Technology in Society 83 (2025) 102956

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

Unveiling the multifaceted concept of cognitive security: Trends, perspectives, and future challenges☆

Fran Casino *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Avinguda dels Països Catalans, 26, 43007, Tarragona, Spain
Information Management Systems Institute, Athena Research Centre, Artemidos 6, Marousi 15125, Greece

ARTICLE INFO

Keywords:
Cybersecurity
Cognitive security
Artificial intelligence
Human–computer interaction
Information systems

ABSTRACT

As a transversal concept tied to human evolution, security has increased its relevance at the same pace as development and digitisation. With the advancement of artificial intelligence (AI) and the sophistication of advanced persistent threats, the emerging paradigm of cognitive security (i.e., defined by some authors as the use of self-aware and adaptable AI with learning capabilities to detect and mitigate security threats) gains momentum. Nevertheless, cognitive security is a complex concept that requires a more granular description. In this article, we redefine cognitive security by first analysing the state of the art to derive the current state of practice and the definitions of cognitive security. Next, we expand the concept of cognitive security by analysing its multiple pillars, including learning theories, AI technologies, human–computer interactions, and the ethical and legal aspects impacting its development and implementation. The latter is crucial towards understanding cognitive security, providing insight into its potential and prerequisites towards its realisation while emphasising its multidisciplinary nature. In addition to such a description, we analyse the current challenges in three closely interconnected fields, namely cybersecurity, digital forensics, and digital investigations, to provide a taxonomy that can be used to assess the current challenges and limitations of cognitive security and understand its potential better. Finally, we propose future research directions, aiming to develop cognitive systems capable of continuous learning, adaptation, and ethical compliance in dynamic cybersecurity environments. Our findings highlight the role of cognitive computing systems in enhancing cybersecurity, discussing the integration of human cognition and AI for proactive and resilient security solutions.

1. Introduction

Delinquency and criminality have always been part of our societies. The concept of security has evolved throughout history by adopting different approaches and technologies, considering multiple contexts and concerns tied to human evolution, as summarised in Fig. 1. Nevertheless, the digitisation, among others, has enabled perpetrators to expand their outreach, automate their actions, and scale their operations, even on a global scale.

Nowadays, the advent of artificial intelligence (AI), leveraged by high computing capabilities, creates many opportunities and services for interconnected and dynamic systems. In this context, AI helps resource-intensive security operations by using technologies such as machine learning, pattern recognition, and natural language processing, which can ingest terabytes of unstructured data to enhance response times and expand the capacities of security operations. However, such a fast and continuous evolution has several drawbacks. One of them is the lack of security countermeasures (i.e., timely detection and mitigation mechanisms when a security incident occurs).

☆ This work was supported by the European Commission under the Horizon Europe Programme, as part of the projects SAFEHORIZON (Grant Agreement no. 101168562) and LAZARUS (Grant Agreement no. 101070303). This work was also supported by the European Union's Internal Security Fund as part of the ALUNA project (Grant Agreement no. 101084929). This work was partially supported by Ministerio de Ciencia, Innovación y Universidades, Gobierno de España (Agencia Estatal de Investigación, Fondo Europeo de Desarrollo Regional -FEDER-, European Union) under the research grant PID2021-127409OB-C33 CONDOR. Fran Casino was supported by the Spanish Ministry of Science and Innovation, Spain under the "Ramón y Cajal" programme (RYC2023-044857-I), and by AGAUR with the project ASCLEPIUS (2021SGR-00111). The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

* Correspondence to: Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Avinguda dels Països Catalans, 26, 43007, Tarragona, Spain.
E-mail address: franciscojose.casino@urv.cat.

https://doi.org/10.1016/j.techsoc.2025.102956
Received 20 May 2024; Received in revised form 23 April 2025; Accepted 28 May 2025
Available online 13 June 2025

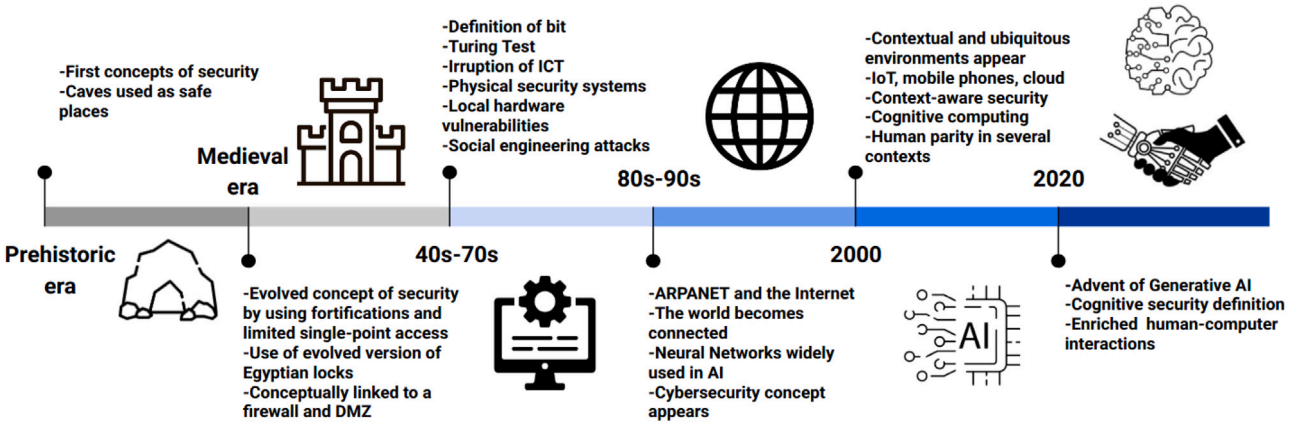

Fig. 1. Evolution of the technologies and approaches related to the security concept through history.

As cybersecurity evolves, so does cybercrime in terms of scale and sophistication (The European Union Agency for Cybersecurity (ENISA), 2023; United Nations International Computing Centre, 2023). For instance, criminal activity has transformed with the advent of 'Crime-as-a-Service' (CaaS) (Manky, 2013), presenting unprecedented challenges, transforming malware proliferation into a commoditised industry (Patsakis, Arroyo, & Casino, 2025; Wall, 2024). This evolution leverages service-based models, where complex tools and services are made available through an accessible, subscription-based framework that advanced persistent threat (APT) groups use. APTs are highly organised, well-resourced adversaries that execute prolonged and stealthy cyber operations, often for espionage, financial gain, or geopolitical influence. This commodification of cybercrime lowers the barrier to entry, allowing even less technically skilled threat actors to carry out sophisticated attacks while enabling APTs to scale and optimise their operations (Georgoulias, Pedersen, Falch, & Vasilomanolakis, 2023). Moreover, by leveraging outsourced capabilities, APT groups can obfuscate their origins, diversify attack vectors, and rapidly adapt to defensive measures, making them more elusive and persistent (Casino et al., 2022a).

In parallel, neither governments nor private companies are nowadays fully equipped with the expertise and resources to prevent and overcome cybercrime, as proved by its continuous growth, especially due to its dynamic and cross-jurisdictional nature (Casino et al., 2022b; Forum, 2023). The above can be attributed to the lack of unity of effort and the varying levels of technical expertise, which hinder uniform and robust countermeasures. Note that many of the APT groups are state-sponsored. As a result, cybercriminals find more possibilities and multiple attack vectors to exploit. Thus, the current landscape demands a more proactive and holistic approach to security.

One of the main aspects required by the current cybersecurity context is more sophisticated interactions, regardless of whether they are human-to-human, human-to-machine, or machine-to-machine (Wu et al., 2022). The need for systems that can operate autonomously and adapt to changing environments has led to the emergence of Agentic Artificial Intelligence (Agentic AI), which is defined by its ability to set complex goals in uncontrolled situations and pursue them through autonomous management of its resources (Acharya, Kuppan, & Divya, 2025). The latter is further enhanced by cognitive computing systems (Wang, 2002), which appear as a natural evolution of classical human-machine interactions (Gutierrez-Garcia & López-Neri, 2015). Cognitive computing systems simulate human capabilities to think, reason, learn, and adapt, covering the entire process of transforming contextual data into wisdom through adaptive processes and interactions. These systems can learn from data patterns, prior experience, and distributed knowledge (George, 2005), and adapt their actions using adaptive control mechanisms. In addition, the ability to use tools and semantic memory allows cognitive systems to retain contextual information and optimise ongoing tasks (Singh, Ehtesham, Kumar, & Khoei, 2024). This evolution has reached the milestone of such systems passing the Turing test (Biever, 2023).

