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

VOLUME 11

ARTIFICIAL INTELLIGENCE, ROBOTICS & DATA SCIENCE

Catálogo de publicaciones de la Administración General del Estado:
https://cpage.mpr.gob.es

Editorial CSIC:
http://editorial.csic.es (correo: *publ@csic.es*)

ISBN Vol. 11: 978-84-00-10758-1
ISBN O.C.: 978-84-00-10736-9
e-ISBN Vol. 11: 978-84-00-10759-8
e-ISBN O.C.: 978-84-00-10734-5
NIPO: 833-21-091-1
e-NIPO: 833-21-092-7
DL O.C: M-2426-2021

Diseño y maquetación: gráfica futura

CSIC SCIENTIFIC CHALLENGES: TOWARDS 2030

VOLUME 11
ARTIFICIAL INTELLIGENCE, ROBOTICS & DATA SCIENCE

Topic Coordinators
Sara Degli Esposti & Carles Sierra

CSIC SCIENTIFIC CHALLENGES: TOWARDS 2030

What are the major scientific challenges of the first half of the 21st century? Can we establish the priorities for the future? How should the scientific community tackle them?

This book presents the reflections of the Spanish National Research Council (CSIC) on 14 strategic themes established on the basis of their scientific impact and social importance.

Fundamental questions are addressed, including the origin of life, the exploration of the universe, artificial intelligence, the development of clean, safe and efficient energy or the understanding of brain function. The document identifies complex challenges in areas such as health and social sciences and the selected strategic themes cover both basic issues and potential applications of knowledge. Nearly 1,100 researchers from more than 100 CSIC centres and other institutions (public research organisations, universities, etc.) have participated in this analysis. All agree on the need for a multidisciplinary approach and the promotion of collaborative research to enable the implementation of ambitious projects focused on specific topics.

These 14 "White Papers", designed to serve as a frame of reference for the development of the institution's scientific strategy, will provide an insight into the research currently being accomplished at the CSIC, and at the same time, build a global vision of what will be the key scientific challenges over the next decade.

VOLUMES THAT MAKE UP THE WORK

1 *New Foundations for a Sustainable Global Society*
2 *Origins, (Co)Evolution, Diversity and Synthesis of Life*
3 *Genome & Epigenetics*
4 *Challenges in Biomedicine and Health*
5 *Brain, Mind & Behaviour*
6 *Sustainable Primary Production*
7 *Global Change Impacts*
8 *Clean, Safe and Efficient Energy*
9 *Understanding the Basic Components of the Universe, its Structure and Evolution*
10 *Digital and Complex Information*
11 *Artificial Intelligence, Robotics and Data Science*
12 *Our Future? Space, Colonization and Exploration*
13 *Ocean Science Challenges for 2030*
14 *Dynamic Earth: Probing the Past, Preparing for the Future*

CSIC scientific challenges: towards 2030
Challenges coordinated by:
Jesús Marco de Lucas & M. Victoria Moreno-Arribas

Volume 11
Artificial Intelligence, Robotics and Data Science

Participating Researchers and Centres

Topic Coordinators
Carles Sierra (IIIA-CSIC) and Sara Degli Esposti (IPP-CSIC)

Challenges Coordinators
Felip Manyà (IIIA-CSIC), Adrià Colomé (IRI-CSIC); Nardine Osman (IIIA-CSIC), Daniel López (IFS-CSIC); José Javier Ramasco Sukia (IFISC-CSIC), Lara Lloret Iglesias (IFCA-CSIC); Guillem Alenyà (IRI-CSIC), Jorge Villagra (CAR-CSIC); M. Dolores del Castillo (CAR-CSIC), Marco Schorlemmer (IIIA-CSIC); Pablo Noriega (IIIA-CSIC), Txetxu Ausín (IFS-CSIC); Teresa Serrano (IMSE-CNM-CSIC), Arantza Oyanguren (IFIC-CSIC); David Arroyo Guardeño (ITEFI-CSIC), Piedad Brox Jiménez (IMSE-CNM-CSIC).

Participating Researchers
Centro de Automática y Robótica (CAR, CSIC)
Centro de Investigaciones Biológicas (CIB, CSIC)
Instituto de Análisis Económico (IAE, CSIC)
Instituto Cajal (IC, CSIC)
Instituto de Ciencias del Espacio (ICE, CSIC)
Instituto de Ciencias Matemáticas (ICMAT, CSIC-UAM-UC3M-UCM)
Geociencias Barcelona (GEO3BCN, CSIC)
Instituto de Ciencias de la Vid y del Vino (ICVV, CSIC)
Instituto de Economía, Geografía y Demografía (IEGD, CSIC)
Instituto de Física de Cantabria (IFCA, CSIC– UC)
Instituto de Física Corpuscular (IFIC, CSIC – UV)
Instituto de Física Interdisciplinar y Sistemas (IFISC, CSIC – UIB)
Instituto de Filosofía (IFS, CSIC)
Instituto de la Grasa (IG, CSIC)
Instituto de Investigación en Inteligencia Artificial (IIIA, CSIC)
Instituto de Investigaciones Marinas (IIM, CSIC)
Instituto de Lengua, Literatura y Antropología (ILLA, CSIC)
Instituto de Microelectrónica de Barcelona (IMB-CNM, CSIC)
Institución Milá y Fontanals (IMF, CSIC)
Instituto de Microelectrónica de Sevilla (IMSE-CNM, CSIC – US)
Instituto de Neurociencias (IN, CSIC – UMH)
Instituto de Gestión de la Innovación y del Conocimiento (INGENIO, CSIC-UPV)
Instituto de Políticas y Bienes Públicos (IPP, CSIC)
Instituto de Química Orgánica General (IQOG, CSIC)
Instituto de Robótica e Informática Industrial (IRII, CSIC)
Instituto de Tecnologías Físicas y de la Información Leonardo Torres Quevedo (ITEFI, CSIC)
Avantopy

