTION

The achievements of AI systems have been outstanding, outperforming humans in some cognition tasks such as the well-known Jeopardy contest where the IBM Watson computer (Ferrucci et al., 2010) defeated its human competitors with almost 40-fold difference in reaction time. However, *human brains still largely outperform the most advanced supercomputers when we compare physical size and energy efficiency.* Coming back to Watson's competition, Watson had 2880 computing cores, occupying the equivalent physical size of 10 refrigerators, and consumed 80 kW, while human brain occupies less than 2 liters and consumes 10-25 W.

In recent years, we have witnessed the growth of AI applications and algorithms. Typically, these algorithms require a heavy computational load and huge data storage which is performed in remote servers on the cloud. Consequently, these systems require an intense data communication between the individual front-end devices running the application and the cloud servers performing the computations and storing BD. These internet communications consume a great amount of energy. As an example, just two google searches emit the same CO_2 amount than boiling a kettle (Warman, 2009). It has been predicted than following the current trend, the internet communications will consume 20% of the world's electricity and will generate 5.5% of the world's carbon emissions by 2025 (Vidal, 2017), and by 2030 the communication technology will consume 51% of total electricity and will contribute up to 23% of the global gas emissions (Andrae and Edler, 2015). Currently, there are already important efforts to move part of the computation from the cloud to embedded AI processors to be used as 'edge' computing devices (Chen et al., 2019).

Furthermore, the big super computation servers are based on traditional Von Neumann computer architectures where a high-speed computation unit sequentially performs operations on data that are read from a separated memory unit. Data have to be transferred to and from the memory at high rates, consuming a high amount of energy in the data transference. This Von Neumann bottleneck, also known as memory bottleneck, is accepted to be one of the main limitations of the performance of conventional processors when running artificial neural networks algorithms (Backus, 1978) . In this sense, the use of dedicated AI low power parallel computing hardware accelerators (Edge TPC, 2019), (Servethehome, 2019), (Extremetech, 2019), (Kim et al., 2019), (Lee et al., 2019), (Cho et al., 2019), (Sano et al., 2020), processing in memory architectures (Li et al., 2016), or commercially available parallel hardware platforms such as *graphic processing units* (GPUs) and *field programmable gate arrays* (FGPAs) have been proposed to implement more efficient AI processors.

The astonishing ability of *biological brains* to solve cognitive problems using low-power low-speed noisy computational neurons has motivated that engineers had looked for inspiration in biology to look into ways of achieving efficient cognitive systems. The brain is very different from conventional computing systems in terms of technology, architecture, as well as signal processing and coding. It is composed of slow components (neurons time constants and synaptic delays are in the order of milliseconds [$10^{-3}s$]) compared to CMOS technology (with operating frequencies in the gigahertz range [that is, $10^{-9}s$]). The brain components suffer from high noise and high variability among them, whereas digital computers operate on high precision digital numbers. However, the on-line plasticity of the devices allows compensating variability and faulty components. The implementation of on-line adaptation in CMOS technology consumes many resources. Alternative beyond CMOS technologies like memristive nanodevices exhibit the capability of on-line adaptation following biologically inspired learning rules while are able to implement very compact memory with no leakage (Linares-Barranco and Serrano-Gotarredona, 2009; Camuñas et al., 2019). Thus, dedicated AI processors combining CMOS technologies with emerging devices (memristors, spintronics, etc.) are being investigated.

Opposed to the high-speed sequential computing paradigm of digital computers, human brain is a highly parallel architecture composed of a high number of parallel processing units or neurons (in the order of 10^{10} 10^{11} neurons) operating at lower speed. Furthermore, the neural processing units are highly interconnected through synaptic weight connections (in the order of 10^{14} synapses) which have an on-line adaptive memory. So that, opposed to what happens in conventional computer architectures, memory and computation are closely interleaved in biological brains.

Not only are there architectural differences between brains and conventional computers, but the signal coding and processing principles are also quite different. In standard computers as well as in conventional AI systems, data are sampled at periodic time intervals and the recognition algorithms are applied sequentially to these sampled data. For example, in conventional AI vision systems, the visual input is a video sequence which is composed of a sequence of static representations of the visual scene or frames which are acquired at periodic time intervals (typically 20-30*ms*). In this conventional AI systems, also known as *artificial neural networks* (ANNs), time is not an explicit variable but it is implicitly contained in the sampled data (Ghosh-Dastidar and Hojjat, 2009). However, in biological brains data are communicated and processed as asynchronous electrical pulses or spikes. In the retina, there are no frames; sensed data are represented as the asynchronous occurrence in time of a flow of spikes that are communicated through the optic nerve along time. Spikes generated in retina travel through the optic nerve and are processed and propagated asynchronously by all the layers in the visual cortex. It has been demonstrated (Thorpe, Fize, and Marlot, 1996) that recognition of a familiar object occurs in the upper layer of the cortex with just the delay of a single spike front traveling through all the cortical layers. A class of AI systems or neuromorphic systems where signals are represented by spikes have emerged and are nowadays also known as *spiking neural networks* (SNNs) or the third generation of neural networks (Maass, 1996; Ghosh-Dastidar and Hojjat, 2009). In this kind of systems, computation is no longer driven by a periodic sampling clock, but computation is driven by the occurrence of spikes resembling the spike-driven computation of biological neural systems (Farabet et al., 2012). It is believed that the parallel computation, the close interaction between computation and memory, an extremely efficient information coding, and on-line adaptation are some of the clues of the high efficiency of biological brain computation.

Different approaches at different levels of abstraction have emerged to design bioinspired low power dedicated hardware to implement more efficiently different aspects of the complex perception, cognition and actuation abilities of the biological neural systems. The different specific low power AI dedicated hardware approaches can be complementary with conventional digital AI systems to optimize the efficiency and abilities of any specific AI system depending on the particular application.

3. IMPACT IN BASIC SCIENCE PANORAMA AND POTENTIAL APPLICATIONS

The development of low-power edge computing dedicated smart sensors and new processors will potentially impact the emergence of many new applications and more efficient solutions of intelligent systems in many application areas. Some of the areas where low power AI hardware can have an important impact and where CSIC can contribute based on the expertise of its research groups are presented in the following.

3.1. High-Performance Energy-Efficient Smart Systems: Autonomous Driving, Drones or Robots

The development of intelligent autonomous systems such as autonomous vehicles or robots demands high-speed low power acquisition and processing of massively parallel input data and high-speed real-time generation of decision and control signals. The high-speed demand is specially challenging in applications such as drones where high-speed reaction is a must. Low-power operation is specially demanding in drones or robots that must operate in remote or inaccessible areas.

This demanding specifications cannot be covered with conventional computing systems. Neuromorphic engineering sensors and processors are being demonstrated able to efficiently acquire massive data and perform high-speed extraction of features using low-power budget. Furthermore, neuromorphic engineering systems implementing spatiotemporal data coding and learning algorithms may lead to compact real-time efficient implementations of recurrent neural networks which are currently very computationally demanding. To obtain energy efficient and high-speed systems, the neuromorphic engineering computation principles can be combined with the use of novel emerging technologies to further improve the system performance. A possible approach is to combine the electronic computation technologies with photonic technologies.

Despite the improvements in the communication networks and the supporting dense computing technology, data acquisition and signal processing are currently firmly dependent on the electronic infrastructure. Even though most of the global data traffic flow is routed nowadays through fiber-optic channels as optical signals, the intelligence that is applied in the optical domain to make any kind of processing is minor. Transferring even a small part of intelligent computations to the optical domain can reduce the unsustainable energy footprint of currently implemented concepts, training methods and AI chips. Moreover, the computational speed has the potential to be increased dramatically, limited only by electrical bottlenecks introduced by optoelectronic conversions. In addition, this may impact a plethora of applications through:

- the increase of the *transfer rates* and *response times* from and within data centers and the incorporation of AI concepts and techniques on-the-fly in the optical domain,
- the *incorporation of dedicated hardware* AI implementations to the next generation of low power photonic integrated circuits for information processing.

In the case of autonomous driving, the biggest challenge to massively deploy autonomous vehicles is finding solutions that address a high degree of uncertainty in highly complex environments. Both artificial perception and decision-making aspire to respond to this challenge and have a critical cross-domain element for the appropriate management of a wide variety of situations: *machine learning* (ML). However, to embed this kind of software approaches, often too computational-intensive, into on-board safety-critical computing devices, which have to face challenging energy and space requirements, motivates the rise of new hardware paradigms. Indeed, ***the advent of powerful multi processors system-on-chip, combining different processing technologies, such as GPUs, FPGA or neuromorphic chips, will enable affordable fault-tolerant hardware architectures.*** As a result, more dependable systems will be available, thus inspiring user trust and engagement, the most difficult hurdle to make this new mobility paradigm a reality.

3.2. Real Time Analysis of Big Data Scientific Systems

Basic science systems where a huge amount of data hit detectors have often to be quickly analyzed for possible storage or for alert. Nuclear and high energy physics, astrophysics and cosmology can greatly benefit from the development of algorithms and ML tools in specific hardware for fast event reconstruction, discrimination and selection. The availability of dedicated hardware able to perform real-time ML on BD may allow scientific discoveries in the cosmology field, discovery of rare and new events in particle physics or even alert to the presence of hazardous events (asteroids). In a similar way, we foresee potential applications in many other basic scientific fields facing the problem of real time BD analysis.