As previously highlighted, human–computer interactions are evolving at an unprecedented pace, enabling AI systems that can become self-aware (Andrade & Torres, 2018) and learn from past experiences by using direct human inputs (Wang et al., 2023). When applied to the security context, cognitive systems pave the way for a paradigm change, namely, cognitive security. The latter requires different key intrinsic components to enable continuous learning and adaptation, enhancing their resilience.

In this work, we redefine the multidisciplinary concept of cognitive security (i.e., defined by some authors as using self-aware and adaptable AI with learning capabilities to detect and mitigate security threats) by considering its inherent dimensions and challenges. The current state of the art often frames cognitive security as an AI-driven cybersecurity paradigm, emphasising self-learning and adaptive intelligence to counteract cyber threats. However, such a technocentric perspective risks narrowing its conceptual scope by overlooking its cognitive, socio-political, and epistemological dimensions. Following Alvesson and Sandberg's problematisation approach (Alvesson & Sandberg, 2011), this article challenges key assumptions in existing cognitive security research, arguing that it is not merely an extension of AI-based cybersecurity but a distinct paradigm requiring interdisciplinary exploration. Thus, by analysing the current state of the art, we expand current definitions of cognitive security by pointing out its main dimensions and the requirements towards its realisation. Next, we provide a taxonomy of challenges according to the state of the art and our observations. Our analysis is further expanded by considering the interconnected challenges of cybersecurity, digital forensics (i.e., digital forensics is the science of identifying, acquiring, preserving, analysing, and reporting on digital evidence to establish facts in a criminal or civil investigation), cross-border investigations, and cognitive security. Finally, we discuss future research paths, some of which have yet to be unveiled. To the best of our knowledge, this is the first time that a survey on cognitive security provides such a multifaceted definition and a taxonomy of challenges while providing potential countermeasures to them, becoming a timely article to foster advancement in the field.

The rest of the article is organised as follows. Section 2 provides the research methodology and an analysis of the state of the art, and Section 3 describes the main dimensions of cognitive security. Section 4 provides a taxonomy and a discussion of the challenges of the current cybersecurity landscape, including cognitive security barriers. Section 5 is devoted to analysing the requirements towards the realisation of cognitive security, enabling technologies, and research paths that can solve present and future challenges. Finally, Section 6 concludes the article and identifies potential directions for future research.

Table 1
Summary of research questions and the corresponding sections devoted to answering them.

Research question	Objective	Discussion
RQ1: What are the fundamental components and defining characteristics of cognitive security?	The aim is to identify the core constructs, dimensions, and elements that constitute cognitive security, refining its multidisciplinary nature.	Sections 2, 3
RQ2: How do cognitive security systems interact with emerging cybersecurity technologies to mitigate advanced threats?	The goal of this question is to examine how cognitive security integrates with AI, human–computer interactions, and regulations to enhance security measures.	Sections 3, 4
RQ3: Why is the current cybersecurity landscape insufficient in addressing the growing sophistication of cyber threats, and how can cognitive security address these gaps?	This question focuses on the challenges and limitations of current cybersecurity methodologies and explores how cognitive security can bridge these gaps.	Sections 4, 5
RQ4: Who are the key stakeholders involved in the implementation and regulation of cognitive security systems, and how do their roles influence adoption and effectiveness?	The goal is to identify relevant actors (e.g., researchers, regulators, agencies, industry) and assess their influence on cognitive security adoption.	Sections 5, 6
RQ5: Where are cognitive security frameworks most applicable, and how do different environments (e.g., critical infrastructure, financial systems, law enforcement) impact their design and effectiveness?	We aim to determine the most relevant domains for the applicability of cognitive security and how current instruments, protocols, and contextual factors shape its realisation.	Sections 3, 5
RQ6: When should cognitive security mechanisms be deployed in cybersecurity workflows, and which strategies can be used to address the cybercrime fight in the near future?	This question aims to define optimal deployment strategies for cognitive security systems based on risk levels, attack types, and environmental conditions to ensure timely measures and adoption.	Sections 5, 6

2. Research methodology

Cognitive processes encompass the mental activities associated with acquiring, processing, understanding, and creating knowledge (Bayne et al., 2019). When these processes are integrated into artificial intelligence, we can transfer human cognitive processes to, e.g., machine learning algorithms, large language models. By adding humans in the knowledge and feedback loop (human-in-the-loop) (Dautenhahn, 1998; Mosqueira-Rey, Hernández-Pereira, Alonso-Ríos, Bobes-Bascarán, & Fernández-Leal, 2023; Wu et al., 2022), human reasoning is integrated into creating cognitive systems, creating continuous learning processes. Cognitive security is a paradigm change with implications from several perspectives, including human, workflow, and culture. Nevertheless, according to the literature, cognitive security is still in its infancy.

Our review methodology is based on the five steps of Denyer and Tranfield (2009) for a systematic literature review. More precisely, the steps are the following: (1) Define the scope of the review, (2) Define the research questions, (3) Search literature databases, (4) Apply inclusion and exclusion criteria, and (5) Synthesise and report the results of the literature analysis.

2.1. Defining the scope of the review

A systematic literature review follows standardised processes for searching, screening, analysing, and synthesising the available literature in a systematic and reproducible way (Tranfield, Denyer, & Smart, 2003). Systematic reviews help build a reliable knowledge base by aggregating information from a range of relevant studies (Tranfield et al., 2003). In addition, this article partially falls into the category of review for understanding (Rowe, 2014), as it synthesises the main challenges of cognitive security and proposes paths to solve them.

To further refine our investigation, we apply Whetten's framework (Whetten, 1989) in structuring six research questions addressing *What, How, Why, Who, Where,* and *When* applied to cognitive security. Such research questions are described in Table 1.

2.2. Search strategy

As previously stated, our overall survey process is based on several predefined research questions relevant to the cognitive security literature. To this end, we performed a systematic literature search without time constraints in January 2025. The main search engines used were Web of Science (WoS) and Scopus, which were used to locate all scientific-related literature due to their multidisciplinary coverage and scope (Pranckute, 2021). Thus, we queried WoS and Scopus using the following query:

```
TITLE-ABS-KEY (‘‘cognitive security’’ AND (‘‘AI’’ OR ‘‘artificial intelligence’’) AND ‘‘cybersecurity’’)
```

Note that we framed the search with the previous query so that additional studies can be found afterwards. It should be noted that the first bulk search query yielded 178 results after removing duplicates. The full article was retrieved and evaluated for relevance when a study's abstract was unavailable. Moreover, we retrieved the full text of all relevant articles. Given the broad selection of articles, we discovered additional studies using the so-called backward and forward snowball effect, which involved searching the references of critical articles and reports for additional citations (Vom Brocke et al., 2015). For instance, additional literature was discovered by manually searching the reference lists in several articles.

2.3. Applying inclusion and exclusion criteria

We evaluated the eligibility of the selected literature based on specific inclusion/exclusion criteria. First, we excluded all non-English articles. The next step included screening the selected articles (title and abstract reading). For the rest of the articles, we performed a full reading. It should be noted that several articles were excluded during the last two steps (title/abstract screening and full paper reading). Our exclusion criteria aimed to fulfil the scope of our search; thus, we excluded articles that did not name or discuss cognitive security in the context of AI and cybersecurity.

As summarised in Fig. 2, after collecting all relevant sources and applying our methodology, 28 research articles passed the title and abstract screening. Of these, six were discarded after a full review, leaving 22 research articles relevant to the scope of our article, which were complemented with 18 more found through the snowball search described above. Subsequently, the year-wise distribution of the 40 selected publications is depicted in Fig. 3. Note that these articles were used to compute the current number of publications in the field and showcase that it has received little attention, identify the current definitions of cognitive security, and extract the challenges of cognitive security, later analysed in Section 4.

2.4. Content analysis and reporting

We used qualitative analysis software for the thematic content analysis of the selected literature (MAXQDA) (maxqda). We applied narrative synthesis to classify the extracted data comprehensively, combining the qualitative findings of multiple studies. For example, we identified several key concepts inherent to cognitive security that have already been discussed in the literature, such as self-learning AI, which refers to AI capable of learning and adapting on its own by, e.g., training itself using unlabelled data. However, there is a lack of holistic solutions towards cognitive security. Nowadays, industry solutions apply AI to

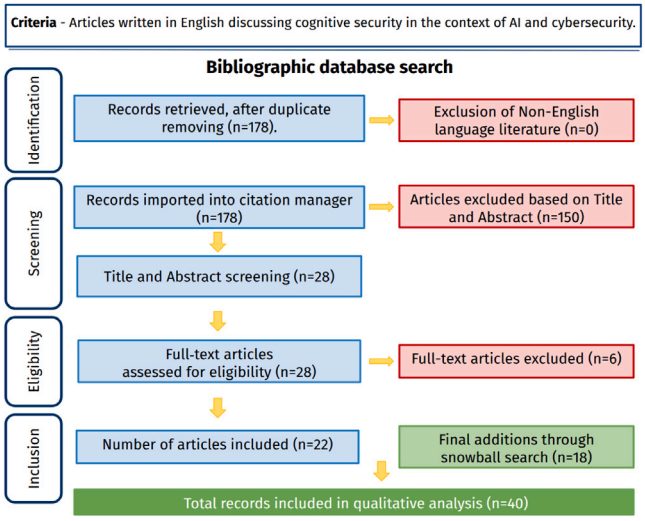

Fig. 2. Flowchart of the statistics for each step according to the inclusion criteria.