VOLUME 11 SUMMARY

CHALLENGE 8

SMART CYBERSECURITY

Coordinators

D. Arroyo Guardeño (ITEFI, CSIC)

P. Brox Jiménez (IMSE-CNM, CSIC – US)

Participant researchers and centers

J. Godoy (CAR, CSIC – UPM)

J. Villagra (CAR, CSIC – UPM)

H. Mueller (IAE, CSIC)

V. Gallego (ICMAT, CSIC-UAM-UC3M-UCM)

A. Kosgodagan (ICMAT, CSIC-UAM-UC3M-UCM)

R. Naveiro (ICMAT, CSIC-UAM-UC3M-UCM)

D. Rios Insua (ICMAT, CSIC-UAM-UC3M-UCM)

D. Rodríguez González (IFCA, CSIC – UC)

S. Hidalgo Villena (IMB-CNM)

S. Degli Esposti (IPP, CSIC)

P. Noheda Marín (IQOG, CSIC)

1. EXECUTIVE SUMMARY

The relationship between AI and cybersecurity is of conflicting nature. While threat detection and containment can be perfected by the use of new AI tools and methodologies, the current data deluge and the increasing complexity of *information and communication technologies* (ICT) make almost impossible to properly protect them without the guidance of automatic decision making solutions. However, an excess of confidence in such solutions can compromise security and enclose safety risks for information systems. Moreover, neglecting the treatment of these risks could undermine the different modalities of governance that configure some pillars of our democracy. Fundamental citizens' rights such as privacy or accountability in public procurement are major components of the conundrum related to the alignment of technological possibilities with ethical, legal and normative regulations. This chapter summarizes the CSIC approach to tackle the above challenges in the crossed domain of AI and cybersecurity, ***underlining the need for an integral strategy for the deployment of secure and safe AI systems.*** This approach encompasses the whole stack associated with the design and implementation of AI solutions, ranging from the hardware to the application layer, and considering the

theoretical underpinnings to adequately bridge AI functionality and cybersecurity requirements.

2. INTRODUCTION AND GENERAL DESCRIPTION

ICT systems need to integrate approaches to minimise security risks. In order to forecast, monitor, and update the *security of ICT systems, techniques based on AI and other automated tools contribute to threat detection* in massive data processing with minimal human intervention in order to protect national critical infrastructures. According to the EU NIS Directive (2016/1148), the 2019 Spanish Cybersecurity Strategy and global standards (e.g. ISO27001, ISO27005), cyberspace needs to be protected from malicious and illicit activities of all kinds. The resilience and operational continuity of critical infrastructures (water, energy, transport, financial and health sectors) need also to be ensured. Computer Security Incident Response Team must be equipped with the appropriate advanced cybersecurity tools to adequately respond to attacks or system failures.

All private and public actors must contribute to achieve the following objectives. First, the protection of computer systems, networks between computer systems, networks of networks (social networks included) against eventual attacks to their hardware, software and electronic data, as well as the disruptions or failures of the services they provide (computer technology). For instance, *Global Navigation Satellite Systems* (GNSS) are strategic to provide geo-spatial positioning. Second, in order to ensure the trustworthy development of interconnected digital technologies, we need two things. A significant improvement in the quality and safety of AI-assisted services facilitating the interaction between humans and between humans and bots. The safe, proportionate and ethical processing of massive amounts of data through a robust infrastructure to enable extensive and intensive digital communication activity both terrestrial and through satellites.

One of the added problems that advanced cybersecurity has to handle is that there is no overarching government ruling in the cyber-physical domain. It has generated and generates numerous conflicts over boundaries, hierarchies of rights and priorities over how the technology should be designed and used. Additionally, the irruption of available efficient *machine learning* (ML) algorithms (in special, *deep learning* [DL] algorithms) and their ubiquitous implementations in public and private services (including, health, education, justice, governance, public and administration, citizen security, wellness, mobility,

business management and optimization, marketing and demography), together with autonomous computer security bots that automatically repair security vulnerabilities without human intervention (cybersecurity reasoning systems), increase the need for advanced cybersecurity research with capability of facing new kinds of global safety vulnerabilities, risks and threats.

Governments cannot only rely on private companies to develop next-generation cybersecurity solutions, but need to have their own experts to ensure common security criteria are respected in product development, cryptographic protocols are robust, backdoors are not built into critical systems. Starting from a vision of *cybersecurity as a public good,* public-private partnerships on cybersecurity need to maximise and advance theoretical and operational knowledge on risks and vulnerabilities of those digitally automated systems on which society increasingly relies to function.

As a huge variety of applications is being automated through ML algorithms, it is essential that these techniques are robust and reliable as many decisions are based on their outputs. State-of-the-art ML algorithms perform extraordinarily well on standard data but, recently, have been shown to be vulnerable to adversarial examples, data instances targeted at fooling them (Goodfellow, Shlens and Szegedy, 2015). Algorithms designer should take into account the possible presence of adversaries to be robust against such data manipulations. The work in (Comiter, 2019) provides a review from the policy perspective showing how many AI societal systems, including content filters, military systems, law enforcement systems and autonomous vehicles (AV), to name but a few, are susceptible and vulnerable to attacks. The proper alignment between the benefits of the new cyberphysical reality and the ***protection against cyberthreats using AI techniques is a major challenge.*** To face this properly, cybersecurity solutions have to be evaluated.

Traditionally, information security has been structured around the protection of the *confidentiality, integrity and availability* (CIA) of information. The complexity of living in an hyper-connected environment demands going beyond the CIA triad (Vacca, 2013) to include security objectives such as *authentication, authorization and auditability* (3Au scheme) procedures (Krutz and Vines, 2002). This implies the deployment of security policies focused on the delimitation of a security perimeter supported by access control policies (e.g. traffic routers, firewalls) that enable network traffic analysis and control. The correct application of policies and procedures to protect the CIA and 3Au properties guarantee the proper functioning of information systems and communication

channels. Thus, information assets are protected through the preservation of the CIA+3Au, and the adequate creation of a secure perimeter. Security can only be achieved if there is a coherent and consistent articulation of such paradigms in relation to the solutions for achieving a good security perimeter. Moreover, the integration of AI as part of an integral cybersecurity strategy should acknowledge the contributions from the conventional information security framework, but it has to be extended to treat other aspects arising form (cyber)safety, physical harms, cyberattacks targeted at hindering assets as reputation and trust in institutions, organizations and governments, as it occurs in the case of hybrid warfare. Data-driven methodologies and tools should be devised to further extend the CIA+3Au and construct a smart perimeter to articulate cybersecurity and cybersafety more coherent and consistent manner.

The variable and dynamic nature of various modes of information exchange makes it necessary to reinterpret the *security perimeter,* in which the concept of digital identity plays a central role. A digital identity is nothing more than an application that assigns a user a place in the cyberspace. That application is an operation performed on the basis of something users know (password) and something that identify them (a biometric feature or an ID). These digital identity management mechanisms require the involvement of some sort of central authority, which is responsible for checking the authentication information provided by the user and authorizes access to the system/service if it is valid. In practice this means that this central authority has stored information that enables the verification of digital identities. If the central authority loses this information (e.g., credential theft in cloud services such as Dropbox) there is a serious security problem. In addition, the central authority has the ability to record all the activity of a digital identity and, therefore, of the corresponding user. The latter infringes on users' privacy, something that may have legal and/or regulatory consequences, in addition to a possible deterioration in users' confidence in ICTs. This impact is of major relevance if we take into consideration AI applications based on the automatic treatment of personal data to sustain or deliver automatic decision making. In other words, the outsourcing of data storage and computing encloses a challenge in terms of governance, which should be properly considered and managed to adequately include AI as part of any integral cybersecurity strategy. As an important component of the CSIC smart cybersecurity plan, blockchain and distributed ledger technologies are discussed as a means to monitor AI activity, foster accountability by means of advanced digital evidence recording and treatment, and promote transparency in the context of e-governance.