3.3. Monitoring Systems for Health and Security

The development of edge *smart multisensor-fusion* systems can be applied to miniaturized, adaptive instrumentation for real-time industrial/environmental monitoring in order to reduce maintenance and analytical costs, offer long-term stability in front of continuously changing environmental conditions (e.g. temperature, interference effects, sensor drifts), and provide immediate assessment in critical situations like terrorist threats or contamination. This technology will also impact the development of wearable, cognitivebiomonitoring devices for personalized diagnosis and preventive health care, providing continuous multimodal monitoring of individuals in natural environments and therefore allowing to study how dynamically-changing markers can be correlated with their physiological states. Brain mapping and brain-machine interfaces closed-loop neural implants are a promising solution for the treatment of neurological diseases (Bergey, 2015; Zimmermann and Jackson, 2014), as well as for the implementation of sensorimotor brain-machine interfaces. Neural signals captured by the implant are internally processed by means of AI techniques and, based on the outcome, stimulation patterns are triggered either for ameliorating the impact of the disease (e.g., by stopping uncontrolled epileptic seizures) or for restoring lost senses after a neural injury. Signal processing in neural implants should, on the one hand, reduce dimensionality and extract features able to provide clinically relevant information (Yoo et al., 2013) and, on the other, exhibit low power consumptions to not exhaust the presumably short energy resources available. Among the many different operators which have been proposed for the extraction of neural features, those suitable for the quantification of brain functional connectivity are gaining growing interest (Sakkalis, 2011). *Functional connectivity* refers to the statistical evaluation of coupling strengths between brain regions to identify those involved in sensory responses, motor activity or intellectual/emotional processing. Functional connectivity can be assessed on neural signals captured at different spatial resolutions: from singleunit responses acquired by micro-electrode arrays to aggregated signals obtained through electroencephalography or functional magnetic resonance imaging measurements. Besides providing means for brain mapping, clinical studies have demonstrated that functional connectivity also gives relevant information to distinguish between normal and pathological brain states (Bruña, Maestú and Pereda, 2017). Indeed, it has been shown that abnormal synchronization patterns are associated with different neurological disorders, such as, epilepsy (Jiruska et al., 2012), Alzheimer's disease

(Knyazeva et al., 2012), Parkinson's disease (Schnitzler and Gross, 2005) or schizophrenia (Uhlhaas and Singer, 2010).

The referred applications in neurophysiology, disease monitoring and therapy, and brain-machine interfaces suggest the development of dedicated integrated circuits for AI processing based on functional connectivity (Delgado-Restituto, Romaine, and Rodriguez-Vazquez, 2019); ***these applications would be especially valuable in scenarios demanding power efficient solutions*** (e.g., implantable neural prostheses), fast on-site operation to avoid neural activity transfer, and off-line computations by a host computer (e.g., neurostimulators for seizure abortion).

3.4. Emerging Memristive Technologies

As previously mentioned, one shortcoming of conventional computing systems is the separate implementation of the memory and computing units and the time and power consumption required to interchange the data between the memory and the computation part. This makes this conventional systems particularly unsuitable to implement neural networks systems.

In neural networks systems, the synaptic elements are in-memory computing elements, which compute the modulated signal transferred between neural units and at the same time store the knowledge of the system. Furthermore, the synaptic devices have to be massively implemented to achieve a global distributed knowledge of the system. The electronic implementation of synapses is still a challenge for the scientific community.

Memristors are two terminal devices whose resistance changes as a function of the current or voltage applied to their terminals. Memristors devices are probably the best placed to replicate neuronal synapses. This is due to two reasons; the first is the capability of these devices of modulating their electrical resistance thus emulating adaptable connectivity between neurons. The second reason is their small size, few nanometers, which allows many devices to be integrated in a very small area, allowing the number of connections between neurons to be comparable to that of a biological system.

However, as the *memristive technology is still emerging* and several issues concerning reduced variability, low energy operation, device endurance, and current limitation to avoid device damaging have to be addressed to push them into a mature technology that can be densely integrated with CMOS to achieve tightly coupled dense in-memory computing devices. To address this challenge different approaches are being explored, either in terms of device operating conditions or in the optimization and design of new materials combinations. The memristive technology will have impact on:

- *Electronic synapses in neuromorphic systems.* The capability of tuning the memristor resistance together with the potential device scaling of metal-insulator-metal structures in a crossbar configuration are features that are promising for these applications.
- *Non-volatile memories* The scalability, endurance and CMOS compatibility of resistive random access memories based on filamentary resistive switching devices are being considered as potential alternatives to current flash memory technology.
- *Security applications.* The cycle-to-cycle and the device-to-device variabilities are being explored for physical unclonable functions for hardware security applications.

4. KEY CHALLENGING POINTS

4.1. High-Speed Sustainable Hardware for AI

Recently, the availability of large amount of data, the increment of the computing capabilities of CPU processors and the use of high-performance parallel GPU units have motivated a rapid increase in the performance of automatic ML methods. Nowadays, conventional ANNs achieve impressive classification accuracy in tasks such as image recognition.

However, the realization of these classification tasks is very computationally and power demanding even for the most powerful CPUs. For this reason, the use of dedicated AI low power parallel computing hardware accelerators (Edge TPC, 2019), (Servethehome, 2019), (Extremetech, 2019), (Kim et al., 2019), (Lee et al., 2019), (Cho et al., 2019), (Sano et al., 2020), and processing-in-memory architectures (Li et al., 2016), have been proposed.

Furthermore, the interaction with the environment in real time is still very limited even when using hardware accelerators what limits its application in environments where high reaction speed is required like automatic driving or drone control.

On the other hand, SNNs that use a spatio-temporal coding of the signals while still achieving lower recognition rates compared with their sampled or frame-based ANN counterparts have demonstrated their ability to achieve much lower recognition latency and energy than using conventional frame-based vision recognition architectures (Pérez-Carrasco et al., 2013). This high-speed capability and power efficiency performance has motivated emergence of vlsi implementation of complex large scale spiking processing and learning SNN architectures (Furber, 2016).

In parallel, novel neuromorphic sensors for vision, tactile, olfactory and audition have been developed. In particular, novel vision sensors based on sensing the illumination temporal contrast (commonly known today as *dynamic vision sensors* (DVS)) have reached a development maturity that has arisen the interest of the industry in the neuromorphic technology (Posch et al., 2014). These sensors have reached an integration density level (megapixel DVSs are commercially available), low interpixel variability, high-dynamic range (higher than 120 dB), and very low latency (lower than 1ms). Furthermore, the coding of the temporal changes compresses the information reducing the computing requirements of the subsequent processing system.

These developments of *neuromorphic sensors and processors* have risen the interest of the field in the last years as they are promising candidates to implement AI systems that require interacting with the environment in real time such as robotic applications, real time surveillance, autonomous driving... in a more efficient way. Recently, many neuromorphic technology spin-off companies have been launched (such as CelexPixel, Insightness, IniVation, GraiMatterLabs, Prophesee, BrainChip, CortexAI...) and other big companies such as Samsung, Sony, IBM or Intel have begun researching and developing the field.

One of the reasons why SNNs still achieve lower recognition performance than conventional ANNs is the lack of training methods as effective as *backpropagation.* The nondifferential nature of the spikes prevents the backpropagation rule do be directly applicable to SNNs. Other biologically inspired learning rules such as *spike-timing-dependent plasticity* (STDP) can be used for on-line training SNN systems but with lower performance results. However, backpropagation based trained ANNs are not appropriate to incorporate novel learning during operation and suffer from what is known as *catastrophic forgetting.* That is, when an ANN previously trained to recognize a set of objects is re-trained for recognition of a new object set, it gets a serious degradation in recognition of the former training set and one has to retrain it for the complete dataset (although sometimes it is sufficient to retrain the higher layers only). Recently, it has been demonstrated that combining a supervised back-propagation trained ANN with an unsupervised STDP learning SNN classifier, robust learning avoiding catastrophic forgetting is achieved (Muñoz-Martín et al., 2019).

In this context, the development of ***hybrid ANN/SNN architectures exploiting and combining the best capabilities of each approach*** is a key challenging point to achieve high-speed reaction high-accuracy recognition systems with robust learning capabilities. The developed systems should be optimally partitioned depending on the target application to exploit the advantage of high-speed, natural processing of dynamic contents and temporal redundancy suppression of neuromorphic SNN systems with the advantages of more mature learning and computational algorithms of conventional ML systems.

In particular, this challenge requires research on:

- Development of system partition rules for SNN and conventional ANN systems. SNN are more efficient for massive data acquisition and feature extraction processing.
- Study of signal optimal conversion and interaction between SNN and conventional ANN systems.
- Learning algorithms for SNN exploiting spatio-temporal correlations.
- Development of efficient signal temporal coding in SNN to minimize redundancy in signal representations.
- Implementation of efficient hardware friendly on-line learning for SNN.
- Development of combined DNN and SNN learning systems.

One of the possible fields where ***hybrid SNN/ANN systems may have a breakthrough is in real time analysis of massive parallel data.*** The real-time analysis of massive parallel data is a challenging task for conventional sequential processing units even for the more advanced CPUs with many core architectures. This is the case for data coming from highly parallel sensors such as visual data coming from high-resolution vision sensors, data coming from large deployments of IoT sensors, or data for *high energy physics* (HEP) experiments of the *Large Hadron Collider* (LHC) at CERN. Nowadays, the parallelism of GPUs and FPGAs can provide a higher computational power than traditional CPUs and these platforms are currently used to implement ANN processing systems of highly parallel input data. Furthemore, in this kind of massive parallel input data systems there is a large data redundancy in space and time resulting in that only a small fraction corresponds to interesting signals. Trigger systems to decide which data have to be persisted making use of new advanced technologies on computing accelerator platforms such as GPUs and FPGAs are currently used. SNN data processing techniques exploiting efficient spatio-temporal coding of the signal eliminating temporal redundancy in the data and exhibiting low-latency and low-power offer unique capabilities for trigger systems which have to do fast selection decisions on the data of interest. Conventional ML and ANN architectures and methods have been demonstrated to be very efficient in data selection and they can be deployed in the final stages of data processing. Nevertheless, despite the complexity, they are potentially very interesting to be implemented in specific hardware such as FPGAs. This work is starting *to be developed in the HEP field, where ML algorithms are being implemented in FPGAs to improve the performance of the current state of the art reconstruction* (Ortiz Arciniega, Carrió, and Valero, 2019). The experiments of the future high luminosity upgrade of the LHC will use ML algorithms in real time reconstruction accelerator systems such as FPGAs to achieve the expected performance of the experiments, whose trigger systems require the highest possible processing rate and bandwidth and can be implemented using SNN dedicated hardware or SNN architectures programmed in FPGA hardware. In this experiments, the large number of interactions per bunch crossing, happening at a rate of 40 MHz, will make very challenging the distinction of key signatures, which are usually recognized using topological characteristics of the event and kinematic properties of the reconstructed objects. Real-time particle track reconstruction will be crucial to perform fast selection decisions and to record potentially interesting data events for higher level of processing. This poses a major challenge because of ***the large combinatorial and the size of the associated information flow, requiring unprecedented massively parallel pattern-recognition algorithms.*** For this purpose, SNN neurobiology-inspired algorithms, such as Retina (Abba et al., 2016) are of great interest for this field. These SNN bioinspired algorithms can be programmed on FPGAs to offload the most repetitive and logically simple tasks of the trigger systems such as detector decoding, clustering and tracking reconstruction. In particular, GPUs with a many-core architecture for fast particle reconstruction, is also being proposed as a solution for real-time data analysis at some of the decision software stages (Halyo et al., 2013; Aaij et al., 2020)].