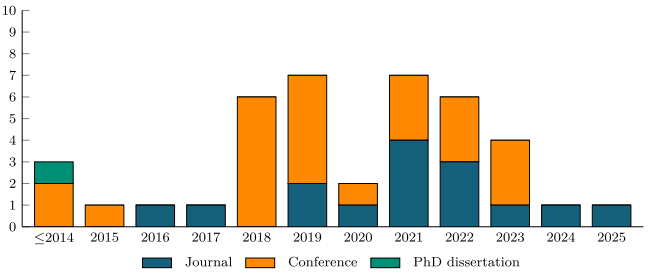

Fig. 3. Year-wise analysis of the selected literature. Note that most publications are conference papers (24), followed by journal articles (15).

constrained scenarios where learning capabilities are limited (Amann & James, 2015). Moreover, the term "cognitive" is commercially used in traditional AI systems with learning capabilities in specific, reduced contexts (Schroer, 2023), but still far from reaching fully functional cognitive solutions. In parallel, the term "cognitive security" is used by some authors to refer to the security of humans' cognitive processes towards threats (Doherty, 2023) (i.e., in particular, 21 articles that we later discarded refer to that topic). We distinguish cognitive security from these works and the concept of cognitive warfare (Claverie & du Cluzel, 2022; NATO, 2023), which refers to influencing the perception and behaviour of the population via disinformation campaigns and other strategies enabling the exploitation of people's cognitive biases.

The earliest discussions of cognitive security approaches appear in works focusing on personal devices and sensor networks. Greenstadt and Beal (2008) introduced the concept of continuous user verification through multi-modal biometrics, moving beyond static password-based authentication. Around the same time, Muraleedharan and Osadciw (2009) proposed a cognitive security protocol for vehicular sensor networks, emphasising swarm intelligence and real-time adaptivity to prevent denial-of-service attacks. Later, Sreekumaridevi (2011) explored how resource-constrained sensor networks can defend themselves through combined cognitive and swarm-based strategies, underscoring the balance between timely responses and limited computational resources.

As cognitive security gained traction, subsequent studies began formalising its principles through frameworks and architectures. Jagadeesan, Mc Bride, Gurbani, and Yang (2015) highlighted the potential of real-time analytics and self-adapting network functions in virtualised telecom environments, and Zheng, Moini, Lou, Hou, and Kawamoto (2016) shifted attention to the use of behavioural and contextual cues for continuous authentication in mobile networks. Yılmaz, Güvenkaya, Furqan, Köse, and Arslan (2017) addressed cognitive security at the physical layer, supporting adaptive radio strategies against eavesdropping or jamming. These concepts were further refined in studies that emphasise self-awareness and broader frameworks, such as Andrade and Torres (2018) and Andrade, Torres, and Flores (2018), who discussed holistic cognition encompassing real-time monitoring, risk management, and reduced analyst workload. In parallel, Silva and Hernández-Alvarez (2017), and Andrade, Torres, and Tello-Oquendo (2018) established concrete applications of cognitive and big-data-driven techniques to detect threats like ransomware and enhance security analysts' decision-making processes in large-scale infrastructures.

As a more general overview of cognitive security and its applications, Sreedevi, Harshitha, Sugumaran, and Shankar (2022) provide a glimpse of how cognitive computing is applied to domains such as healthcare, cybersecurity, big data, and IoT. While authors do not formally define cognitive security, their work underscores the growing significance of leveraging artificial intelligence techniques to manage large data volumes towards adaptive AI-driven defence mechanisms. Several works have placed even greater emphasis on developing holistic frameworks and applying these to diverse environments. Andrade and Yoo (2019) define cognitive security at the intersection of cognitive science and cybersecurity operations, underscoring the value of situational awareness and human decision-making processes. Their study highlights the need for analysts to make sense of large volumes of ambiguous data, emphasising collaboration, automation, and an evolving adversarial focus as core elements of next-generation cyber defence. In a related effort, Andrade, Torres, and Cadena (2019) propose integrating cognitive techniques into the entire incident management lifecycle, resulting in a more adaptive and automated response capability. This approach ties policy or compliance considerations into an AI-based analytical framework, enabling security teams to operate more efficiently under conditions of uncertainty. In healthcare-oriented cyber–physical systems, Abie (2019) argues that simulating human cognitive processes can address the advanced security needs of IoT-driven health ecosystems. By leveraging layered architectures that incorporate AI-based prediction, dynamic risk assessment, and policy enforcement, the proposed approach underscores both the urgency and the feasibility of cognitive defence strategies in sensitive environments where patient privacy and safety are at stake. The work of Cinque, Cotroneo, and Pecchia (2019) discusses combining SIEM-like systems with comprehensive analytics, building towards what they term "cognitive security defence". Their emphasis lies in automated situational awareness and data correlation to reduce the workload on human analysts and accelerate threat detection. Similarly, Andrade, Fuertes, Cazares, Ortiz-Garcés, and Navas (2022) refine and broaden this concept by presenting a cognitive cybersecurity model that integrates user behaviour, adversarial focus, and advanced analytics (e.g., text mining). Their exploration highlights a persistent research gap in measuring cognitive load and achieving real-time data integration. Jiang and Atif (2021) propose an ensemble-based machine-learning architecture to unify diverse security data repositories. Their selective ensemble approach shows measurable gains in predictive accuracy, showing how an intelligent fusion of AI techniques and human insight may strengthen real-world security analytics. Tariq, Ahanger, Nusir, and Ibrahim (2021) outline a co-design framework based on pervasive computational intelligence. This research shows how cognitive security concepts, including resource-aware AI, adaptive threat detection, and real-time data processing, can be successfully combined with rigorous industrial IoT constraints. Devadarshini et al. (2023) focus on leveraging machine learning alongside cognitive defence approaches to predict cyber threats in enterprise environments. Their work reinforces the shift towards proactive security, where threat prediction and resource-constrained AI solutions reduce vulnerabilities before they appear.

Several recent efforts highlight cognitive security's practical impact across diverse domains. In the context of Software-Defined Networking (SDN), Kavitha, Priya, and India (2019) present a machine-learning-driven intrusion recognition strategy, while El-Sayed, Toony, Alqahtani, Alginahi, and Said (2025) extend this idea to software-defined IoT by adopting a P4-powered architecture with multi-controller scalability. Focusing on mobile-centric security threats, Al-Kadhimi, Singh, and Khalid (2023) systematically review AI-based detection of APTs, whereas Martínez Santander, Yoo, and Moreno (2018) propose a cognitive patterns classifier for web-based attacks by combining honeypots and machine learning. In terms of large-scale IoT authentication, Chen, Yang, and Chou (2021) propose a cognitive gateway for trust evaluation, and Demertzis and Iliadis (2023) present a self-learning neural architecture employing adversarial training to enhance resilience. There have also been works related to smart grids, such as the one proposed by Sen et al. (2023) in which authors explore synthetic cyberattack data generation for ML-based IDS in smart grids, and in Butakova, Chernov, and Shevchuk (2019), where authors use distributed reasoning for situational awareness in critical infrastructure. In other contexts, Benzaid, Taleb, and Song (2022) propose a hierarchical AI-based security management framework for next-generation mobile networks, and Jayaganesh and Parvees (2022) propose the use of cognitive intelligence to mitigate identity threats in web applications.

Additional studies refine the notion of cognition in more specialised settings, integrating concepts such as adaptive decision-making and human-aware threat modelling. Chouhan, Chen, Hussain, and Beard (2021) apply semantic-rich situation calculus to automate cyberattack responses in IoT environments, leveraging knowledge-driven representations to determine the right mitigation actions. Lakhdhar, Rekhis, and Sabir (2020) propose a game-theoretic model that balances the cost of enhanced forensics with the benefits of preventing stealthy attacks, showcasing how cognitive security measures can promote adaptive strategies in adversarial contexts. Highlighting the human factor, Køien (2021) discusses the role of augmented cognition in accelerating threat detection, specifically in derailing the cyber kill chain, and Tantar, Tantar, Kantor, and Engel (2018) emphasise anomaly detection powered by cognitive techniques in SDN environments. At the optical layer, Furdek et al. (2020) illustrate how machine learning can detect real-time jamming or fibre-tapping attacks. Ogiela (2021) complements the previous perspective by examining cognitive cryptography for advanced pattern recognition and semantic inference.