3. IMPACT IN BASIC SCIENCE PANORAMA AND POTENTIAL APPLICATIONS

Information security in general, and cybersecurity in particular, are built upon the intertwining of multiple disciplines and knowledge domains. Ranging from communication systems to networks, any software application cannot be considered neither safe nor secure unless the underlying hardware is properly analyzed and validated, which also calls for an exhaustive pondering of each single functional and non-functional requirement. In the very case of AI and cybersecurity, the dual nature of their connection demands the articulation of holistic approaches. Advances in AI disciplines as *natural language processing* (NLP), recommender systems, people analytics, DL, and chatbots have increased the effectiveness of different types of autonomous agents in scenarios as digital on-boarding in banking systems, AI-powered messaging in e-commerce and targeted advertising, robotic autonomous systems, autonomous vehicles, unmanned aerial vehicles, cyber weaponry, and automatic writing. In all these contexts, information is not only captured, transformed and exchanged. Certainly, automation is imbricated in the very generation of new pieces of information that could be further fed into (automatic) decision making/taking processes. Security and safety in all these areas demands knowledge construction and sharing between professionals and researchers with very different background and interests. Therefore, there is an urge to foster interdisciplinary theoretical innovation in cybersecurity, which has a special relevance in the case of the crossed domain of AI and cybersecurity.

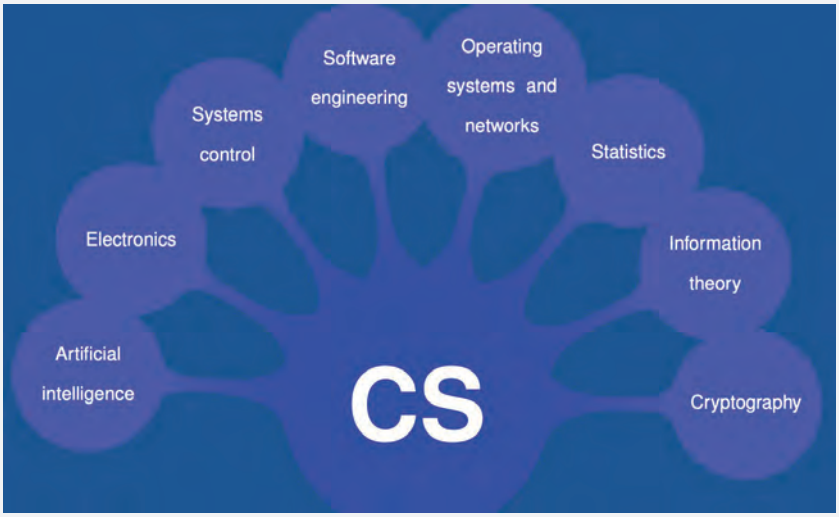
3.1. Fundamentals

Mathematical, logical and physical foundations

Many mathematical sciences aspects are relevant at the AI-cybersecurity interface. To name but a few, statistical and ML methods are essential for properly uncovering patterns and trends in attacks; risk analysis provides methodologies for the proper management of threats; game-theoretic methods facilitate modelling the presence of intelligent adversaries. Advances in these areas facilitate progress in cybersecurity analytics; in turn, complex problems in the cybersecurity domain motivate advances in mathematical sciences.

Smart cybersecurity is going to be articulated in cyberphysical enviroments where any mathematical framework should be properly enhanced and adapted by physical underpinnings. The application of formal methods, ϖ calculus and communication sequential processes, is of major relevance to bridge

FIGURE 8.1–Knowledge domains involved in the deployment of an holistic cybersecurity policy.



mathematical tools and the requirements and limitations of different operational contexts. The latter has been successfully applied in cybernetics and control systems to verify safety and liveness properties. Nonetheless, there should be a roadmap to better bridge AI, cybersecurity and control theory.

Threat modeling

AI can be applied to the analysis of malicious activity in information systems and networks. This could contribute to extend and enhance current methods to characterize computer security threats, as STRIDE. Indeed, AI can aid in grasping more adaptively the tactics and strategies of potential attackers. It is worth noting the difficulties associated with modeling insider threats and byzantine faults, the threats associated with users with privileged information about an organization, but also the problems associated to spurious activity and unintentional misuse of information systems. Along with *data mining* (DM) and ML, game theory and formal methods based on ϖ calculus should be integrated into general frameworks for the characterization of attack vectors and their impact in information systems (Diaz et al., 2019).

With the purpose of *enforcing cybersecurity,* there exist several security threats whose subsequent analysis can be very helpful to protect the systems at the

end. *Big data* (BD) tools should be applied in the design, implementation and validation of adaptive systems for the protection of cyberspace and, therefore, of our physical world (Curry et al., 2013). To this end, the first step is to carry out an in-depth review of the uses of DM, ML and AI, both in the characterization of abuses in the use of ICT and in the establishment of safeguards and other proactive mechanisms for a safer and more satisfactory experience in the ICT field (Stamp, 2017). Successful cases of DM-ML-AI can be found in the context of cryptography (Rivest, 1991; Bost et al., 2015), the identification of malicious software (Ucci, Aniello and Baldoni, 2017) and intruders (Huang, He and Dai, 2015) or the attribution of responsibility in the diffusion of software (Caliskan-Islam et al., 2015). Based on the recognition of the work in each of the cybersecurity domains where some element of the DM-ML-AI triplet has been applied, we also need to realize the need for comprehensive approaches to analyze heterogeneous data sources to identify security threats quickly and effectively (de la Torre, Lago and Arroyo, 2019). Likewise, we have to take into consideration that such an integral approach must be continuously adjusted, to react in a convenient way to short-term or structural modifications of the operation context.