The following research should be developed to achieve the real-time analysis and reconstruction of this big physics data:

- Fast object reconstruction and pattern recognition for alignment and calibration of detectors in real time.

- Fast, highly parallelized, seeding algorithms which fulfil timing requirements in trigger systems.
- Fast and efficient data-driven discrimination for data storage.
- Fast and accurate sample classification for selection and online analysis.

Another application field than can benefit from the use of high-speed low power SNN pre-processing algorithms is the sensor fusion. *Multisensor integration* exploits the extended coverage of multiple detectors to increase perceptual confidence in smart systems (Hall et al., 2009; Margarit-Taule et al., 2019), but embedded implementations are yet in their infancy due to the lack of hardware able to infer from the multivariate, nonlinear, time-dependent and noisy signals supplied by modern sensors. Current commercial instrumentation is only able to analyze a small fraction of the markers targeted in everyday applications, generally employs one selective sensor per parameter, and requires periodical calibration in front of environmental changes and sensor non-idealities. Most analyses are still performed in laboratories, resulting in increased costs, hindered logistics, and delayed detection. By using principles of how biological systems promptly combine multisensory information and generate meaningful features in dynamic and uncontrolled real-world conditions, spiking neuromorphic networks are emerging as a powerful, VLSI-amenable computing paradigm to accelerate sensor fusion and enable continuous learning and context awareness under these constraints (O'Connor et al., 2013; Diehl et al., 2015; Li et al., 2019). Such networks can be embedded in smart systems to fuse multivariate data from a set of Si-based microsensors on the edge. The resulting cognitive multisensor system is targeted to be faster, smaller, energy efficient, and highly resilient to noise, nonlinearities, matrix effects, and drifts associated with the sensors. ***To achieve the target edge smart multisensor fusion system, the following research points should be addressed:***

- Accurate sensor modeling to adjust readout/processing circuit design to their static and dynamic characteristics.
- Definition of sensor fusion algorithms tolerant to sensor variability, drift, and crosssensitivities.
- Development of energy-efficient analog front-ends and analog/digital encoding-decoding architectures.
- Implementation of algorithms for incremental and local learning in embedded neuromorphic networks.
- Joint integration of different sensing technologies and VLSI readout/neural processing circuits.

4.2. Development of Next Generation Technology for AI

For efficiently implementing high-speed low-power systems that can perform intelligent tasks in a similar manner as human do and at the same time consume an affordable amount of energy and computing resources, ***hardware parallel architectures alternative to sequential Von Neumann architectures and overcoming the scaling problems of CMOS technology have to be devised.***

The development of low power highly parallel in-memory computing hardware platforms specifically tailored for the class of neural network computation and based on the hybrid combination of conventional CMOS technology and novel computing technologies such as nano memristive devices or photonic computing systems is a key challenge to achieve artificial systems that can emulate the real-time highly parallel computing capabilities of natural cognitive systems.

To develop such a system requires a jointly research effort addressing both architectural and technological questions.

At the ***technology level,*** one of the most promising paradigms to obtain dense neural networks with massive synaptic interconnections among the neural units is the combination of CMOS technologies with new memristive devices. Their existence was theoretically hypothesized by Leon Chua in the 1970s based on circuit theory and it was in 2008 when memristive behaviour was first demonstrated in nanoscale devices by William's group at HP Labs (Camuñas et al., 2019). They are right now the most promising candidates to implement low power in-memory computation and hybrid CMOS-memristive architectures have already been demonstrated using them as binary memory devices. In the neuromorphic engineering field, memristors have been considered as artificial synapses as it has been demonstrated that *when stimulated with electric pulses at their terminals, they exhibit a learning rule of their resistive state which resembles the spike-timing-dependent (STDP) learning rule observed in biological synapses* (Camuñas et al., 2019). Several on-line learning neural networks parallel hardware architectures have been proposed combining CMOS neurons interconnected with massive memristive devices.

However, most of the proposed systems are still based on software simulations whereas only very small demonstrators have been built and mostly based on a binary storage capability of the synaptic devices.

Another challenge that limits the dense integration between arrays of synaptic memristive devices fabricated on top of CMOS neurons and densely interconnected with them is the need of placing a current limiting CMOS transistor in series with every memristor (the so called 1T-1R structures) to avoid damaging of the memristive devices. The implementation of current limitation techniques using the same nanotechnology is series with each device is a must to achieve *high synaptic integration density*.

The are other processing steps in the memristive technology that should be optimized in order to increase the device yield and to improve device performance, such as retention time, resolution of memory states, noise, endurance and variability. In the case of filamentary resistive switching devices, set and reset switching processes are of a stochastic nature leading to a large device-to-device and cycle-to-cycle variability. These processes need to be fully understood for reliable applications. These emerging devices call for new characterization techniques to assess the device electrical behavior. The development of these new techniques will allow to gain a deeper understanding of the physical mechanisms responsible for the device behavior, and to extract parameters that will be necessary in the development of simulation tools for circuit designers.

Photonics is another beyond CMOS technology emerging in the field of neural network hardware implementation. The concepts of optical information processing and computing were introduced several decades ago, with a vision for solving image (pattern) recognition problems (Stroke, 1972; Abu-Mostafa and Psaltis, 1987). Since then, optical processing and computation have spread to different communities, providing revolutionary solutions to their problems, as well as leading to novel applications.

Key factor for this success has been the diversity of available building blocks that introduce computation functionalities in optics and photonics, with low-energy consumption and small footprint. Such attributes have been based lately on semiconductor devices, photonic crystals, fiber-optics, photonic integrated circuits or photonic nano-devices, providing great flexibility to select the most compatible and incorporate it in the various platforms of technological infrastructures (Minzioni et al., 2019). On the one hand, such building blocks can be used in on-demand designs of dedicated photonic neural network configurations that scale up computational capabilities. On the other hand, fiber-optic topologies—supported by a mature technology from the telecom industry—provide a huge flexibility to experiment and test numerous, but relatively simple configurations that addressed diverse classification and time-series prediction problems. For example, the transfer of reservoir computing concepts in all-optical, electro-optical and photonic integration hardware has been a successful paradigm of the last decade, for fast and efficient task-solving computation.

Regarding the development of photonic neural processing systems the following research points should be addressed:

- Introduce versatile concepts of optical computing as an emerging technology for AI hardware accelerated computing with low power consumption.
- Incorporate novel photonic solutions in the optical interfaces of fiber-optic communication infrastructure (data centers / supercomputing centers).
- Implement standalone ML photonic substrates as fundamental ML building blocks (such as photonic perceptrons).
- Design, test and validate photonic integrated circuits with a minimum energy footprint that offer parallel advanced capabilities through AI implementations.
- Design and implement photonic accelerators with AI capabilities for ultra-fast data pre-processing and volume reduction of big data obtained in real-time (eg. from optical sensing and communications networks) or offline.

At an ***architectural level,*** research should be conducted to develop new system architectures tailored for the specifities of the neural network systems. Based on those specifities new in-memory computing basic building blocks, routing and communication blocks and parallelization strategies should be designed. The architecture should look for a ***compromise between versatility and efficiency.*** A versatile hardware is desired to be able to implement different signal codings, different learning algorithms, different neuron models and different neural architectures. However, adding versatility results in a more complex architecture less area and power efficient. At the same time, the architecture should divide the system functionalities among the different technological platforms (CMOS, memristive and photonics) to optimize the system performance.

VOLUME 11 **REFERENCES**

Aaij, R., Albrecht, J., Belous, M. et al. (2020). Allen: A High-Level Trigger on GPUs for LHCb. *Computing and Software for Big Science*, *4*(1), 7. DOI:10.1007/s41781-020-00039-7

Abba, A., Bedeschi, F., Citterio, M. et al. (2016). The artificial retina for track reconstruction at the LHC crossing rate. *Nuclear and Particle Physics Proceedings*, *273–275*, 2488–2490. DOI:10.1016/j.nuclphysbps.2015.09.434

Abu-Mostafa, Y. S., & Psaltis, D. (1987). Optical Neural Computers. *Scientific American*, *256*(3), 88–95. DOI:10.1038/scientificamerican0387-88

ACT-R Research Group. (2013). ACT-R Publications & Models. http://act-r.psy.cmu.edu/publication/, 2002-2013.