Real-time anomaly detection is a critical aspect of cognitive security, and thus, several authors focus on that. Al Amin (2022) explores log-stream classification under dynamic adversarial conditions. Milosevic et al. (2022) introduce BACS, a deep learning-based suite for anomaly detection tailored to edge-fog-cloud deployments. Prabavathy and Supriya (2021) explore the use of fog nodes for faster detection of IoT threats by simulating real-time behaviour. Finally, Garcés, Cazares, and Andrade (2019) showcase cognitive security's role in phishing defence through automated URL classification.

While the above works discuss cognitive security, its benefits and applications, we identified some authors who have addressed the definition of cognitive security in the past. Andrade et al. consider cognitive security a procedural enhancement of both humans and machines, by leveraging AI with self-awareness capabilities (Andrade et al., 2022; Andrade & Torres, 2018; Andrade, Torres, & Tello-Oquendo, 2018; Andrade & Yoo, 2019). In general, several authors such as Al-Kadhimi et al. (2023), Chen et al. (2021), Prabavathy and Supriya (2021), Silva and Hernández-Alvarez (2017), and Garcés et al. (2019) highlight concepts such as awareness and reactive reasoning as pillars of cognitive security. Other authors such as Chouhan et al. (2021) and Jiang and Atif (2021) provide a special focus on human processes and computer interactions as relevant aspects of cognitive security systems. Greenstadt and Beal (2008) and Zheng et al. (2016) focus on behavioural and environmental aspects tied to real-time, as relevant pillars of proactive cognitive systems. Jagadeesan et al. (2015), and Sreekumaridevi (2011) suggest that data analytics and self-adaptive learning are key enablers of cognitive security. Finally, Huang and Zhu (2023) describe cognitive security as a method to enhance the capabilities of cybersecurity analysis through AI and behavioural analysis.

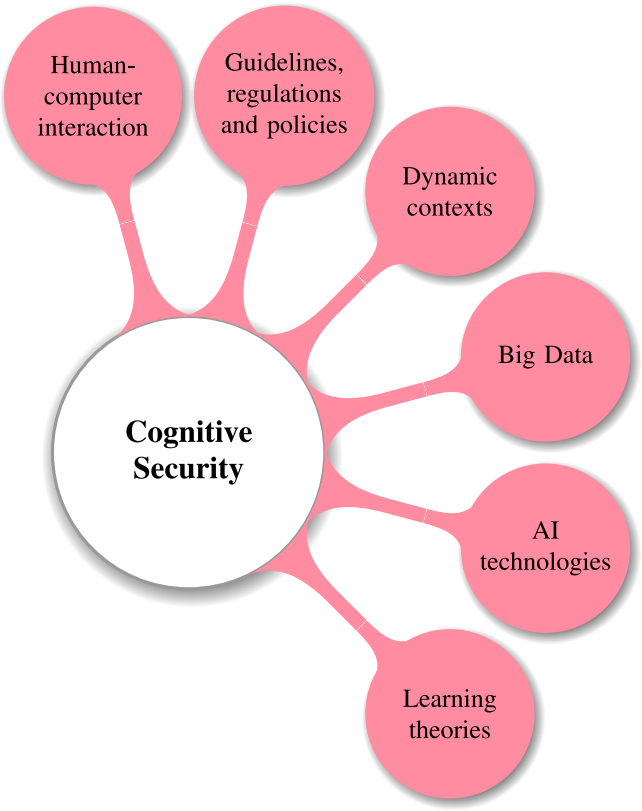

Fig. 4. Main pillars of cognitive security.

Nevertheless, in general, the ethical and explainability aspects of AI, which are gaining momentum, are barely highlighted, human-machine interactions are discussed at a very high level, and the learning theories behind mimicking human processes are recalled only a few times. Thus, we propose redefining the cognitive security concept by identifying and highlighting its core components in the following section.

3. Towards cognitive security and its dimensions

Given the previous context, cognitive security is often referred to as the application of self-learning and self-aware artificial intelligence technologies to detect and mitigate security threats. However, this article considers a more granular approach, identifying and describing cognitive security as a multidimensional concept. Therefore, we fine-grain the multi-faceted concept of cognitive security (see Fig. 4) by considering six main pillars, drawing on the strengths of artificial intelligence, human cognition, ethics and regulations, human–computer interactions, and cybersecurity to forge a robust defence against threats. The latter highlights the multidisciplinary nature of cognitive security and the necessity to collaborate and interrelate different expertise and backgrounds towards its realisation.

Learning theories: Historically, the evolution of learning theories (Ertmer & Newby, 1993) has seen a progression from Behaviourism, which focuses on observable behaviours and their responses to specific stimuli, to Cognitivism, which emphasises the role of mental processes in understanding and acquiring knowledge. Constructivism further advanced this understanding by positing that learners construct knowledge by integrating new information into their existing cognitive structures. The advent of Connectivism, as proposed by

George Siemens, marked a significant shift in the understanding of learning in the digital age by highlighting the relevance of networks and connections, where knowledge potentially resides (George, 2005). Nevertheless, unlike traditional learning theories primarily focusing on human-centric learning, cognitive systems, particularly in the cybersecurity domain, emphasise a symbiotic relationship where humans and machines acquire, process, and refine knowledge through interactions (Şahin, 2012). This bi-directional flow of information and learning ensures that the system is continuously updated, adapting to new threats and challenges. Within such a connected landscape, the ability to locate, organise, evaluate, and analyse information via digital technologies becomes a pre-condition for meaningful participation. More concretely, digital literacy is defined as the ability to locate, organise, understand, evaluate and analyse information using digital technology (Şahin, 2024). With rapid advances in technology, lack of digital literacy appears to be a problem, especially for certain groups of people who are older or digitally illiterate. Thus, pairing digital literacy initiatives with explainability, openness, and interpretability in algorithmic design makes the underlying reasoning pathways visible both to experts and to less skilled users. In terms of cultural bias, sociocultural constructivism (Jaramillo, 1996) concludes that learning processes differ across cultures. Consequently, data need to be curated and audited specifically to detect and mitigate cultural or regional biases, including explainable descriptions of the underlying reasoning of the models to foster robust interpretations. Techniques such as representational parity, group-based performance metrics, and post-processing fairness restrictions help ensure that the trained models protect users equitably, especially those from under-represented communities, while maintaining operational effectiveness and usability. Paired with the latter, ethical and policy guidelines need to be embedded into learning loops and policy controls, ensuring robustness and regulatory adherence (Benraouane, 2024; Jackson, 2024). By converting high-level norms into machine-readable safeguard constraints, cognitive systems internalise societal expectations and dynamically adapt to regulatory changes.

As our understanding of biological and artificial intelligence deepens, these systems evolve to emulate learning frameworks of other life forms or even entirely novel intelligent structures. Such advancements could lead to cognitive systems that can address and resolve specific cybersecurity challenges more efficiently than humans, embedding digital literacy support, cultural fairness, and enforceable ethical constraints directly into the learning procedures.

AI technologies: Advanced AI methods are pivotal for evolving cognitive security. In this sense, these methods should possess several essential characteristics to realise cognitive security systems. For instance, the ability to learn in a distributed manner by using, e.g., federated learning approaches (Banabilah, Aloqaily, Alsayed, Malik, & Jararweh, 2022; Mothukuri et al., 2021) is crucial in dynamic environments where endpoints have varying hardware capabilities and contexts with changing behaviours (Xu, Qu, Xiang, & Gao, 2023). Transparency and interpretability are further key components of these AI technologies. In an age where ethical considerations and legal mandates demand clarity in AI decision-making, it is imperative that these systems are not "black-boxed" (Bommasani, Klyman, et al., 2023). They should offer a level of explainability that aligns with these emerging ethical and legal frameworks, ensuring that stakeholders can understand and trust the decisions made by AI, so that malicious uses and unwanted risks for downstream applications (Khodabandehloo, Riboni, & Alimohammadi, 2021; Mozes et al., 2023; Zhao et al., 2024) can be detected more effectively. Furthermore, the nature of cognitive security systems lies in their ability to interact seamlessly with humans. While traditional machine learning models can be retrained, cognitive systems elevate this by being self-aware of their learning needs. Instead of being unidirectionally instructed, they should be able to autonomously generate hypotheses and identify when they require retraining or upgrading. This notion of self-awareness in AI technologies is becoming more tangible with the emergence of large language models (LLMs), which have redefined the boundaries of human–computer interaction. Given the contextual understanding of an LLM, prompts can be strategically employed to foster an interaction in which both humans and computers mutually enrich their knowledge base, which is, in turn, one of the primary purposes of cognitive systems. Furthermore, Retrieval-Augmented Generation (RAG) techniques enable LLM-based agents to dynamically retrieve external knowledge, improving the relevance and context of their responses, especially useful in real-time decision-making systems (Acharya et al., 2025). Beyond LLMs and RAG, cognitive security could significantly benefit from applying consciousness-inspired computational solutions, such as artificial neural networks and deep learning approaches derived from computational neuroscience (Guerrero, Castillo, Arango-Lopez, & Moreira, 2025). As these techniques model neural interactions similar to human cognition, they could facilitate more robust and dynamic threat recognition and response capabilities.