In this context, DL is presented as a key element in providing such a comprehensive approach. In recent years, deep neural network models have broken records in many ML application areas, from NLP to machine vision, and have recently begun to be applied in the context of cybersecurity. The main hypothesis of the medium-term objectives consists of considering DL techniques as the fundamental basis of the integral scheme for early detection of threats in information security management systems. This approach takes as a reference, in the first place, the advantages of deep networks as mechanisms for the automatic codification of features and their capacity to take advantage of existing structural and spatial dependencies in data to build reliable models. Secondly, it takes into account the positive results of DL in the framework of cybersecurity, in applications such as intrusion detection(Kim and Aminanto, 2017), privacy protection assessment (Rimmer et al., 2017), malware identification (Bisio et al., 2017; Yu et al., 2017; Lison and Mavroeidis, 2017; Su et al., 2018), continuous user authentication (Shoshitaishvili et al., 2017), etc. Finally, the comprehensive scheme for dealing with security events and incidents must be supported by an adequate articulation of its life cycle. In this respect, a methodology for the correction of the hyperparameter space of DL models will be implemented based on recent results in the field of *adversarial ML* (AML) and algorithmic fairness (Lepri, 2018).

Adversarial machine learning

A basic hypothesis of ML theory is that their systems rely on *independent and identically distributed* data for both the training and testing phases. However, security aspects, which conform the field of AML, question such hypothesis given the presence of adaptive adversaries ready to modify the data and obtain a benefit, potentially degrading the performance of ML algorithms with important consequences. Practically all ML methodologies have been touched upon from an adversarial perspective including, to quote just a few, naive Bayes; logistic regression; support vector machines; latent Dirichlet allocations; or deep neural networks.

AML is a difficult area rapidly evolving through an arms race in which the community alternates a cycle of proposing attacks and implementing defences. Thus, it is crucial to develop sound techniques. Note though that, stemming from the pioneering (Dalvi et al., 2004), most of this research has been framed within a standard game-theoretic approach pervaded by Nash equilibrium and refinements. However, these entail common knowledge assumptions which are hard to maintain in the security realm. We could argue that *common knowledge* is too commonly assumed. A new paradigm seems to be required.

Disintermediation and decentralization: beyond blockchain

A technology with great potential in AI governance is the blockchain (Werbach, 2018), the base technology of bitcoin. In the case of bitcoin, blockchain is oriented towards creating a record of financial transactions based on a distributed consensus protocol and, therefore, without a central authority (the database administrator or backing services). Today, these distributed networks solve the problem of the custody of information assets, without relying on a trusted third party. *Distributed ledger technologies* (DLT) offer a whole range of solutions for disintermediation in the processes of information management and value exchange (ISO). These solutions lead to new ways of generating and negotiating trust and tracing and linking operations. These facilitate new forms of managing digital identities that are consistent with the requirements of the *General Data Protection Regulation* (GDPR) (Voigt and Von dem Bussche, 2017) and the *Payment Services Directive* PSD-2 (Cortet, Rijks and Nijland, 2016), and with the specifications of the regulation on electronic identification and trust services in electronic transactions in the internal market (electronic identity recognition scheme (European Union)).

The design of new consensus protocols in asynchronous distributed systems configures an essential step in the definition of new decentralized trust management schemes. DLT and, particularly, blockchain are built upon consensus algorithms that enable state validation and replication without the participation of any sort of central authority. This opens up the possibility of establishing data governance schemes with different levels of traceability, linkability, transparency, and accountability. Among the most adopted consensus algorithms, the *proof-of-work* (PoW) is the most popular public permissionless blockchain of the kind applied in the case of bitcoin and ethereum. In addition, there exist plenty of alternatives to the PoW, such as the proof of stake or the practical byzantine fault tolerant. From a theoretical point of view, it must be established the relation between the safety and security of blockchain ecosystems and the properties of the underlying consensus protocols. Moreover, consensus protocols along with the peer-to-peer communication network and the cryptographic layer of a blockchain determine the on-chain governance of this technology. However, it is important to notice that the blockchain technology does not impose restrictions by itself on the behaviour of external agents. As a result, it is necessary to take into consideration different off-chain governance schemes to ensure the right balance between conflicting interests, business dynamics, and legal and normative requirements. Thus, the study of consensus algorithms and protocols should be conducted not only from the perspective of the theory of asynchronous distributed systems, but should also encompass contributions from other fields, such as game theory, corporate governance and complex networks. This type of analysis is especially required in the case of public permissionless blockchains, since the efficient and effective collaboration among stakeholders cannot be achieved unless efficient and effective rewarding policies are clearly and robustly defined.

Blockchain can play a key role in IT governance. Taking as a starting point the *International Standard for Corporate Governance of Information Technology* (ISO/EIC 38500), while also acknowledging the fundamental role ICTs plays in e-government, there is a need to envision new ways of incorporating IT design decisions into the complexities of corporate, social and political governance. The design, implementation and validation of different technological means to foster citizens' participation in decision making, public procurement monitoring, and other expressions of digital transparency and accountability, is necessary to promote trust in e-governance schemes and ensure active verification of the trustworthiness of trustees.

The following areas of research can be identified within the blockchain domain:

- Blockchain protocols to bear data curation processes and protect the data lifecycle.
- Advanced blockchain governance schemes to monitor AI-based decision making.
- Development of schemes based on blockchain for the correct protection of information assets and the privacy of sensitive user data (Finck, 2018).
- Use of blockchain as a distributed storage system for event capture in highly dynamic environments, such as IoT networks (Iglesias García, Diaz, and Arroyo, 2019).

In addition, blockchain scalability and usability should be tackled in each of the above research objectives, otherwise the technology and, in specific, its integration in the AI ecosystem will be critically hindered.

Privacy-utility models for data sharing and minimisation

Privacy is a very elusive, context-dependent concept, difficult to be expressed in formal terms. Privacy can be interpreted as confidentiality, as control, and as practice. One major challenge in the data deluge era is the development of a solid theoretical framework to define the possibilities and limitations of data exploitation models. In conjunction with traditional sanitization and pseudo-anonymization procedures, the differential privacy approach seems promising in delivering privacy respectful ML and DM. Along the development of cryptographic tools, this kind of proposal can be further extended by means of, for instance, homomorphic encryption and multiparty computation. Furthermore, trust management, specifically the definition of *dis-joint trust domains* is very relevant to promote trustworthiness across different models of data outsourcing, storage and computation. A number of proposals have been made to protect people's and information privacy. Some examples are: group signatures; network traffic anonymization through Tor; CryptDB for querying encrypted data using homomorphic encryption; or the theoretical framework of differential privacy (widely used in biomedical data processing). Indeed, users' privacy can be protected using *privacy enhancing technologies* (PET). PETs include cryptographic tools for managing anonymity and processing encrypted information, as well as solutions for information obfuscation. While PET technologies enable the protection of users' identities, they can also be abused by malicious actors, as it happens with bitcoin and Tor, generally associated with ransomware attacks and

payments (e.g. WannaCry, Petya or Ryuk attacks). These technologies are also associated to illegal market activities on the dark web. As any technology, PETs can be abused. Anonymity, which be used as a camouflage by criminals, is a cornerstone of electronic voting systems (Querejeta Azurmendi et al., 2019) and of privacy-respectful e-commerce platforms (Diaz et al., 2019). Undertaking an analysis of privacy-security tradeoffs (Arroyo et al., 2015; Diaz et al., 2018), and of privacy-surveillance tradeoffs (Ball et al., 2019) is necessary to enable to new forms of interventions of security agencies into encrypted communications and services. Striking a balance between the need to fight malicious activities and the safeguard of people's data protection rights, as required by the GDPR (Voigt and Von dem Bussche, 2017), is a huge challenge that requires interdisciplinary research, technological innovation, and policy development.