Adams, S. S., Arel, I., Bach, J. et al. (2012). Mapping the landscape of human-level artificial general intelligence. *AI Magazine*, *33*(1), 25–41. DOI:10.1609/aimag.v33i1.2322

Anderson, J. R. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*(4), 439–462. DOI:10.1207/s15327051hci1204_5

Anderson, M. L. (2003). Embodied Cognition: A field guide. *Artificial Intelligence*, 149 (1), 91–130. DOI:10.1016/S0004-3702(03)00054-7

Andrae, A., & Edler, T. (2015). On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges*, *6*(1), 117–157. DOI:10.3390/challe6010117

Andrejczuk, E., Bistaffa, F., Blum, C. et al. (2019). Synergistic team composition: A computational approach to foster diversity in teams. *Knowledge-Based Systems*, *182*, 104799. DOI:10.1016/j.knosys.2019.06.007

Andriella, A., Alenyà, G., Hernández-Farigola, J., & Torras, C. (2018). Deciding the different robot roles for patient cognitive training. *International Journal of Human Computer Studies*, *117*, 20–29. DOI:10.1016/j.ijhcs.2018.03.004

Andriella, A., Torras, C., & Alenyà, G. (2020). Cognitive System Framework for Brain-Training Exercise Based on Human-Robot Interaction. *Cognitive Computation*, *12*(4), 793–810. DOI:10.1007/s12559-019-09696-2

Antonucci, A., Cholvy, L., & Papini, O. (2017). Symbolic and Quantitative Approaches to Reasoning with Uncertainty. *Springer International Publishing*. DOI:10.1007/978-3-319-61581-3

Arciniega, J. L. O., Carrió, F., & Valero, A. (2019). FPGA implementation of a deep learning algorithm for real-time signal reconstruction in particle detectors under high pile-up conditions. *Journal of Instrumentation*, *14*(9). DOI:10.1088/1748-0221/14/09/P09002

Arroyo, D., Diaz, J., & Gayoso, V. (2015). On the difficult tradeoff between security and privacy: Challenges for the management of digital identities. *Advances in Intelligent Systems and Computing*, *369*, 455–462. DOI:10.1007/978-3-319-19713-5_39

Artila, V., & Lloyd, D. (2014). Subjective time: The philosophy, psychology, and neuroscience of temporality.

Artunedo, A., Villagra, J., & Godoy, J. (2019). Real-Time Motion Planning Approach for Automated Driving in Urban Environments. *IEEE Access*, *7*, 180039–180053. DOI:10.1109/ACCESS.2019.2959432

Backus, J. (1978). Can Programming Be Liberated from the von Neumann Style? A Functional Style and Its Algebra of Programs. *Communications of the ACM*, *21*(8), 613–641. DOI:10.1145/359576.359579

Bair, J., Bellovin, S., Manley, A. et al. (2017). That Was Close! Reward Reporting of Cybersecurity "Near Misses." *Colorado Technology Law, 16, 327*

Ball, K., Degli Esposti, S., Dibb, S. et al. (2019). Institutional trustworthiness and national security governance: Evidence from six European countries. *Governance*, *32*(1), 103–121. DOI:10.1111/gove.12353

Ball, N. M., & Brunner, R. J. (2009). Data Mining and Machine Learning in Astronomy. *International Journal of Modern Physics D*, *19*(7), 1049–1106. DOI:10.1142/S0218271810017160

Banks, D. L., Rios, J., & Insua, D. R. (2015). Adversarial risk analysis. *Adversarial Risk Analysis*. DOI:10.1201/b18653

Bansal, S., & Tomlin, C. J. (2019). Control and Safety of Autonomous Vehicles with Learning-Enabled Components. *Springer, Cham*, 57-75. DOI:10.1007/978-3-319-97301-2_4

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. DOI:10.1146/annurev.psych.59.103006.093639

Batty, M. (2018). *Inventing Future Cities.* MIT Press. https://books.google.es/books/about/Inventing_Future_Cities.html?id=zs0StQEACAAJ&redir_esc=y

Bengio, Y., Lodi, A., & Prouvost, A. (2018). Machine Learning for Combinatorial Optimization: a Methodological Tour d'Horizon. *European Journal of Operational Research, 290*(2), 405–421. http://arxiv.org/abs/1811.06128

Bent, R., & Van Hentenryck, P. (2018). Online Stochastic Combinatorial Optimization. *The MIT Press*. DOI:10.7551/mitpress/5140.001.0001

Bergen, K. J., Johnson, P. A., De Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science* 363 (6433). DOI:10.1126/science.aau0323

Berger, J. O., Insua, D. R., & Ruggeri, F. (2000). Robust Bayesian Robustness. *Lecture Notes in Statistics,* 152. DOI:10.1007/978-1-4612-1306-2_1

Bergey, G. K., Morrell, M. J., Mizrahi, E. M. et al. (2015). Long-term treatment with responsive brain stimulation in adults with refractory partial seizures. *Neurology, 84*(8), 810–817. DOI:10.1212/WNL.0000000000001280

Billari, F., & Zagheni, E. (2017). Big Data and Population Processes: A Revolution? DOI:10.31235/osf.io/f9vzp

Bing, Z., Meschede, C., Röhrbein, F. et al. (2019). A survey of robotics control based on learning-inspired spiking neural networks. *Frontiers in Neurorobotics, 12,* 35. DOI:10.3389/fnbot.2018.00035

Bischoff, R., & Guhl, T. (2010). The strategic research agenda for robotics in Europe. *IEEE Robotics and Automation Magazine, 17*(1), 15–16. DOI:10.1109/MRA.2010.935802

Bisio, F., Saeli, S., Lombardo, P. et al. (2017). Real-time behavioral DGA detection through machine learning. *Proceedings - International Carnahan Conference on Security Technology,* 1–6. DOI:10.1109/CCST.2017.8167790

Bistaffa, F., Blum, C., Cerquides, J. et al. (2019). A Computational Approach to Quantify the Benefits of Ridesharing for Policy Makers and Travellers. *IEEE Transactions on Intelligent Transportation Systems, 1-12.*

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2016). Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association, 112*(518), 859–877. DOI:10.1080/01621459.2017.1285773

Blum, C., & Roli, A. (2003). Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. *ACM Computing Surveys,* 35(3), 268–308). DOI:10.1145/937503.937505

Blum, C., Pinacho, P., López-Ibáñez, M., & Lozano, J. A. (2016). Construct, Merge, Solve & Adapt: A New General Algorithm For Combinatorial Optimization. *Computers & Operations Research,* 68, 75–88. DOI:10.1016/j.cor.2015.10.014

Bost, R., Ada Popa, R., Tu, S., & Goldwasser, S. (2015). Machine Learning Classification over Encrypted Data.

Bonet, B., Palacios, H., & Geffner, H. (2009). Automatic Derivation of Memoryless Policies and Finite-State Controllers Using Classical Planners. *ICAPS*.

Brachman, R. J., & Levesque, H. J. (2004). Knowledge Representation and Reasoning. *Elsevier Inc*. DOI:10.1016/B978-1-55860-932-7.X5083-3

Brook, A., & DeVidi, R. C. (2001). Self-reference and Self-awareness. *Benjamins.*

Bruña, R., Maestú, F., & Pereda, E. (2018). Phase locking value revisited: Teaching new tricks to an old dog. *Journal of Neural Engineering, 15*(5). DOI:10.1088/1741-2552/aacfe4

Buchanan, B. G. (2005). A (Very) Brief History of Artificial Intelligence. *AI Magazine,* 26 (4), 53. DOI: 10.1609/aimag.v26i4.1848

Buchanan, B., & Sutherland, G. (1969). Heuristic Dendral: A program for generating explanatory hypotheses in organic chemistry. *Machine Intelligence, 4,* 209–254.

Buchanan, B. G., & Shortliffe, E. H. (1984). Rule-based expert system. *Reading: Addison-Wesley*

Byrd, R. H., Hansen, S. L., Nocedal, J., & Singer, Y. (2016). A stochastic quasi-Newton method for large-scale optimization. *SIAM Journal on Optimization, 26*(2), 1008–1031. DOI:10.1137/140954362

Caliskan-Islam, A., Greenstadt, R., Harang, R. et al. (2015). De-anonymizing Programmers via Code Stylometry. *24th {USENIX} Security Symposium,* 255–270.

Campbell, M., Hoane, A. J., & Hsu, F. H. (2002). Deep Blue. *Artificial Intelligence, 134*(1–2), 57–83. DOI:10.1016/S0004-3702(01)00129-1

Camuñas-Mesa, L. A., Linares-Barranco, B., & Serrano-Gotarredona, T. (2019). Neuromorphic spiking neural networks and their memristor-CMOS hardware implementations. *Materials*, 12 (17), 2745. DOI:10.3390/ma12172745

Canal, G., Alenyà, G., Torras, C. et al. (2019). Adapting robot task planning to user preferences: An assistive shoe dressing example. *Autonomous Robots*, 43 (6), 1343–1356.

Castelló, X., Eguíluz, V. M., & San Miguel, M. (2006). Ordering dynamics with two non-excluding options: Bilingualism in language competition. *New Journal of Physics*, *8*(12), 308. DOI:10.1088/1367-2630/8/12/308

Cegielski, W. H., & Rogers, J. D. (2016). Rethinking the role of Agent-Based Modeling in archaeology. *Journal of Anthropological Archaeology*, *41*, 283–298. DOI:10.1016/j.jaa.2016.01.009

Centola, D., González-Avella, J. C., Eguíluz, V. M., & San Miguel, M. (2007). Homophily, Cultural Drift, and the Co-Evolution of Cultural Groups. *Journal of Conflict Resolution*, *51*(6), 905–929. DOI:10.1177/0022002707307632

Chatzilygeroudis, K., Vassiliades, V., Stulp, F. et al. (2020). A survey on policy search algorithms for learning robot controllers in a handful of trials. *IEEE Transactions on Robotics*, *36*(2). DOI:10.1109/TRO.2019.2958211ï

Chi, P., Li, S., Xu, C. et al. (2016). PRIME: A Novel Processing-in-Memory Architecture for Neural Network Computation in ReRAM-Based Main Memory. *Proceedings - 2016 43rd International Symposium on Computer Architecture, ISCA 2016*, 27–39. DOI:10.1109/ISCA.2016.13

Christoph Posch, B., Member IEEE, S., Serrano-Gotarredona, T. et al. (2014). Retinomorphic Event-Based Vision Sensors: Bioinspired Cameras With Spiking Output. *Proceedings of the IEEE*, 102, 1470–1484 DOI:10.1109/JPROC.2014.2346153

Chuyakin, A., Schmidt, K., Phillips, C. (2012). Logging and log management: the authoritative guide to understanding the concepts surrounding logging and log management. *Newnes.*

Cintula, P, Fermüller, C., & Noguera, C. (2015). Handbook of mathematical fuzzy logic-volume 3. *College Publications.*

Cintula, P., Hájek, P., & Noguera, C. (2011). Handbook of Mathematical Fuzzy Logic - volume 1. *College Publications*.

Cintula, P., Hájek, P., & Noguera, C. (2011). Studies in Logic, Mathematical Logic and Foundations. *College Publications.*

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9), 345–351. DOI:10.1016/S1364-6613(99)01361-3

Collobert, R., Weston, J., Com, J. et al. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, 12, 2493–2537.