Big Data: Data are the main background upon which cognitive systems are built and defined. Data can take the form of structured designs like databases and unstructured forms like text documents, social media posts, images, log files, graphs, or an amalgamation of different formats, hindering their processing to extract insights about security threats and malicious actors. Processing big data necessitates sophisticated mechanisms to extract meaningful patterns and actionable intelligence in a reasonable amount of time. However, in addition to the amount of data, the provenance and quality of these data play a pivotal role in determining the efficacy of cognitive security systems. As outlined in Andrade and Yoo (2019), several quality metrics need to be considered:

- **Consistency**: Ensuring that data do not have inherent contradictions and biases and is coherent across different information sources.
- **Accuracy**: The data should be free from errors and accurately represent the real-world scenario it is derived from.
- **Completeness**: No crucial information should be missing, and the data should provide a comprehensive contextual view.
- **Auditability**: The ability to trace back the data to their source, ensuring its authenticity and reliability.
- **Orderliness**: Data should be organised and use standardised notation and representation to facilitate easy retrieval and processing.

Although obtaining quality data is often challenging, its impact on cognitive systems dramatically influences their performance in real-world scenarios, both in terms of accuracy and in terms of potential bias, as highlighted by the cognitive warfare concept (Bagdasaryan & Shmatikov, 2022; NATO, 2023). Thus, a system trained on high-quality data will more likely detect threats accurately and respond effectively and in an unbiased manner. The latter, paired with the capability to handle high volumes of such diverse data types, ensures that the system remains robust and adaptable to cybersecurity threats.

Dynamic contexts: The mutability of contexts and their threats is inherent in cybersecurity. A cognitive security system should be designed to adapt to these changes, exhibiting proactive measures derived from its learning and evolution capabilities. Dynamic contexts may enclose different sets of variables, from data inputs that could dynamically change according to random or predefined setups to evolving threat models that adapt to the latest security measures, as in APTs. Additionally, a cognitive security system can change its decision-making style depending on uncertainty or resource availability. Also, the use of multimodal learning, which integrates information from multiple sources, could improve the versatility of cognitive security systems in the operating context, such as autonomous vehicles operating in complex situations. In this scenario, resiliency is crucial and is deployed through anticipation, adaptability, self-learning capabilities, and the mutability of the underlying cognitive system, ensuring that it remains a step ahead.

Guidelines, Regulations and Policies: Legal, ethical, and regulatory challenges co-exist with the rapidly evolving digital landscape. However, such evolution is not occurring at the same pace. A foundational aspect of the latter is the establishment of verifiable methodologies for training AI systems. The decisions made by these systems, which can have far-reaching consequences, should be rooted in ethical principles, auditable data sources, fairness, transparency, and accountability. In this regard, explainability is closely related to the ethical deployment of AI by enabling verifiable analysis of AI outcomes given specific inputs and ensuring the integrity of the methodologies employed. However, ethical considerations alone are not enough (Jarrahi, 2018; Mosqueira-Rey et al., 2023). They must be translated into tangible guidelines, regulations, and policies, enabling legal systems to seamlessly integrate and adapt to these novel AI-driven technologies (Liwång, 2022). Such a framework mitigates potential threats and vulnerabilities and streamlines investigations when breaches occur or when bureaucratic and methodological drawbacks arise (Casino et al., 2022b). Thus, proactive measures, characterised by fast and timely responses to threats, necessitate the development of AI-based methods that are forensically robust and reproducible, ethically grounded and legally compliant, enabling efficient and timely prosecution.

Human–computer interactions: The symbiotic relationship between humans and machines is central to the evolution of cognitive security. While machines have proved their capabilities compared with humans (e.g., by achieving human parity in several contexts Dupoux, 2018; Mosqueira-Rey et al., 2023, and by being capable of processing vast amounts of data), the human factor remains irreplaceable, especially in scenarios that demand judgement and oversight (i.e., note the excessive agency problem OWASP Foundation, 2023). One of the foundational aspects of this relationship is readability. Systems must be designed to allow humans to interpret and understand their operations. Readability also requires that data-related tasks allow for bi-directional processing so that humans and machines can adapt and learn from each other's inputs. This is crucial in critical infrastructures, specific industrial procedures, and healthcare, where delicate decisions occur. In such scenarios, while machines can analyse and suggest, the final decision should be made by humans (Jarrahi, 2018), thus requiring readability-enabled systems. While closely related to readability, explainability focuses on the structural complexity of AI models and methods (Gunning et al., 2019). Techniques related to feature engineering and analysis, model design, and further benchmarks to evaluate explainability are necessary to shed light on the "black-boxed" nature of AI, especially in critical contexts (Khodabandehloo et al., 2021). For instance, several metrics are devoted to measuring explainability in LLMs according to fine-tuning-based and prompting-based paradigms (Zhao et al., 2024). Finally, in addition to the aforementioned strategies, human–computer interactions need to go a step beyond and adapt to the characteristics of the users. As each user has specific cultural, language, literacy, and cognitive barriers (i.e., these can be linked with specific learning theories to create personalised and inclusive systems), interactions need to be adapted at different levels to ensure understanding, development, and adoption.

While cognitive security has been broadly discussed in terms of self-learning AI applied to cybersecurity threats, its defining components remain an evolving area of study. The multidimensional nature of cognitive security necessitates a structured analysis of its intrinsic features, distinguishing it from conventional AI-based security paradigms. These components include adaptive AI mechanisms, context-aware decision-making, ethical compliance frameworks, and resilient human–computer interactions. Unlike traditional cybersecurity approaches that rely on pre-defined rule sets, cognitive security systems are envisioned as autonomous, self-evolving entities capable of learning from new threats, dynamically adjusting their defence mechanisms, and enhancing decision-making in real-time environments. Thus, based on the above pillars, we redefine cognitive security by expanding it as follows: *Cognitive security is a multidisciplinary paradigm involving self-aware and self-learning artificial intelligence systems, grounded in advanced learning theories and enriched human–computer interactions, which proactively detect, mitigate, and adapt to evolving cybersecurity threats while ensuring ethical compliance, transparency, and explainability*.

4. Open issues and challenges

To systematically analyse the open issues and challenges in cognitive security, as well as in the cross-domain cybersecurity landscape, we develop a taxonomy that classifies and organises these challenges into distinct categories. This taxonomy is based on an extensive review of the literature, including academic papers, cybersecurity threat reports, and policy documents, as well as our analysis of emerging trends in cognitive security. First, we present the challenges related to cognitive security, systematically classifying them into categories that reflect their interdisciplinary nature, in Section 4.1. Next, we collect the challenges related to digital forensics, cybercrime prosecution and harmonisation, and cybersecurity, and provide a structured visualisation of these challenges, grouping them into five domains: *Data Management & Acquisition*, *Forensic Methodologies*, *Jurisdictional Constraints*, *Cooperation & Interoperability*, and *Advanced Threats*. These five domains are used to map the current landscape and its impact on cognitive security in Section 4.2.

4.1. Cognitive security challenges

Cognitive security is a multidisciplinary approach that requires technologies to evolve and adapt to this new paradigm. In this section, we analyse the collected literature as reported in Section 2, and extract the challenges of cognitive security (Andrade & Yoo, 2019; Huang, Zheng, Shang, & Xue, 2023; Rajtmajer & Susser, 2020). Next, we abstract such challenges to create a set of categories. Finally, we add our perceptions and knowledge to perform an in-depth analysis of the current state of practice, which is summarised in the following paragraphs.