3.2. Applications

Data-driven threat models and containment in information systems

Over the last two decades, DM and ML techniques have increasingly been used to gather and interpret digital evidence, and enhance auditing and forensics capabilities on both networks and devices. *The protection of the security perimeter has evolved:* from the simple use of routers, firewalls and other network devices, we have witnessed the proliferation of behavioural, data-driven operational models focused on intrusion and malicious activity detection. Some examples are: intrusion detection systems, web application firewalls, and security information and event management systems. These solutions rely on BD analytics, thus requiring further improvements of DM, NLP and other ML techniques. Moreover, the elusive nature of evolving cyberthreats, and the heterogeneity of agents operating in cyberphysical systems, demand the integration of information sources placed outside corporate or government control. In the case of disinformation campaigns and other information operations, the study and characterization of social media and open data is key to leverage open source intelligent. There is a widespread demand for effective procedures to monitor the quality of information travelling online. As in previous cases, information gathering is an essential element that require to improve sensorization and event recording/treatment along the entire data lifecycle. The proper integration of hardware and software tools and methodologies is needed to ensure system and AI trustworthiness in line with the security-by-design principle.

AI can foster the creation of robust authentication, authorization, audit and accountability solutions. In this vein, a line of research that needs to be

expanded is the use of DL architectures for the extraction, processing and modelling of information from system and security logs (Chuvakin, Schmidt and Phillips, 2012), but also from open sources on security alerts and social networks (Bair et al., 2017; Sabottke et al., 2015; Khandpur et al., 2017). Recurrent long short term memory networks (Hochreiter and Jürgen Schmidhuber, 1997) are currently used, which are very useful for modelling sequences, although other techniques such as variational autoencoders (Kingma and Welling, 2013) or generative adversarial networks (Goodfellow et al., 2014) should also be taken into consideration.

Misinformation, fake news and cyber-attribution

Users are bombarded by the flow of digital information coming not only from newspapers, TV and radio, but also from digital media and user generated content. The new media environment impacts the way people form their opinions as well as the mechanisms leading to democratic consensus. Increasingly divergence of opinions and polarization reinforced by echo chambers is fragmenting public opinions and dividing citizens even when discussing common-sense daily issues.

Hybrid threats include methods of warfare, such as propaganda, deception, sabotage and other non-military tactics which have long been used to destabilise adversaries. Foreign governments may use social media platforms to influence public views and polarise political opinions by targeting opposite political candidates during election campaigns. Spain has been the target of various disinformation campaigns and information operations sponsored by foreign states. Historical cleavages and local conflicts extend the vulnerability surface of society and offer opportunities to certain groups or to external attackers to easily exploit people's vulnerability to misinformation at their advantage. Nowadays, it is possible to monitor information operations thanks to Twitter's election transparency initiative and the publication of thousands of tweets associated with propaganda or disinformation.

To better understand the relationship between fake news, post-truth, misinformation and disinformation it would be useful to clarify the terminology. *Disinformation* refers to motivated faking of news or other information as part of cyberwarfare. To have disinformation we need an attacker or malicious entity suing inaccurate information to intentionally deceive the audience. *Misinformation,* in contrast, only refers to the presence of inaccurate information. Disinformation campaigns are also called 'information operations' and

are mostly conceived to affect the stability of political institutions in another country. A promising line of research at the crossroad between computer science and social science explores the role social media platforms play in building political polarisation starting from affective polarisation and ideology. Understanding the drivers of political divisions and opinion polarisation is a fundamental research priority to safeguard our democracies from hybrid threats and information operations (Westwood et al., 2018).

Besides understanding how people consume and interpret information, another important line of research in computer science focuses on misinformation spreading in complex networks through human accounts and bots, and also on assessing information accuracy and source trustworthiness. Evaluating information quality and reliability on digital platforms demands the development of automated AI tools for the early detection and reaction to disinformation campaigns. The coexistence of official and unofficial sources creating and spreading contents online complicates the task of assessing the veracity of a piece of information. As today, the detection of misinfornation is based on singling out malicious spreading strategies but not so much on the content itself. Another challenge to be faced in the fake news detection area, is the recent interest in text-generating AI systems, raised partially due to the creation of the GPT-2 system by OpenAI. This fact manifests the urgent need to introduce real time and automatic fake news detection methods.

News curation and cyber-attribution are major challenges in the digital media environment. Instances of misinformation (inaccurate or contested information) coexist with hybrid threats and cyberwarfare. Individual users may unintentionally amplify the effects of specific information operations by keeping sharing malicious or inaccurate contents. While some users are more vulnerable to phishing attacks, other users are more vulnerable to disinformation campaigns promoting forms of reasoning rooted in conspiracy theories. Filter bubbles and radicalization dynamics can also reinforce the effects of misleading communications thought the by polarisation of public views.

Mis/disinformation bring in specific challenges related to: (a) the reliability of information sources and the identification and forensic attribution of the attacker; (b) the classification of misleading information and its tracking through complex-network analysis; and, (c) an in-depth understanding of peoples' vulnerability to mis/disinformation and the deployment of accurate and effective counter-deception communications.

Addressing these challenges require analysis of psychological mechanisms and public opinion reactions, combined with the deployment of detection tools capable of identifying malicious sources and misinformation spreading on a variety of channels, from chatbots to encrypted communication systems. Social media and instant messaging applications are constantly under scrutiny as likely misinformation channels. However, the identification of misleading information on encrypted channels entails controversial privacy-security tradeoffs. Current solutions try to incorporate a human-in-the-loop logic into the automatic detection of misleading information through NLP and complex network analysis. Additional measures to improve security without compromising privacy envision user reporting mechanisms and cyberawareness.