Colomé, A., & Torras, C. (2020). Reinforcement Learning of Bimanual Robot Skills. *Springer International Publishing*, 134. DOI:10.1007/978-3-030-26326-3

Comiter, M. (2019). Attacking Artificial Intelligence AI's Security Vulnerability and What Policymakers Can Do About It. *Belfer Center Paper.*

Conte, R., Gilbert, N., Bonelli, G. et al. (2012). Manifesto of computational social science. *European Physical Journal: Special Topics*, *214*(1), 325–346. DOI:10.1140/epjst/e2012-01697-8

Confalonieri, R., Pease, A., Schorlemmer, M. et al. (2018). Concept Invention: Foundations, Implementation, Social Aspects and Applications. *Springer, series in Computational Synthesis and Creative Systems.* DOI:10.1007/978-3-319-65602-1

Coral (2019). Edge TPC. https://coral.withgoogle.com/docs/edgetpu/faq/

Cortes, J., & Egerstedt, M. (2017). Coordinated Control of Multi-Robot Systems: A Survey. *SICE Journal of Control, Measurement, and System Integration*, *10*(6), 495–503. DOI:10.9746/jcmsi.10.495

Cortet, M., Rijks, T., & Nijland, S. (2016). PSD2 : the digital transformation accelerator for banks. *Journal of Payments Strategy & Systems*, *10*(1).

Cowan, R. S. (1987). The Whale and the Reactor: A Search for Limits in an Age of High Technology. Langdon Winner. *Isis*, *78*(2), 296–297. DOI:10.1086/354443

Curry, S., Kirda, E., Schwarts, E. et al. (2013). Big Data Fuels Intelligence-Driven Security, *RSA Security Brief.*

Czaplicka, A., Toral, R., & San Miguel, M. (2016). Competition of simple and complex adoption on interdependent networks. *Physical Review E, 94*(6). DOI:10.1103/PhysRevE.94.062301

Dabney, W., Kurth-Nelson, Z., Uchida, N. et al. (2020). A distributional code for value in dopamine-based reinforcement learning. *Nature, 577*(7792), 671–675. DOI:10.1038/s41586-019-1924-6

Dalvi, N., Domingos, P., Mausam, S. et al. (2004). Adversarial classification. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 99–108. DOI:10.1145/1014052.1014066

Darwiche, A. (2018). Human-level intelligence or animal-like abilities? *Communications of the ACM, 61*(10), 56–67. DOI:10.1145/3271625

Dastani, M., El Fallah Segrouchni, A., Leite, J., & Torroni, P. (2010). Languages, Methodologies, and Development Tools for Multi-Agent Systems. *Springer Berlin Heidelberg*, 6039. DOI:10.1007/978-3-642-13338-1

De la Torre-Abaitua, G., Lago-Fernández, L. F., & Arroyo, D. (2021). On the application of compression-based metrics to identifying anomalous behaviour in web traffic. *Logic Journal of the IGPL, 28*(4), 546–557. DOI:10.1093/JIGPAL/JZZ062

De Santos, P. G., Garcia, E., Ponticelli, R., & Armada, M. (2009). Minimizing energy consumption in hexapod robots. *Advanced Robotics, 23*(6), 681–704. DOI:10.1163/156855309X431677

Deb, K., & Myburgh, C. (2016). Breaking the Billion-Variable Barrier in real-world optimization using a customized evolutionary algorithm. *Proceedings of the 2016 Genetic and Evolutionary Computation Conference,* 653–660. DOI:10.1145/2908812.2908952

Delgado-Restituto, M., Romaine, J. B., & Rodríguez-Vázquez, Á. (2019). Phase Synchronization Operator for On-Chip Brain Functional Connectivity Computation. *IEEE Transactions on Biomedical Circuits and Systems, 13*(5), 957–970. DOI:10.1109/TBCAS.2019.2931799

Dennett, D. C. (2018). Facing up to the hard question of consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 373(1755). DOI:10.1098/rstb.2017.0342

Diaz, J., Choi, S. G., Arroyo, D. et al. (2019). A methodology for retrofitting privacy and its application to e-shopping transactions. *Advances in Cyber Security: Principles, Techniques, and Applications*. 143–183. DOI:10.1007/978-981-13-1483-4_7

Diaz, J., Choi, S. G., Arroyo, D. et al. (2018). Privacy in e-Shopping Transactions: Exploring and Addressing the Trade-Offs. *International Symposium on Cyber Security Cryptography and Machine Learning,* 206–226.

Diehl, P. U., Neil, D., Binas, J. et al. (2015). Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. *Proceedings of the International Joint Conference on Neural Networks*. DOI:10.1109/IJCNN.2015.7280696

Dobchev, D., Pillai, G., & Karelson, M. (2014). In Silico Machine Learning Methods in Drug Development. *Current Topics in Medicinal Chemistry, 14*(16), 1913–1922. DOI:10.2174/1568026614666140929124203

Dwork, C. (2013). Differential privacy: A cryptographic approach to private data analysis. *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, 296–322. DOI:10.1017/CBO9781107590205.018

D'Inverno, M., Luck, M., Noriega, P. et al. (2012). Communicating open systems. *Artificial Intelligence, 186,* 38–94. DOI:10.1016/j.artint.2012.03.004

Eguíluz, V. M., Zimmermann, M. G., Cela-Conde, C. J., & San Miguel, M. (2005). Cooperation and the emergence of role differentiation in the dynamics of social networks. *American Journal of Sociology, 110*(4), 977–1008. DOI:10.1086/428716

Ekin, T., Naveiro, R., Torres-Barrán, A., & Ríos-Insua, D. (2019). Augmented Probability Simulation Methods for Non-cooperative Games. *ArXiv*

Esteva, M., Rodríguez-Aguilar, J.-A., Sierra, C. et al. (2001). On the Formal Specification of Electronic Institutions. *Springer Berlin Heidelberg*. 126-147. DOI:10.1007/3-540-44682-6_8

EU (2014). Electronic identification, authentication and trust services. https://eurlex.europa.eu/legal-content/en/txt/pdf/?uri=celex:32014r0910&from=en.

ExtremeTech. (2019). Intel Details Its Nervana Inference and Training AI Cards. https://www.extremetech.com/computing/ 296990 -intel-nervana-nnp-i-nnp-t-a-training-inference.

Farabet, C., Paz, R., Pérez-Carrasco, J. et al. (2012). Comparison between frame-constrained fix-pixel-value and frame-free spiking-dynamic-pixel convNets for visual processing. *Frontiers in Neuroscience, APR*. DOI:10.3389/fnins.2012.00032

Faratin, P., Sierra, C., & Jennings, N. R. (1998). Negotiation decision functions for autonomous agents. *Robotics and Autonomous Systems, 24*(3–4), 159–182. DOI:10.1016/S0921-8890(98)00029-3

Farinelli, A., Iocchi, L., & Nardi, D. (2004). Multi-Robot Systems: A classification focused on coordination. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(5),2015–2028.

Feigenbaum, E., Feldman, J. (1963). Computers and Thought. *McGraw Hill.*

Fernández-Gracia, J., Suchecki, K., Ramasco, J. J. et al. (2014). Is the Voter Model a Model for Voters? *Physical Review Letters, 112*(15). DOI:10.1103/PhysRevLett.112.158701

Fernández, R., Montes, H., Salinas, C. et al. (2013). Combination of RGB and multispectral imagery for discrimination of Cabernet Sauvignon grapevine elements. *Sensor, 13*(6), 7838–7859. DOI:10.3390/s130607838

Fernández, R., Montes, H., Surdilovic, J. et al. (2018). Automatic detection of field-grown cucumbers for robotic harvesting. *IEEE Access, 6,* 35512–35526. DOI:10.1109/ACCESS.2018.2851376

Ferrucci, D., Brown, E., Chu-Carroll, J. et al. (2011). Building watson: An overview of the deepQA project. *AI Magazine, 31*(3), 59–79. DOI:10.1609/aimag.v31i3.2303

Finck, M. (2018). Blockchains and Data Protection in the European Union. *European Data Protection Law Review, 4*(1), 17–35. DOI:10.21552/edpl/2018/1/6

Flaminio, T., Godo, L., & Hosni, H. (2017). On boolean algebras of conditionals and their logical counterpart. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10369 LNAI,* 246–256. DOI:10.1007/978-3-319-61581-3_23

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review, 1*(1). DOI:10.1162/99608f92.8cd550d1

Fodor, J. A. (1975). *The Language of Thought.* Harvard University Press.

French, S., & Rios, D. (2007). Statistical Decision Theory, *Wiley.* DOI:10.1007/978-0-387-73194-0_3

Furber, S. (2016). Large-scale neuromorphic computing systems. *Journal of Neural Engineering*, 13(5). DOI:10.1088/1741-2560/13/5/051001

Futami, F., Sato, I., & Sugiyama, M. (2018). Variational Inference based on Robust Divergences. *International Conference on Artificial Intelligence and Statistics,* 813-822

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement. *Neuron,* 94(5), 978–984. DOI:10.1016/j.neuron.2017.05.025

Gallego, V., Naveiro, R., & Insua, D. R. (2019). Reinforcement learning under threats. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 9939–9940. DOI:10.1609/aaai.v33i01.33019939

Garcia-Camacho, I., Alenya, G., Kragic, D. et al. (2020). Benchmarking Bimanual Cloth Manipulation. *IEEE Robotics and Automation Letters, 5*(2), 1111–1118. DOI:10.1109/LRA.2020.2965891

García-Pérez, L., García-Alegre, M. C., Ribeiro, A., & Guinea, D. (2008). An agent of behaviour architecture for unmanned control of a farming vehicle. *Computers and Electronics in Agriculture, 60*(1), 39–48. DOI:10.1016/j.compag.2007.06.004

Geffner, H., & Bonet, B. (2013). A concise introduction to models and methods for automated planning. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 22,* 1–141. DOI:10.2200/S00513ED1V01Y201306AIM022

Ghallab, M., Nau, D., & Traverso, P. (2016). Automated Planning and Acting. *Cambridge University Press*. DOI:10.1017/CBO9781139583923

Ghosh-Dastidar, S., & Adeli, H. (2009). Third generation neural networks: Spiking neural networks. *Advances in Intelligent and Soft Computing, 61,* 167–178. DOI:10.1007/978-3-642-03156-4_17

Godoy, J., Pérez, J., Onieva, E. et al. (2015). A driverless vehicle demonstration on motorways and in urban environments. *Transport, 30*(3), 253–263. DOI:10.3846/16484142.2014.1003406

Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies, 10*(1), 1–311. DOI:10.2200/S00762ED1V01Y201703HLT037

Gómez Esteban, P., Liu, S., Ríos Insua, D., & González-Ortega, J. (2020). Competition and cooperation in a community of autonomous agents. *Autonomous Robots, 44*(3–4), 533–546. DOI:10.1007/s10514-019-09867-y

González-Ruibal, A., Chilton, E., Kristiansen, K., & Niklasson, E. (2014). Towards a new paradigm? The Third Science Revolution and its Possible Consequences in Archaeology. *Current Swedish Archaeology, 22*(4), 11-71.