Data inputs and quality: Ensuring ethical, high-quality, and realistic data is crucial in cognitive systems to guarantee the trustworthiness of cognitive technologies, especially when considering the context-aware nature of some scenarios. Information overload is a challenge in itself, which is augmented by the proliferation of misinformation and disinformation. In this context, AI is a potential objective of such a threat, which could be used maliciously to manipulate facts, beliefs, and behaviour, and to hinder such systems (e.g., as a denial of service attack by providing incorrect or biased responses, affecting the usability of the system). The latter is also related to ensuring effective updating and upgrading mechanisms to solve these issues rapidly and providing users with properly trained models, avoiding potentially harmful, outdated ones.

Adversarial attacks: AI models can be intentionally manipulated using specially crafted input data, leading them to produce incorrect or biased results. The process typically involves introducing subtle, almost imperceptible changes to inputs, such as images or text, exploiting the inherent vulnerabilities of the model, and causing it to make erroneous predictions. Some of the most notable use cases of adversarial attacks span across domains such as image classification, where images are manipulated to mislead models affecting sectors like autonomous driving and medical imaging; text generation, where deceptive text can trick sentiment analysis tools or produce biased responses; and malware evasion, where malware is designed to bypass machine learning-based detection systems (Casino et al., 2022a). Moreover, conducting effective feature engineering with domain expertise is another way to improve detection accuracy (Guo, 2023). Attackers may circumvent detection if the attack features are similar to legitimate ones. Thus, it is essential to integrate domain experts' knowledge into the feature engineering process to ensure the new attacks will reveal

Table 2
Cross-domain mapping of the challenges and threats of digital forensics (DF), cybercrime prosecution and harmonisation (CH), cybersecurity (CY) and the corresponding dimensions relevant to cognitive security. Checkmarks indicate the relevance of each cognitive security dimension to each challenge category.

DF	CH	CY	Category	Cognitive Security Dimension					
				Learning theories	AI technologies	Big Data	Dynamic contexts	Regulations and policies	Human–computer interactions
•	•	•	Data management and acquisition			✓		✓	✓
•			Forensic methodologies and standards	✓	✓		✓	✓	
•	•	•	Jurisdictional, legal and ethical constraints					✓	
•	•	•	Cooperation and interoperability					✓	✓
•		•	Advanced threats and technological vulnerabilities	✓	✓	✓	✓		✓

themselves in the designated feature space. By addressing these vulnerabilities, there is potential to refine model training, enhance design, and strengthen overall security, ensuring the trustworthiness and reliability of cognitive security systems.

Human influence and cognitive bias risks: Humans possess inherent cognitive biases resulting from perceptual shortcuts, affecting processes like search, understanding, selection, and memory. The cognitive domain is transitioning from being centred around the "biological brain" to a fusion of the "biological brain" and the "digital brain". This evolution impacts human perception, understanding, learning, and decision-making, especially when models are trained without considering context constraints such as cultural heritage. The latter can be exploited by governments, media, and others to push cognitive warfare, hindering the security of cognitive and behavioural processes. Thus, over-reliance on AI for cognitive tasks may diminish human autonomy in the long term and create further issues that must be considered.

Lack of generalisation and abstraction: While related to the previous challenge, this one focuses on AI models and their inherent limitations when training with specific data sources, language models, or particular systems, which hinders generalisation across multiple contexts (Triguero, Molina, Poyatos, Del Ser, & Herrera, 2024). In this regard, it is crucial to leverage the proper mechanisms to improve the adaptability of cognitive security systems in different environments.

Dynamic digital environments and constrained contexts: Software updates, device upgrades, and new applications introduce potential vulnerabilities continuously, in addition to the potential exploitation of zero-day vulnerabilities (Guo, 2023). Despite the continuous evolution of cybersecurity methodologies, existing frameworks struggle to counteract the scale, complexity, and automation of emerging cyber threats. One primary gap is the reactive nature of conventional cybersecurity strategies, which often rely on signature-based detection mechanisms that fail against zero-day exploits and AI-driven attacks. Cognitive security addresses these limitations by incorporating continuous learning algorithms, behavioural anomaly detection, and predictive analytics, allowing for proactive defence strategies. Moreover, cognitive security introduces a hybrid human-AI collaboration model, ensuring that cybersecurity operations benefit from both machine-driven efficiency and human interpretability in decision-making. Thus, cognitive security systems should be able to provide effective and adaptable solutions that comply with, e.g., regulations and explainability. This reinforces the aim of producing standardised and robust approaches towards more resilient systems.

Decision-making and analysis of risks: Current cognitive systems have yet to fully mature in decision-making and risk analysis, especially in critical contexts where incorrect decisions can be fatal. Moreover, this decision-making can be applied to socio-economic factors, culture, political climates, and individual behaviours (Burton, 2023; Echterhoff, Liu, Alessa, McAuley, & He, 2024). Hence, nowadays, human intervention remains essential for comprehensive risk analysis and final decision-making.

Lack of guidelines and regulations: The lack of procedural guidelines is not a particular issue of cognitive systems but a generalised concern for novel devices and technologies (e.g., drones, novel cognitive-based humanoids, generative AI). As the discussion regarding the necessity of regulations is a trend (Hacker, Engel, & Mauer, 2023; Meskó & Topol, 2023), and regulations towards generative AI are starting to appear (Engler, 2023), cognitive systems, and in particular, cognitive security, require regulations that go beyond AI. This also includes methodologies towards data collection and curation, data processing, AI models and their explainability, the learning capabilities of cognitive AI, as well as other aspects coming from the different dimensions of cognitive security, which require further analysis and research to update current regulations holistically. Cognitive systems should be designed to comply with forensic readiness strategies and standards, enabling faster investigation and prosecution should they have been used maliciously. The latter is paired with the increasing attempts to monitor and control digital spaces, often at the cost of individual freedoms. It is thus necessary to devote efforts towards reducing these issues and establishing effective guidelines so that cognitive security systems can be analysed and audited with all the ethical, legal, and privacy guarantees. In this category, we also consider evaluation as a guarantee of such compliance, especially depending on the application scenario (Chang et al., 2023). Thus, the lack of consensus on AI red-teaming's scope to identify and mitigate emerging threats, e.g., generative AI, its structure, and assessment criteria, showcases the need for standardised methodologies (Feffer, Sinha, Lipton, & Heidari, 2024).

4.2. Interconnections with current cybersecurity landscape

To increase the value of our analysis and to enable holistic strategies that can go one step ahead of the current landscape, we used recent bibliography to identify the challenges, limitations and threats of cybersecurity (i.e., as determined by the European Union Agency for Cybersecurity (ENISA) (The European Union Agency for Cybersecurity (ENISA), 2023), the United Nations International Computing Centre (UNICC) (United Nations International Computing Centre, 2023) and Crowdstrike (Crowdstrike, 2024)), the digital forensics state of practice (Casino et al., 2022a; Javed et al., 2022), cross-border cybercrime prosecution (Blažič & Klobučar, 2020; Casino et al., 2022b) and cognitive security (Andrade & Yoo, 2019; Huang et al., 2023; Rajtmajer & Susser, 2020), as seen in Fig. 5. Our methodology focused on identifying the challenges, limitations, and threats across these interconnected fields, which were selected for their critical and complementary roles in addressing modern cyber threats and supporting cross-border cybercrime prosecution. Next, we provide a taxonomy of the challenges by using five categories to classify them. Note that several challenges may fall into different categories, yet we selected the most fitting category for each for clarity. For more granular descriptions of each particular challenge, we refer the interested reader to the referenced bibliography.