Cyber threats and cyber insurance

As discussed above, all kinds of organisations are critically impacted by cyber threats. *Risk analysis* is a fundamental methodology to help manage such issues. With it, organizations can assess the risks affecting their assets and what security controls should they implement. Numerous frameworks support cybersecurity risk management. Similarly, several compliance and control assessment frameworks, provide guidance on the implementation of cybersecurity best practices. They have many virtues; however, much remains to be done regarding risk analysis from a methodological point of view: a detailed study of the main approaches to cybersecurity risk management reveals that they often rely on risk matrices, with well documented shortcomings, potentially inducing suboptimal cybersecurity resource allocations.

In this context, a complementary way for dealing with cyber risks through risk transfer is emerging: cyber insurance products have been introduced in recent years. However, cyber insurance has yet to take off. Thus, all advances proposed, addressed towards improving smart cybersecurity risk management should help enhancing better grounds for cyber insurance adoption. Moroever, dynamic AI approaches to cyber risk management may lead to enhanced dynamic cyber insurance products.

Security, transparency and accountability in the Fintech sector

The transition into a cashless society and the implementation of open bank infrastructures entail a series of challenges. On this concern, it is of major relevance the required effort to achieve adequate tradeoffs between the protection of citizens' rights (e.g., in the context of GDPR and PSD-2) and compliance with fraud detection and prevention requirements. The convenient implementation

of know your customer and *digital onboarding* solutions by banks and other financial institutions is a must to meet the requirements of EU laws on anti money laundering and counter terrorism financing. The proliferation of cryptocurrencie and, specifically, of those enabling financial transactions in a (pseudo) anonymous way. The analysis of public permissionless blockchains is critical to identify and prosecute criminal activities, as it is the case of drug trafficking or cyberattacks as ransomware campaigns or the implementation of command and control systems targeted to access privilege information from institutions and organization. Network analysis and sociophysics provide a set of methodologies that pave the way to monitor and trace cryptoassets, and eventually enable the identification of illegitimate financial flows. NLP is another useful framework to correlate dynamics in underground fora in the deep web and capital flow by means of cryptocurrencies as bitcoin. Moreover, the aggregated analysis of open data and other cryptosignals is demanded to scaffold adequate governance schemes and accountability solutions in cryptoeconomy.

Secure computing on dependable systems

Lemmas such as "design for test" or "design for manufacturability" have recently been replaced by the novel "design for trust" imperative. To ensure the dependability of systems, both software and hardware and increase Spanish and European technological independence, a set of new methods supported by open hardware and software modules have been proposed to guarantee full protection of personal and confidential information. Specific hardware modules can increase the security of digital processing systems, while trusted and open *systems-on-chip* (SoC) platforms can increase European technological self-determination. Dependable systems built from the root of trust to the interface level are suitable for a wide range of applications; from tiny and ultra-low power internet-of-things devices, to the most advanced high performance computing systems. The open RISC-V initiative is an example of trustworthy hardware development meant to improve security, while reducing risks, and allowing the exploration of new micro-architectural extensions for specific application domains also enabling sustainable energy-efficient tools and methodologies.

Embedded systems are major components of the physical layer of any information system, and thus of any AI system. The hybrid (hardware/software) nature of embedded processors can be exploited to increase the overall level security of AI systems. In certain scenarios, there is an urge to update the management of cryptograhic keys and remote attestation to enhance conventional secure and trusted hardware architectures based on trusted platform

modules. For instance, these hardware modules are not affordable to be integrated in devices with limited resources. One option is to build hardware root-of-trust, also known as hardware anchor, with the inclusion of hardware dedicated modules. The *intrinsic digital identity* (ID) derived by hardware is used as an anti-counterfeit mechanism to detect the impersonation of devices (Martínez-Rodríguez et al., 2018). The idea is that the unique ID is used to regenerate cryptographic keys as many times as necessary without having to store them. Therefore, the underlying hardware circuitry is used as the basics to build high level cryptographic protocols running on software programs on the processors.

Another way to protect electronic devices is to monitor the system during its normal operation to detect eventual anomalies. The design of hardware modules that collect micro architectural information of the system is a mechanism to diagnose it (Tang, Sethumadhavan and Stolfo, 2014). Modern processors and SoCs include the so-called *performance monitoring units* (PMUs). This idea can be extended with the design of hardware dedicated modules to measure execution times and events using hardware/software co-design methodologies. Therefore, PMUs provide real-time feedback to diagnose bugs, identify anomalies, or bottlenecks during program execution. This information can be used as training set for AI techniques. Additionally, other hardware modules to test the hardware platform can be included to reinforce the decision-making accuracy of AI techniques. For instance, the inclusion of aging sensors that measure the performance degradation of circuitry over time. Or the integration of on-line testing techniques to guarantee the reliability of the output response during the generation of digital ID.

Furthermore, AI-based penetration testing can be carried out to achieve resistance against *reverse engineering* (RE) techniques and attacks. Actually, RE techniques of *integrated circuits* (ICs) can be applied with different security purposes. For instance, the certification of cryptographic algorithms implemented in hardware, the identification of hardware trojans inserted during the fabrication process, and the hacking of ICs to access sensitive information for forensic purposes (Quijada et al., 2018). Whereas the non-invasive *side channel attacks* (SCAs) exploit vulnerabilities on software or hardware implementations of cryptographic algorithms to recover the secret key (Hettwer, Gehrer and Güneysu, 2019), the use of methodologies based on AI techniques reduce the time required to complete RE and SCA procedures. In the case of SCA attacks, published works have demonstrated that DL based SCAs are very

efficient when targeting cryptographic implementations even protected with the common side-channel countermeasures (Maghrebi, 2019). In short, AI can be used to conduct RE on ICs with a wide range of security purposes as:

- Pattern and object identification. Logical block and security mechanism identification on IC surface.
- IC segmentation. IC decomposition into basic logical blocks.
- Intellectual property modules detection. Proprietary logical blocks identification using graphical convolutional networks.
- Image super-resolution. Use of low-quality images to reconstruct high-quality ones.
- Image restoration. Damaged image reconstruction using *generative adversarial networks* GANs.
- Power consumption patterns. Functional patterns identification using recurrent neural networks.
- Voltage contrast IC signals identification. Application of NLP procedures to pattern identification and functionality inference from IC signals and video analysis.
- IC Synthesis tools. Total or partial IC generation using GANs.
- Style Transfer. IC design and technology identification in order to emulate its functionality using computer-aided libraries for IC design.
- Adversarial attacks. Use of evolved DL models in IC reconstruction.