Goodfellow, I., Bnegio, Y., & Courville, A. (2018). Deep learning. *Genetic Programming and Evolvable Machines, 19*(1–2), 305–307. DOI:10.1007/s10710-017-9314-z

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M. et al. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems,* 27.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *3rd International Conference on Learning Representations-Conference Track Proceedings*.

Guggenmos, M., Schmack, K., Veer, I. M. et al. (2020). A multimodal neuroimaging classifier for alcohol dependence. *Scientific Reports, 10*(1). DOI:10.1038/s41598-019-56923-9

Guiochet, J., Machin, M., & Waeselynck, H. (2017). Safety-critical advanced robots: A survey. *Robotics and Autonomous Systems, 94,* 43–52. DOI:10.1016/j.robot.2017.04.004

Hall D., Llinas J., Liggins II, M., Poularikas, A. (2017). Handbook of Multisensor Data Fusion. *CRC Press.* DOI:10.1201/9781420053098

Halyo, V., Hunt, A., Jindal, P. et al. (2013). GPU Enhancement of the Trigger to Extend Physics Reach at the LHC. *Journal of Instrumentation, 8*(10). DOI:10.1088/1748-0221/8/10/P10005

Hamrick, J. B., Ballard, A. J., Pascanu, R. et al. (2017). Metacontrol for Adaptive Imagination-Based Optimization. *5th International Conference on Learning Representations*

Hanea, A. M., Nane, G. F., Bedford, T., & French, S. (2021). *Expert Judgement in Risk and Decision Analysis*. Springer International Publishing. DOI:10.1007/978-3-030-46474-5

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics, 4*(2), 100–107. DOI:10.1109/TSSC.1968.300136

Havens, J., Bielby, J., & Alvarez, M. A. P. (2019). Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems. *IEEE* https://www.academia.edu/38746211/ETHICALLY_ALIGNED_DESIGN_A_Vision_for_Prioritizing_Human_Wellbeing_with_Artificial_Intelligence_and_Autonomous_Systems

Hettwer, B., Gehrer, S., & Güneysu, T. (2020). Applications of machine learning techniques in side-channel attacks: a survey. *Journal of Cryptographic Engineering, 10*(2), 135–162. DOI:10.1007/s13389-019-00212-8

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. DOI:10.1162/neco.1997.9.8.1735

Hodges, A. (2014). Alan Turing: The Enigma. *Penguin Books Australia.*

Hooker, J. N. (2019). Logic-Based Benders Decomposition for Large-Scale Optimization. *Springer Optimization and Its Applications,* 149, 1–26. DOI:10.1007/978-3-030-22788-3_1

Hu, Y., & De Giacomo, G. (2011). Generalized Planning: Synthesizing Plans that Work for Multiple Environments. *Twenty-Second International Joint Conference on Artificial Intelligence.*

Huang, Y., He, X., & Dai, H. (2015). Systematization of metrics in intrusion detection systems. *ACM International Conference Proceeding Series*. DOI:10.1145/2746194.2746222

Iglesias García, J., Diaz, J., & Arroyo, D. (2020). Hyot: Leveraging Hyperledger for Constructing an Event-Based Traceability System in IoT. *Advances in Intelligent Systems and Computing, 951,* 195–204. DOI:10.1007/978-3-030-20005-3_20

ISO (2016). ISO/TC 307 Blockchain and distributed ledger technologies. https://www.iso.org/committee/6266604.html.

Jenny Gu, & Philip E. Bourne. (2009). *Structural Bioinformatics, 4.*

Jevtic, A., Flores Valle, A., Alenya, G. et al. (2019). Personalized Robot Assistant for Support in Dressing. *IEEE Transactions on Cognitive and Developmental Systems, 11*(3), 363–374. DOI:10.1109/TCDS.2018.2817283

Jimenez, A. R., Seco, F., & Torres-Sospedra, J. (2019). Tools for smartphone multi-sensor data registration and GT mapping for positioning applications. *International Conference on Indoor Positioning and Indoor Navigation, 1-8*. DOI:10.1109/IPIN.2019.8911784

Jiruska, P., de Curtis, M., Jefferys, J. G. R. et al. (2013). Synchronization and demapping: Controversies and hypotheses. The *Journal of Physiology*, 591(4), 787–797. DOI:10.1113/jphysiol.2012.239590

Jiruska, P., de Curtis, M., Jefferys, J. G. R. et al. (2013). Synchronization and desynchronization in epilepsy: Controversies and hypotheses. *Journal of Physiology*, 591(4), 787–797). DOI:10.1113/jphysiol.2012.239590

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence, 101*(1–2), 99–134. DOI:10.1016/s0004-3702(98)00023-x

Kennedy, P. (2019). Huawei Ascend 910 Provides a NVIDIA AI Training Alternative.

Khandpur, R. P., Ji, T., Jan, S. et al. (2017). Crowdsourcing Cybersecurity: Cyber Attack Detection using Social Media. *Proceedings of the International Conference on Information and Knowledge Management,* 1049–1057. DOI:10.1145/3132847.3132866

Kim, K., & Aminanto, M. E. (2018). Deep learning in intrusion detection perspective: Overview and further challenges. *2017 International Workshop on Big Data and Information Security,* 5–10. DOI:10.1109/IWBIS.2017.8275095

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*.

Knyazeva, M. G., Carmeli, C., Khadivi, A. et al. (2013). Evolution of source EEG synchronization in early Alzheimer's disease. *Neurobiology of Aging, 34*(3), 694–705. DOI:10.1016/j.neurobiolaging.2012.07.012

Koller, D., & Friedman, N. (2009). Probabilistic Graphical Models: Principles and Techniques. *MIT Press*

Kotseruba, I., & Tsotsos, J. K. (2020). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review, 53*(1), 17–94. DOI:10.1007/s10462-018-9646-y

Kowalski, R. (1979). Logic for Problem Solving. *North Holland*

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. In *Nature Neuroscience*, 21(9), 1148–1160. DOI:10.1038/s41593-018-0210-5

Krutz, R. L., & Dean Vines, R. (2002). The CISSP Prep Guide, Gold Edition. *John Wiley & Sons.*

Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: the embodied mind and its challenge to Western thought. *Basic Books*.

Lakshmanan, V., Gilleland, E., McGovern, A., Tingley, M. (2015). Machine learning and data mining approaches to climate science. *Proceedings of the 4th International Workshop on Climate Informatics.* DOI:10.1007/978-3-319-17220-0

Langley, P. (2017). Progress and Challenges in Research on Cognitive Architectures. *Proceedings of the AAAI Conference on Artificial Intelligence,* 31(1), 4870–4876

Lappas, T., Liu, K., & Terzi, E. (2009). Finding a team of experts in social networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 467–475. DOI:10.1145/1557019.1557074

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature,* 521, 436–444. DOI:10.1038/nature14539

Lee, J., Kim, C., Kang, S. et al. (2019). UNPU: An energy-efficient deep neural network accelerator with fully variable weight bit precision. *IEEE Journal of Solid-State Circuits, 54*(1), 173–185. DOI:10.1109/JSSC.2018.2865489

Lee, J., Lee, J., Han, D. et al. (2019). 7.7 LNPU: A 25.3TFLOPS/W Sparse Deep-Neural-Network Learning Processor with Fine-Grained Mixed Precision of FP8-FP16. *IEEE International Solid-State Circuits Conference,* 142–144. DOI:10.1109/ISSCC.2019.8662302

Lepri, B., Oliver, N., Letouzé, E. et al. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philosophy and Technology, 31*(4), 611–627. DOI:10.1007/s13347-017-0279-x

Li, X., Deng, S., Wang, S. et al. (2018). Review of Small Data Learning Methods. *Proceedings - International Computer Software and Applications Conference, 2,* 106–109. DOI:10.1109/COMPSAC.2018.10212

Li, X., Neil, D., Delbruck, T., & Liu, S. C. (2019). Lip reading deep network exploiting multi-modal spiking visual and auditory sensors. *Proceedings - IEEE International Symposium on Circuits and Systems,* 1-5. DOI:10.1109/ISCAS.2019.8702565

Liakos, K. G., Busato, P., Moshou, D. et al. (2018). Machine learning in agriculture: A review. *Sensors* 18(8):2674. DOI:10.3390/s18082674

Liese, F., & Miescke, K.-J. (2007). Statistical Decision Theory, 1-52. DOI:10.1007/978-0-387-73194-0_3

Lighthill, J. (1973). Artificial intelligence: A general survey. *Artificial Intelligence: a paper symposium*.

Linares-Barranco, B., & Serrano-Gotarredona, T. (2009). Exploiting memristance in adaptive asynchronous spiking neuromorphic nanotechnology systems, 601-604.