Next, we elaborated a high-level abstraction and mapped each cognitive security dimension that could be explored towards solving them, as seen in Table 2 (i.e., for the sake of coherence we used the respective categories identified in Fig. 5 for the mapping). Table 2 presents a detailed cross-domain mapping of the challenges and threats in digital forensics (DF), cybercrime prosecution and harmonisation (CH), and cybersecurity (CY), and how the different dimensions of cognitive security relate to them. The table aims to illustrate the potential of cognitive security to address current issues by highlighting how its various dimensions, including learning theories, AI technologies, Big

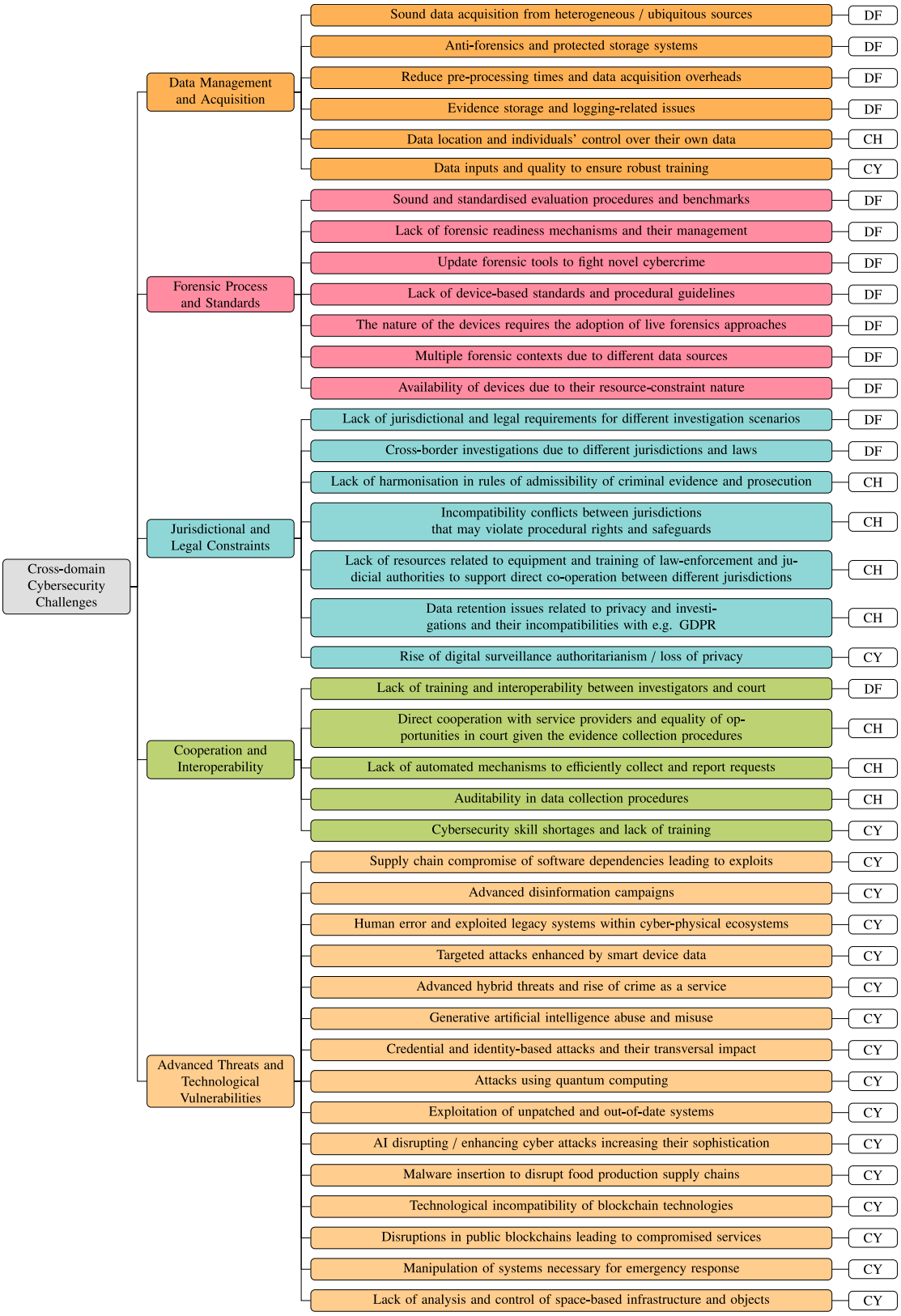

Fig. 5. High-level abstraction of the challenges in the state of the art. We grouped the challenges into categories, easing their identification. Next, we classified each challenge into a specific family, namely digital forensics (DF), cybercrime prosecution and harmonisation (CH), and cybersecurity (CY).

Data management, adaptation to dynamic contexts, ethical guidelines, regulations, and policies, and human–computer interactions, can be leveraged to create solutions. Each challenge category, such as data management and acquisition or jurisdictional and legal constraints, is analysed in terms of its relationship to these dimensions. For example, challenges in data management and acquisition are related to AI technologies for processing large volumes of information, Big Data for handling these volumes, and policies for ensuring data quality, privacy, and auditability. Similarly, issues related to advanced threats and technological vulnerabilities can be mitigated by leveraging learning theories for adapting to new threats, AI technologies for threat detection, and considering dynamic contexts to understand how these threats change.

Data management and acquisition: Digital investigations are inherently tied to evidence acquisition and thus to data management and acquisition. This challenge also considers the difficulty of acquiring data from heterogeneous/ubiquitous sources, highlighting the increasing number, disparity, and protection levels of digital platforms and tools. Both in digital investigations and when creating a dataset, e.g., to train a model, verifiability and auditability are crucial to guarantee the integrity of such data. Moreover, as data volume grows, automating data processing becomes more challenging due to the performance requirements and the resources required to discern between valid/truthful data and fake data, potentially hindering the behaviour of AI models. The latter also complicates data location and control of users' data, hindering the application of digital rights (Casino et al., 2022b). Establishing procedures so that data collection and curation obey specific guidelines or standards, such as the ones mentioned in Section 3, is crucial to guarantee the trustworthiness of both traditional and cognitive cybersecurity technologies, especially when considering the context-aware nature of some scenarios.

Forensic methodologies and standards: Digital forensic processes require strict methodologies for precision, verifiability, and guaranteeing their validity in court. Nowadays, investigations are hindered due to the lack of forensic readiness strategies and their applicability (Casino et al., 2022a). This implies that most systems do not have the measures to collect, monitor, and process data to facilitate their analysis, especially in cases where live forensics is required (Al-Sharif, Al-Saleh, Alawneh, Jararweh, & Gupta, 2020). Integrating cognitive security with emerging technologies creates a transformative paradigm shift in cybersecurity. As AI-driven cyberattacks become more sophisticated, cognitive security systems must leverage cutting-edge AI models, federated learning, and quantum-resistant cryptographic techniques to counteract these threats proactively. The synergy between cognitive security and autonomous threat detection can enhance real-time anomaly identification, mitigating hybrid threats that exploit human cognition and AI vulnerabilities. Furthermore, the ability of cognitive security to function in dynamic and adversarial environments enables it to complement existing cybersecurity frameworks with self-adaptive resilience mechanisms. Finally, promoting uniformity in investigations and elaborating sound standardised procedures and benchmarks will have a significant impact on efficient investigation and prosecution.

Jurisdictional, legal and ethical constraints : The lack of jurisdictional and legal requirements for different investigation scenarios and the auditability of evidence collection procedures differ across countries. The complexities of cross-border investigations due to different jurisdictions and laws are a well-known problem in international digital law (Casino et al., 2022b) that highlights the convoluted paths that investigators, judges, and prosecutors face. As stated in Section 4.1, the lack of guidelines and regulations can be overcome through stakeholder cooperation, promoting multidisciplinary collaboration. Thus, investigators, legal authorities, researchers, and practitioners, among others, should collaborate to integrate mechanisms (in terms of, e.g., explainability and verifiability) that conform to current legislation. The latter should be performed in a timely manner, as cybercriminals are always one step ahead.

Cooperation and interoperability: As previously stated, effective collaboration necessitates seamless interaction between various entities. Clearly, the lack of training and interoperability between investigators and the court explains the issues in communication and understanding, emphasising the need for common languages and standards, which must be extended across all involved actors, including researchers and developers. The value of creating multidisciplinary partnerships can help reduce the issues regarding cooperation between data providers and investigations, including automation and legal and ethical incompatibilities that should not occur. Cognitive-based AI systems require novel background and research since they are still in their infancy. However, cooperation mechanisms between all the actors involved in cybersecurity and cybercrime prosecution should be promoted, leveraging a multidisciplinary skill set that can be augmented via human–computer interactions.

Advanced threats and technological vulnerabilities: The complexity of security-related threats may involve different contexts creating hybrid attacks, including software vulnerabilities, attacks to specific hardware belonging to, e.g., critical infrastructure or emergency response systems, and the use of information and data to generate, e.g., phishing campaigns. Note that phishing is the main entry point for CaaS models to sustain APTs. More concretely, the motives behind these threats can vary significantly. Criminal actors primarily seek financial gains or disruption, often exploiting vulnerabilities through sophisticated campaigns. On the other hand, state actors often pursue espionage and geopolitical dominance through disinformation campaigns and cyberwarfare, leveraging advanced tools and automation to conduct them. Understanding these motives is crucial for tailoring cognitive security measures and workflows to specific threat landscapes. Therefore, the potential impact of novel campaigns is worse than ever if we consider the digitisation level and the scaling integration of AI in our daily lives. Leveraging cognitive security systems that may cope with those challenges by design, e.g., through enabling forensic readiness and providing timely threat intelligence to orchestrated campaigns. Cognitive systems exhibit particular challenges due to their symbiotic relationship between human cognition and digital processes. For instance, overreliance on cognitive systems may cause cognitive bias risks and a lack of critical thinking as AI evolves and humans become more dependent. Note that these systems' data, processing mechanisms, and explainability can also be subject to bias. Thus, holistic and informed judgements should be applied, which humans should assess. While cognitive systems may perform well in multiple contexts, the trade-off between generality and specificity also applies. Thus, it is vital to leverage the proper mechanisms to improve the adaptability of such systems.