Dependable autonomous vehicles

Although the dream of the SAE for a level 5 *autonomous vehicle* (AV) could be envisioned as completely independent entities without any connection with the outside world, every current player envisions the car of the future with multiple interactions with its environment (*vehicle to environment* [V2E]), including road agents, infrastructure elements and service providers. This V2E communication has to guarantee distributed end-to-end security to ensure robustness against all types of attack vectors. The standards and security practices used today are not adequate for AVs. Functional safety standards like ISO 26262 and its evolution into SOTIF, information sharing like Auto-ISAC, software coding guidelines like MISRA, or overall safety scores like EURO NCAP do not solve the security issues that AVs will have to face in the short term. Upcoming AVs will have to consider securing each AI mechanism system itself, and communications between edge computing devices or vehicles with encryption and authentication mechanisms against attacks.

AI mechanisms for AVs

Cognition-inspired mechanisms and AI techniques are used to learn about the environment, and must detect unpredictable and harmful behavior, including hacking. In this context, the actions of an AI strategy may be limited by how it learns from its environment, how the learning is reinforced and how the exploitation dilemma is addressed. AVs could be exposed to malicious actors trying to manipulate the artificial perception and decision-making systems based on adversarial learning mechanisms that influence the training data for abnormal traffic detection.

Secure V2X communication

The majority of V2X messages are safety-related broadcast, with no restriction on which vehicles within range are allowed to read them. A security-related requirement for safety-related messages, is that senders should be trustworthy and accountable. Therefore, the removal/revocation of detected misbehaving participants from the V2X system is key. As a result, the requirement needs to be supported by appropriate technical measures and certification procedures for the underlying software and hardware.

The two previous aspects need to be seamlessly considered in critical AV functions. A good example of this is the enhanced and collaborative perception, which draw from AI to interpret a larger environment in a more dependable way, and needs consistency checks using ad-hoc V2X C-ITS messages, such as the Collective Perception Message. In any case, the involved components should be secure (i) by design (to ensure CIA); (ii) by default (with the confirmed capability to support these security properties at installation); and (iii) throughout their lifecycle.

Data protection by design and by default

Appropriate business, legal, and technical infrastructures are needed to transform into practice GDPR principles such as *privacy by default* and *by design* (Bender et al., 2014). They require to adopt the data minimization principle from the design-phase of a system and as a default property. According to art. 25 of the EU GDPR, people's personal information should be protected by design and by default. This implies two things. First, proper data minimization procedures should be designed in any information management solution including AI applications. Second, that data minimization principles should represent default features. Because of the key role that data plays in ML, adequate schemes to evaluate the quality and reliability of data are needed. There is a

need to improve the theory of data minimisation to enable utility models for AI consistent with citizens' expectations and data protection rights. It is also necessary to establish adequate data classification procedures to comply with current data protection policies. In addition, specific organizational and technical measures must be established to protect data according to their level of sensitivity. For instance, highly sensitive data should be protected using data loss prevention systems by means of advanced data-driven solutions.

Biomedical research and privacy risks

The availability of healthcare data in digital formats has increased over the years with the introduction in hospitals and clinical practice of digital devices and equipment (e.g. computed tomography, MRI, digital microscopes). The storage and availability of health data bring huge opportunities for biomedical research, but also pose considerable privacy risks. According to GDPR Art. 9, additional organisational and technical measures need to be established for the processing of "personal data revealing racial or ethnic origin, [...] genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation". These special categories of data are highly sensitive but also very valuable in economic and instrumental terms as they can improve dramatically medical research and facilitating the personalisation of treatments. However, health, genetic and biomedical data posits serious anonymisation challenges, which are accentuated by data variety, that is, from the fact that these data come in a variety of formats, which further complicate data management and the de-identification process. For instance, a 3D reconstruction including the face of a person can be easily obtained from MRI acquisition of a person's head. While it is possible to mitigate the privacy risk of re-identification by using brain extraction software or other de-facing techniques, there are instances for which no appropriate de-identification tools are available. Furthermore, brain structures can be matched using expectation–maximization algorithms, so once an adversary is in possession of one MRI view of the brain of a person it can identify other images of the same brain. Genomics and DNA data also present specific privacy protection challenges. The popularisation of consumer genomics services has increased reidentification risks for individuals as well as for ethnic groups. It has been estimated that a large proportion of US population can be identified starting just from a DNA sample even if these people never used a DNA service.

4. KEY CHALLENGING POINTS

We have identified the following major areas at the crossroads of cybersecurity and AI to which CSIC researchers will contribute.

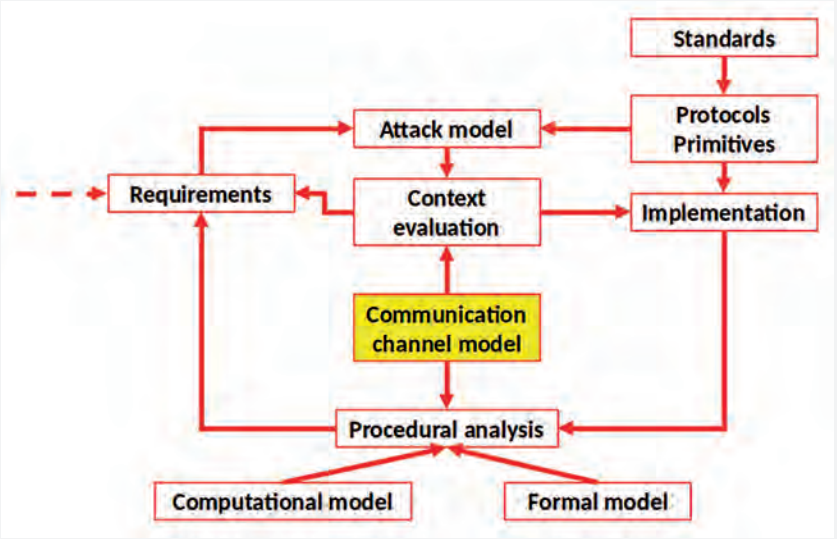
4.1. Fighting Misinformation About Science

The COVID-19 pandemic foregrounded the interconnection between the physical and the digital world and also demonstrated the extent to which the spread of biological and informational viruses bring new challenges to societies. The accelerated pace of scientific production and dissemination during the pandemic and the proliferation of fake news, rumors and hoaxes on those findings put institutions and official sources as well as fact-checking platforms under pressure. People in many countries were victims of misinformation about the contagion or the treatment of the disease. ***The COVID-19 pandemic made evident the need for better crisis management and preparedness strategies incorporating automatic end-to-end content verification tools.*** Acknowledging the capital role of human experts in information curation and verification, we are designing a solution incorporating a variety of data sources. We envision the integration of crowdsourcing approaches, based on knowledge extraction from open social networks and instant messaging apps, and expert knowledge obtained from scientific publications and technical standards. The definition of falsehood and truth is an epistemic and technical challenge, as well as establishing rules for drawing a line between freedom of expression and misinformation. Methodologies for assessing information quality and source reputation and trustworthiness also need to be established. Open questions remain on how to mitigate the damage that fake news and rumors can cause in society. There are several current examples of collective decisions taken under the pressure of misinformation. Collecting data from the different spreading media (online social networks, newspapers and traditional media) is fundamental to characterize and identify the spreading patterns. Modeling and AI can help to design new strategies to reduce the impact and to explore ways to expose the population to un-skewed information.