Lison, P., & Mavroeidis, V. (2017). Automatic Detection of Malware-Generated Domains with Recurrent Neural Models. *Arxiv.*

Litman, T. (2017). Autonomous vehicle implementation predictions. *Victoria Transport Policy Institute*

London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience,* 28, 503–532. DOI:10.1146/annurev.neuro.28.061604.135703

López, D., Monasterio, A., Toboso, M. et al. (2020). Cartography of the Values Involved in Robotics. *Biosystems and Biorobotics,* 25, 98–104.DOI:10.1007/978-3-030-24074-5_18

Love, B. C. (2016). Cognitive Models as Bridge between Brain and Behavior. *Trends in Cognitive Sciences*, 20(4), 247–248. DOI:10.1016/j.tics.2016.02.006

Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review, 3*(3), 127–149. DOI:10.1016/j.cosrev.2009.03.005

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Networks, 10*(9), 1659–1671. DOI:10.1016/S0893-6080(97)00011-7

Maghrebi, H. (2019). *Deep Learning based Side Channel Attacks in Practice. IACR Cryptol. ePrint Arch.,* 578.

Margarit-Taulé, J. M., Giménez-Gómez, P., Escudé-Pujol, R. et al. (2019). Live demonstration: A portable microsensor fusion system with real-time measurement for on-site beverage tasting. *Proceedings - IEEE International Symposium on Circuits and Systems.* DOI:10.1109/ISCAS.2019.8702184

Martínez-Rodríguez, M. C., Prada-Delgado, M. A., Brox, P., & Baturone, I. (2018). VLSI design of trusted virtual sensors. *Sensors, 18*(2). DOI:10.3390/s18020347

Martinez, D., Alenya, G., Ribeiro, T. et al. (2017). Relational Reinforcement Learning for Planning with Exogenous Effects. *Journal of Machine Learning Research,* 18(78),1–44,

Martínez, D., Alenyà, G., & Torras, C. (2017). Relational reinforcement learning with guided demonstrations. *Artificial Intelligence, 247,* 295–312. DOI:10.1016/j.artint.2015.02.006

Mattila, J., Koivumaki, J., Caldwell, D. G., & Semini, C. (2017). A survey on control of hydraulic robotic manipulators with projection to future trends. In *IEEE/ASME Transactions on Mechatronics,* 22(2), 669–680. DOI:10.1109/TMECH.2017.2668604

McCorduck, P., & Cfe, C. (2004). Machines Who Think. *Machines Who Think*. DOI:10.1201/9780429258985

Milanés, V., Villagrá, J., Pérez, J., & González, C. (2012). Low-speed longitudinal controllers for mass-produced cars: A comparative study. *IEEE Transactions on Industrial Electronics, 59*(1), 620–628. DOI:10.1109/TIE.2011.2148673

Minzioni, P., Lacava, C., Tanabe, T. et al. (2019). Roadmap on all-optical processing. *Journal of Optics (United Kingdom), 21*(6), 063001. DOI:10.1088/2040-8986/ab0e66

Mitchell, T. (1997). Machine Learning. *McGraw Hill.*

Mitchie, D. (1979). Expert Systems in the Microelectronics Age. *Edinburgh University Press*

Monasterio Astobiza, A., Toboso, M., Aparicio, M. et al. (2019). Bringing inclusivity to robotics with INBOTS. *Nature Machine Intelligence,* 1(4), 164. DOI:10.1038/s42256-019-0040-5

Müller, J. P., & Fischer, K. (2014). Application Impact of Multiagent Systems and Technologies: A Survey. *Springer*

Muñoz-Martín, I., Bianchi, S., Pedretti, G. et al. (2019). Unsupervised Learning to Overcome Catastrophic Forgetting in Neural Networks. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, 5*(1), 58–66. DOI:10.1109/JXCDC.2019.2911135

Naveiro, R., & Insua, D. R. (2019). Gradient Methods for Solving Stackelberg Games. *International Conference on Algorithmic Decision Theory*, 126–140.

Newell, A., & Simon, H. A. (2011). Computer science as empirical inquiry: symbols and search. *ACM Turing Award Lectures*, 19,113–126. DOI:10.1145/1283920.1283930

Newell, A., & Simon, H. (1972). Human Problem Solving. *Prentice-Hall*

O'Connor, P., Neil, D., Liu, S. C. et al. (2013). Real-time classification and sensor fusion with a spiking deep belief network. *Frontiers in Neuroscience, 7, 178*. DOI:10.3389/fnins.2013.00178

Olivares-Alarcos, A., Foix, S., & Alenyà, G. (2019). On inferring intentions in shared tasks for industrial collaborative robots. *Electronics, 8*(11). DOI:10.3390/electronics8111306

Osborne, M. J., & Rubinstein, A. (1994). A Course in Game Theory.

Osman, N. (2018). Partakable technology. *IJCAI International Joint Conference on Artificial Intelligence*, 5714–5718. DOI:10.24963/ijcai.2018/815

Osterweil, L. J., Ghezzi, C., Kramer, J., & Wolf, A. L. (2008). Determining the Impact of Software Engineering Research on Practice. *Computer*, *41*(3), 39–49. DOI:10.1109/MC.2008.85

Pack Kaelbling, L., Littman, M. L., Moore, A. W., & Hall, S. (1996). Reinforcement Learning: A Survey. *Journal of Artiicial Intelligence Research*, 4, 237–285.

Parsons, S., Sierra, C., & Jennings, N. (1998). Agents that reason and negotiate by arguing. *Journal of Logic and Computation, 8*(3), 261–292. DOI:10.1093/logcom/8.3.261

Pasquale, F. (2015). The Black Box Society. *Harvard University Press.*

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics, 87*(3), 925. DOI:10.1103/RevModPhys.87.925

Pathinarupothi, R. K., Durga, P., & Rangan, E. S. (2018). Data to diagnosis in global health: A 3P approach. *BMC Medical Informatics and Decision Making, 18*(1). DOI:10.1186/s12911-018-0658-y

Pearl, J. (1984). Heuristics. *Addison-Wesley Publishing Company*

Pearl, J., Glymour, M. & Jewell, N. P. (2016). Causal Inference in Statistics: A Primer. *Wiley*

Peralta, A. F., Carro, A., San Miguel, M., & Toral, R. (2018). Analytical and numerical study of the non-linear noisy voter model on complex networks. *Chaos*, *28*(7), 075516. DOI:10.1063/1.5030112

Pérez-Carrasco, J. A., Zhao, B., Serrano, C. et al. (2013). Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing - Application to feedforward convnets. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(11), 2706–2719. DOI:10.1109/TPAMI.2013.71

Polydoros, A. S., & Nalpantidis, L. (2017). Survey of Model-Based Reinforcement Learning: Applications on Robotics. *Journal of Intelligent and Robotic Systems: Theory and Applications*, *86*(2), 153–173. DOI:10.1007/s10846-017-0468-y

Port, R. F., & van Gelder, T. (1995). Mind as Motion: Explorations in the Dynamics of Cognition. *The MIT Press.*

Prosser, S. (2016). Experiencing Time. *Oxford University Press.*

Querejeta-Azurmendi, I., Hernandez Encinas, L., Arroyo, D., & Hernández-Ardieta, J. L. (2020). An Internet Voting Proposal Towards Improving Usability and Coercion Resistance. *Advances in Intelligent Systems and Computing, 951*, 155–164. DOI:10.1007/978-3-030-20005-3_16

Quijada, R., Dura, R., Pallares, J., Formatje, X. et al. (2018). Large-Area Automated Layout Extraction Methodology for Full-IC Reverse Engineering. *Journal of Hardware and Systems Security*, *2*(4), 322–332. DOI:10.1007/s41635-018-0051-4

Radovic, A., Williams, M., Rousseau, D. et al. (2018). Machine learning at the energy and intensity frontiers of particle physics. *Nature*, 560(7716), 41–48. DOI:10.1038/s41586-018-0361-2

Raju, L., Sankar, S., & Milton, R. S. (2015). Distributed optimization of solar micro-grid using multi agent reinforcement learning. *Procedia Computer Science, 46,* 231–239. DOI:10.1016/j.procs.2015.02.016

Renaudin, V., Ortiz, M., Perul, J. et al. (2019). Evaluating Indoor Positioning Systems in a Shopping Mall: The Lessons Learned from the IPIN 2018 Competition. *IEEE Access, 7,* 148594–148628. DOI:10.1109/ACCESS.2019.2944389

Rimmer, V., Preuveneers, D., Juárez, M. et al. (2017). Automated Feature Extraction for Website Fingerprinting through Deep Learning. *ArXiv preprint.*

Rivest, R. L. (1991). Cryptography and Machine Learning. *International Conference on the Theory and Application of Cryptology,* 427–439.

Robertson, D., Giunchiglia, F., Van Harmelen, F. et al. (2008). Open knowledge coordinating knowledge sharing through peer-to-peer interaction. *Lecture Notes in Computer Science, 5118 LNAI,* 1–18. DOI:10.1007/978-3-540-85058-8_1

Rumelhart, D., McCleland, J. (1986). Parallel Distributed Processing. *MIT Press.*

Russell, S., Norving, P. (2010). Artificial Intelligence. *Prentice Hall.*

Sabater, J., & Sierra, C. (2002). Reputation and social network analysis in multi-agent systems. *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent System,* 475-482. DOI:10.1145/544741.544854

Sabottke, C., Suciu, O., Dumitraş, T., & Dumitras, T. (2015). Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits. *24th {USENIX} Security Symposium,* 1041-1056.

Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine, 41*(12), 1110–1117. DOI:10.1016/j.compbiomed.2011.06.020

Salathé, M., Bengtsson, L., Bodnar, T. J. et al. (2012). Digital epidemiology. *PLoS Computational Biology, 8*(7), 1002616. DOI:10.1371/journal.pcbi.1002616

Salazar Gonzalez, J. L., Soria Morillo, L. M., Alvarez-Garcia, J. A. et al. (2019). Energy-Efficient Indoor Localization WiFi-Fingerprint System: An Experimental Study *IEEE Access, 7,* 162664–162682. DOI:10.1109/ACCESS.2019.2952221

Santamaria-Navarro, A., Loianno, G., Solà, J. et al. (2018). Autonomous navigation of micro aerial vehicles using high-rate and low-cost sensors. *Autonomous Robots, 42*(6), 1263–1280. DOI:10.1007/s10514-017-9690-5

Savarimuthu, T. R., Buch, A. G., Schlette, C. et al. (2018). Teaching a Robot the Semantics of Assembly Tasks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48*(5), 670–692. DOI:10.1109/TSMC.2016.2635479

Schnitzler, A., & Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nature Reviews Neuroscience,* 6(4), 285–96. DOI:10.1038/nrn1650

Segovia-Aguas, J., Jiménez, S., & Jonsson, A. (2018). Computing hierarchical finite state controllers with classical planning. *Journal of Artificial Intelligence Research, 62,* 755–797. DOI:10.1613/jair.1.11227

Segovia-Aguas, J., Jiménez, S., & Jonsson, A. (2019). Computing programs for generalized planning using a classical planner. *Artificial Intelligence, 272,* 52–85. DOI:10.1016/j.artint.2018.10.006

Shapiro, L. (2011). Embodied Cognition. *Routledge*

Shoshitaishvili, Y., Weissbacher, M., Dresel, L. et al. (2017). Rise of the HaCRS: Augmenting Autonomous Cyber Reasoning Systems with Human Assistance. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security,* 347-362.

Siciliano, B., & Khatib, O. (2016). Robotics and the handbook. *Springer Handbook of Robotics,* 1-6. DOI:10.1007/978-3-319-32552-1_1

Sierra, C. & Dignum, F. (2001). Agent Mediated Electronic Commerce: The European AgentLink Perspective. *Springer-Verlag Berlin Heidelberg,* 126-147.

Simo-Serra, E., Torras, C., & Moreno-Noguer, F. (2017). 3D Human Pose Tracking Priors using Geodesic Mixture Models. *International Journal of Computer Vision, 122*(2), 388–408. DOI:10.1007/s11263-016-0941-2

Song, J., Cho, Y., Park, J. S. et al. (2019). 7.1 An 11.5TOPS/W 1024-MAC Butterfly Structure Dual-Core Sparsity-Aware Neural Processing Unit in 8nm Flagship Mobile SoC. *Digest of Technical Papers - IEEE International Solid-State Circuits Conference,* 130–133. DOI:10.1109/ISSCC.2019.8662476

Srivastava, S., Immerman, N., & Zilberstein, S. (2008). Learning Generalized Plans Using Abstract Counting. *AAAI,* 8, 991-997.

Stamp, M. (2017). Introduction to machine learning with applications in information security. *Chapman & Hall.*

Stark, P. B. (2016). Privacy, Big Data, and the Public Good: Frameworks for Engagement. *The American Statistician, 70*(1), 119–119. DOI:10.1080/00031305.2015.1068625

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature, 576*(7786), 266–273. DOI:10.1038/s41586-019-1787-x

Stephen Marsland. (2015). Machine Learning: An Algorithmic Perspective. *CRC Pres.*

Stroke, G.W. (1972). Optical computing. *IEEE Spectrum,* 9(12)*, 24-41.*

Su, J., Vargas, D. V., Prasad, S. et al. (2017). Lightweight Classification of IoT Malware based on Image Recognition. *IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC),* 2, 664-669.

Sun, R. (2007). The importance of cognitive architectures: An analysis based on CLARION. *Journal of Experimental and Theoretical Artificial Intelligence, 19*(2), 159–193. DOI:10.1080/09528130701191560

Tang, A., Sethumadhavan, S., & Stolfo, S. (2014). Unsupervised Anomaly-based Malware Detection using Hardware Features. *International Workshop on Recent Advances in Intrusion Detection,109-129.*

Taylor, M. E., & Stone, P. (2009). Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research,*10, 1633-1685.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522. DOI:10.1038/381520a0

Thuruthel, G. T., Ansari, Y., Falotico, E., & Laschi, C. (2018). Control Strategies for Soft Robotic Manipulators: A Survey. *Soft Robotics,* 5(2), 149–163. DOI:10.1089/soro.2017.0007

Tzafestas, S. G. (2018). Mobile Robot Control and Navigation: A Global Overview. *Journal of Intelligent and Robotic Systems: Theory and Applications, 91*(1), 35–58. DOI:10.1007/s10846-018-0805-9

Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers and Security,* 81, 123–147. DOI:10.1016/j.cose.2018.11.001

Uhlhaas, P. J., & Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews Neuroscience,* 11(2),100–113. DOI:10.1038/nrn2774

Vacca, J. (2012). Computer and Information Security Handbook.

Van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy, 92*(7), 345–381. DOI:10.2307/2941061

van Krieken, E., Acar, E., & van Harmelen, F. (2020). Analyzing Differentiable Fuzzy Logic Operators. *ArXiv*.

Varela, F. J., Thompson, E., Rosch, E., & Kabat-Zinn, J. (2016). The embodied mind: Cognitive science and human experience. *The MIT Press.* DOI:10.29173/cmplct8718

Veloso, M., Carbonell, J., Perez, A. et al. (1995). Integrating planning and learning: The prodigy architecture. *Journal of Experimental and Theoretical Artificial Intelligence, 7*(1), 81–120. DOI:10.1080/09528139508953801

Verde, R., & Petrucci, A. (2017). SIS 2017. Statistics and Data Science: new challenges, new generations. *Firenze University Press,* 114. DOI:10.36253/978-88-6453-521-0

Vernon, D. (2014). Artificial Cognitive Systems : A Primer. *Journal of Chemical Information and Modeling,* 53(9),1–35.

Vespignani, A. (2009). Predicting the behavior of techno-social systems. *Science,* 325(5939), 425–428. DOI:10.1126/science.1171990

Vidal, J. (2017). Tsunami of data'could consume one fifth of global electricity by 2025. *Climate Home News, 11.*

Vitali, S., Glattfelder, J. B., & Battiston, S. (2011). The network of Global corporate control. *PLoS ONE, 6*(10). DOI:10.1371/journal.pone.0025995

Vithu, P., & Moses, J. A. (2016). Machine vision system for food grain quality evaluation: A review. *Trends in Food Science and Technology,* 56, 13–20. DOI:10.1016/j.tifs.2016.07.011

Vogt, N. (2018). Machine learning in neuroscience. *Nature Methods,* 15(1), 33. DOI:10.1038/nmeth.4549

Voigt, P., & von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR) A Practical Guide. *Springer International Publishing.*

Vorobeichyk, Y., & Kantarcioglu, M. (2018). Adversarial Machine Learning. *Morgan Clayton.* DOI:10.2200/S00861ED1V01Y201806AIM039

Wallach, H. (2018). Computational social science ? computer science + social data. *Communications of the ACM, 61*(3), 42–44. DOI:10.1145/3132698

Warman, M. (2009). Green I.T.: How many Google searches does it take to boil a kettle?

Watts, D. J. *(2011).* Everything is Obvious. How Common Sense Fails. *Crown Bussiness.*

Weisberg, S. M., & Newcombe, N. S. (2017). Embodied cognition and STEM learning: overview of a topical collection in CR:PI. *Cognitive Research: Principles and Implications,*2. DOI:10.1186/s41235-017-0071-6

Werbach, K. (2018). The Blockchain and the New Architecture of Trust. *The MIT Press.* DOI:10.7551/mitpress/11449.001.0001

Westberg, M., Zelvelder, A., & Najjar, A. (2019). A historical perspective on cognitive science and its influence on XAI research. *Lecture Notes in Computer Science, 11763,* 205–219. DOI:10.1007/978-3-030-30391-4_12

Westwood, S. J., Iyengar, S., Walgrave, S. et al. (2018). The tie that divides: Cross-national evidence of the primacy of partyism. *European Journal of Political Research, 57*(2), 333–354. DOI:10.1111/1475-6765.12228

Yamada, Y., Uchiyama, M., Jobashi, M. et al. (2020). A 20.5 TOPS Multicore SoC with DNN Accelerator and Image Signal Processor for Automotive Applications. *IEEE Journal of Solid-State Circuits, 55*(1), 120–132. DOI:10.1109/JSSC.2019.2951391

Yang, P., Tang, K., & Yao, X. (2019). A Parallel Divide-and-Conquer-Based Evolutionary Algorithm for Large-Scale Optimization. *IEEE Access, 7,* 163105–163118. DOI:10.1109/ACCESS.2019.2938765

Yoo, J., Yan, L., El-Damak, D. et al. (2013). An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor. *IEEE Journal of Solid-State Circuits, 48*(1), 214–228. DOI:10.1109/JSSC.2012.2221220

Yu, B., Gray, D. L., Pan, J. et al. (2017). Inline DGA detection with deep networks. *IEEE International Conference on Data Mining Workshops,* 683–692. DOI:10.1109/ICDMW.2017.96

Zhou, Z., Chen, X., Li, E. et al. (2019). Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing. *Proceedings of the IEEE, 107*(8).

Zimmermann, J. B., & Jackson, A. (2014). Closed-loop control of spinal cord stimulation to restore hand function after paralysis. *Frontiers in Neuroscience, 8*. DOI:10.3389/fnins.2014.00087

CSIC white paper on Artificial Intelligence, Robotics and Data Science sketches a preliminary roadmap for addressing current R&D challenges associated with automated and autonomous machines. More than 50 research challenges investigated all over Spain by more than 150 experts within CSIC are presented in eight chapters. Chapter One introduces key concepts and tackles the issue of the integration of knowledge (representation), reasoning and learning in the design of artificial entities. Chapter Two analyses challenges associated with the development of theories –and supporting technologies– for modelling the behaviour of autonomous agents. Specifically, it pays attention to the interplay between elements at micro level (individual autonomous agent interactions) with the macro world (the properties we seek in large and complex societies). While Chapter Three discusses the variety of data science applications currently used in all fields of science, paying particular attention to Machine Learning (ML) techniques, Chapter Four presents current development in various areas of robotics. Chapter Five explores the challenges associated with computational cognitive models. Chapter Six pays attention to the ethical, legal, economic and social challenges coming alongside the development of smart systems. Chapter Seven engages with the problem of the environmental sustainability of deploying intelligent systems at large scale. Finally, Chapter Eight deals with the complexity of ensuring the security, safety, resilience and privacy-protection of smart systems against cyber threats.

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
CSIC
EDITORIAL CSIC