4.2. Imposing Security-by-Default Along the Computing System by Leveraging AI

Computing infrastructures are constant targets of cyber attacks aimed at gaining access to valuable assets such as data or computing power. To respond to advanced persistent threats, zero-day attacks, hybrid or other emerging threats, there is an ***urgent need to offer integral approaches to secure information***

FIGURE 8.2–Integral methodology for the secure software and hardware development life cycle.



assets and critical infrastructures. An important part of this holistic approach to threat modelling and containment is represented by methodologies to assess the reliability and trustworthiness of hardware and software components along their life cycle. Traditionally the *plan-do-check-act* (PDCA) methodology has been applied to ensure ICT systems quality. However, the complexity of hardware and software production, the dependence on third parties, and the variability of the application contexts make necessary to enlarge and perfect the PDCA methodology (Diaz et al., 2019) (see Figure 8.2). With such a goal, new AI tools are created to identify threats and evaluate security risks along the design and deployment of hardware and software products.

The expertise of CSIC researchers working on smart cybersecurity will be targeted at:

- ***The application of AI techniques*** to detect and avoid side-channel and fault injection attacks, through the adequate application of reverse engineering approaches;
- ***The deployment of computational and formal methods*** to characterise communication channels and attack models in information and communication systems;

- ***The development of procedures for comprehensive risk analysis,*** with a special consideration for automatic decision making and autonomous systems.

4.3. Creating a Formal Model for Adversarial Machine Learning

AML is an area of major importance in security and cybersecurity to protect systems which are increasingly being based on ML algorithms (Comiter, 2019). As we mentioned, there is a need for new paradigms going beyond standard game-theoretic approaches. One possibility is through *adversarial risk analysis* (ARA), which does not entail common knowledge conditions, being therefore much more realistic. As a byproduct, we obtain more robust algorithms. Here are some research challenges in this respect.

Robustifying ML algorithms through Bayesian ideas. Bayesian methods provide enhanced robustness in AML. (Ekin et al., 2019) shows how game theory solutions based on point estimates of preferences and beliefs may lead to unstable solutions, while ARA solutions tend to be more robust better acknowledging uncertainties. Thus, a promising challenge consists of developing efficient algorithms for approximate Bayesian inference with robustness guarantees. Indeed, there are several ways in which the Bayesian approach may increase security of ML systems. Regarding opponent modelling, an agent has uncertainty over her opponent's type initially; as information is gathered, she might be less uncertain about her model through Bayesian updating. Uncertainty over attacks in supervised models can also be considered to obtain a more robust version of adversarial training. Combining this approach with ARA opponent modelling may further increase robustness. Lastly, there are alternative approaches to achieve robustness in presence of outliers and perturbed observations, as through robust divergences for variational inference (Futami, Sato and Sugiyama, 2018). The robust Bayesian literature burgeoned in the period 1980-2000 (Rios Insua and Ruggeri, 2000). In particular, there has been relevant work in Bayesian likelihood robustness, referring to likelihood imprecision, reminiscent of the impact of attacks over data received. Note that Bayesian likelihood robustness focuses around random or imprecise perturbations and contaminations in contrast to the purposeful perturbations in AML.

Modelling and computational enhancements may enhance operational aspects. First, a core element in AML is the choice of the attacker perturbation domain. This is highly dependent on the nature of the data attacked. For example, in computer vision, a common choice is an A_p ball of certain radius

centered at the original input. These perturbations, imperceptible to the human eye, may not be representative of threats actually deployed. Thus, it is important to develop threat models that go beyond A_p norm assumptions. Moreover, it is important to deal with multiple agents in its variants (one defender vs. several attackers, several defenders vs. several attackers) including cases in which agents on one of the sides cooperate.

There is also a need for new algorithmic approaches. Exploring gradient-based techniques for bi-level optimization problems arising in AML is a fruitful line of research, (Naveiro and Ríos Insua, 2019). However their focus on white box attacks. On the other hand, Bayesian methods are also hard to scale to high dimensional problems or large datasets. Recent advances in accelerating stochastic gradient Markov chain Monte Carlo samplers are crucial to leverage the benefits of a Bayesian treatment. The ARA framework essentially goes through simulating from the attacker problem to forecast attacks and then optimize for the defender to find her optimal decision. This may be computationally demanding and we could explore single stage approaches to alleviate computations.

AML methods need to be developed. For example, there is a need to extend ACRA to to discriminative models and multi-class problems. Further research is also required in relation with attacks to unsupervised learning, for example through k-means clustering, autoregressive models and NLP which are just beginning to attract attention.

4.4. Safeguarding Privacy in the Era of Big Data and AI

Privacy-preserving techniques currently available have not resolved the usability-privacy tradeoff. In other words, ***preserving data privacy through obfuscation, de-indentification and encryption create some problems to data utility.*** These alternative techniques need to be complemented with strong communication security protocols and orchestrated into a comprehensive context-sensitive methodology. Identity management represents a fruitful strategy to foster privacy protection through adequate utility and privacy data models and the implementation of suitable anonymization and pseudoanonymization techniques. Starting from basic sanitization procedures (designed to eliminate quasi-identifiers), other methods based on attribute generalization can open the door to the deployment of forms of privacy-respectful querying rooted in *differential privacy* (Dwork, 2014). This can offer privacy protection, but could also reduce research data quality. Furthermore, it cannot be applied to unstructured

data types like medical images, although there exist some important contributions from the field of federated learning. To safeguard data quality, an alternative proposal is to build secure storage and analysis platforms for biomedical data, so called data safe havens, where researchers can analyse data, but cannot extract them. The problem this time is how to ensure platform security, user access control and auditing, network and physical security.

In the field of biomedical research, enabling GDPR-compliant data sharing of health records and genomics data is a key scientific and societal priority. Because of the ***intrinsic identifiable nature of genetic and biometric data, de-identification is an actual challenge.*** A mix of pseudonymisation techniques and access control procedures are usually adopted to protect this type of data, whose sensitivity is high both for individuals and groups. Racial and ethnic information presents risks to group privacy not just to individual privacy. BD adds further risks as it makes easier to identify specific groups.

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