5. Discussion and future trends

The digital landscape is continuously under attack by APTs and sophisticated adversaries. In this context, the realisation and adoption of cognitive security systems should follow a holistic approach towards their architectural design, implementation, and testing. By guaranteeing systems that follow best guidelines and practices (CISA, 2024; ENISA, 2024) by enabling monitoring, awareness, readiness, and timely threat intelligence, cognitive security systems represent an interdisciplinary approach. In what follows, we derive research paths and strategies essential to realising cognitive security.

Improving autonomy and performance at scale continues to be a fundamental challenge. The capability of developing self-learning and adaptive systems can be achieved at different levels of granularity, depending on the level of human interaction required in the process. In this regard, quantum computing has the potential to fundamentally alter AI (Gill et al., 2022) and increase the performance of parallelisation tasks, helping AI systems employed in resource-demanding contexts. In parallel, as stated in Section 4.2, sophisticated attacks may scale in magnitude if large botnets and networks are under the

attacker's control. Thus, every time a new technological paradigm appears, attractive opportunities arise for both attackers and defenders, which require increased attention and readiness (Casino et al., 2022a).

Developing cognitive systems that can adapt to changes in context-aware environments ensures that they are not only reactive, but can proactively adjust to new and emerging threats. These systems should remain applicable and effective amidst constant software updates (i.e., guaranteeing privacy-by-design principles and devsecops practices (The Open Worldwide Application Security Project, 2024)), device upgrades, and the introduction of new applications. The integration of agentic workflows in cognitive security can improve the iterative and reflective process of systems, making them more robust and adaptable (Singh et al., 2024). These workflows, which are based on reflection, the use of tools, planning, and collaboration between multiple agents, are essential for developing systems capable of more sophisticated decision-making and problem-solving. Regarding AI techniques, cognitive security systems require adaptability to learn and evolve with the changing environment without requiring extensive retraining (Singh et al., 2024).

Drawing inspiration from various life forms and evolutionary factors, bio-inspired algorithms have been used in a myriad of contexts with remarkable outcomes (Beegum et al., 2023; Gautam, Kaur, & Sharma, 2019). In addition, exploring genetic algorithms, swarm intelligence, and their compatibility with the use of variational autoencoders (Girin et al., 2021) and efficient upgrading mechanisms such as dynamic distributed learning and transfer learning (Ruiz et al., 2023) can enhance the problem-solving capacities of cognitive security, ensuring the necessary adaptability in dynamic digital environments. Moreover, recent projects such as the Brain initiative (*The brain research through advancing innovative neurotechnologies BRAIN*, 2023), which studies brain and neuroscience and its applications towards medicine, cognition processes and behaviours, are establishing promising research lines.

As the shift towards edge computing intensifies, incorporating lightweight and efficient AI models is critical (Lilhore, Dalal, & Simaiya, 2024). These models facilitate local learning, federated learning, transfer learning, and real-time adaptability, such as in the case of liquid neural networks (Hasani, Lechner, Amini, Rus, & Grosu, 2018), creating AI systems that can operate independently across various nodes in a network. Moreover, due to the current hardware requirements of generative AI methods such as LLMs, exploring and expanding the reach of cognitive security systems (e.g., through quantisation Dettmers, Pagnoni, Holtzman, & Zettlemoyer, 2023 and other optimisations) so that they can be realised locally in privacy-aware environments (Aggarwal, Albert, Hill, & Rodan, 2020), can leverage their adoption. As a result, cognitive security systems can provide efficient resource allocation and timely threat mitigation by combining adaptability and performance.

Guaranteeing the robustness of models against attacks is a challenge affecting multiple disciplines. Several techniques are used to counter adversarial attacks according to their attack strategies (e.g., data poisoning, model poisoning, backdoor attacks) and goals (e.g., confidentiality, integrity, availability) (Akhtar & Mian, 2018), such as Defensive Adversarial Learning (DAL), which aims to fortify models against such attacks, and Generative Adversarial Networks (GANs), a class of machine learning models trained in an adversarial setting (Ren, Zheng, Qin, & Liu, 2020). These can be applied not only to locally crafted models but also when knowledge is transferred from other sources using secure and privacy-preserving federated learning and transfer learning (Mothukuri et al., 2021). Thus, studying specific aspects of cognitive security, such as detecting and responding to threats through the local adaptability of AI systems, is crucial to developing defences against sophisticated AI-powered attacks.

The vast amount of data generated today presents both a challenge and an opportunity. As an AI model's performance is directly related to the data they are trained on, there is a need for procedures that actively detect and correct for biases in the collected data (Bagdasaryan & Shmatikov, 2022). Therefore, researching and developing algorithms that actively scan data for biases, whether gender, racial, or other forms, is paramount. In parallel, AI models should use ethical training protocols (UNESCO, 2021) to guarantee that the decisions made by these AI-driven systems are grounded in ethical principles. In addition to exploring algorithms to detect the quality and veracity of the data, its collection should be audited at any stage, ensuring transparency and credibility. Some enabling technologies for the latter are blockchain technologies, The Interplanetary File System (IPFS) (Benet, 2014) or other transparent, tamper-proof mechanisms. Finally, in an era where data privacy is a major concern, employing privacy enhancement technologies at the edge (Alwarafy, Al-Thelaya, Abdallah, Schneider, & Hamdi, 2020), where data is often collected, is crucial to ensure that cognitive systems have quick access to the required data without any security compromise.

Human–computer interactions should provide comprehensive guidelines to ensure that user feedback is captured and integrated effectively. User interaction and cognitive security systems should promote efficient, transparent, and intuitive mechanisms. Improving explainability in decision-making processes is crucial, particularly in cognitive security systems where AI models must provide transparent justifications for their outputs (Miller, 2019). However, existing explainability techniques often fall short in dynamic cybersecurity environments, where threat landscapes evolve rapidly (Rjoub et al., 2023). For instance, current cybersecurity models lack causal reasoning, leading to uninterpretable alerts. Incorporating causal inference models could improve transparency by mapping security threats to cause-and-effect chains (Lykousas, Argyropoulos, & Casino, 2024). In parallel, AI-generated security alerts must integrate real-time feedback loops where human analysts validate, override, or refine AI decisions, ensuring adaptive and interpretable responses (Shandilya, Datta, Kartik, & Nagar, 2024). Finally, since adoption and explanation are often paired with user readiness, cognitive security systems must provide context-specific justifications rather than generic one-size-fits-all explanations (Paredes, Teze, Simari, & Martinez, 2021) (i.e., note that the financial and critical infrastructure sectors face different security threats). As communication and interaction are bi-directional in nature, it is compulsory to study the learning capabilities of humans and their impact on human–computer interactions to understand potential cybersecurity threats and solutions in a real-world setting better (Kim & Kim, 2024). Exploring how humans learn and retain information, how to enhance the understanding between humans and technology, and how readability and explainability can be better defined are examples of challenges to overcome (Alsharida, Al-rimy, Al-Emran, & Zainal, 2023).

Human learning processes play a critical role in cybersecurity, influencing how individuals interpret, respond to, and mitigate cyber threats (Gutzwiller, Fugate, Sawyer, & Hancock, 2015). However, research in this area remains fragmented, often focusing on either AI automation or human decision-making, without fully integrating the two. One relevant issue is that human analysts exhibit confirmation bias and anchoring effects, leading to misinterpretation of AI-generated security alerts (Aggarwal, Venkatesan, Youzwak, Chadha, & Gonzalez, 2024). Studying how humans cognitively process security warnings can help design bias-aware alerting systems. The latter can be leveraged through adaptive learning environments, where training adjusts to analyst expertise and past decision patterns (Vykopal, Seda, Švábenskỳ, & Čeleda, 2022). That learning rate can be further augmented with real-time interactive simulations, where users experience the consequences of poor cybersecurity habits, potentially triggering higher retention rates and behavioural change (Nespoli et al., 2024). In parallel, next-generation hardware solutions that edge closer to transhumanism are emerging, potentially creating unforeseen forms of human–computer interaction (Al-Emran & Deveci, 2024; Dominijanni et al., 2021; Manzocco, 2019). As technology continues to evolve and embed within

human lives, the interface and communication between humans and cognitive systems must be optimised to guarantee unequivocal understanding, trust, and effectiveness. Thus, exploring research lines focusing on the advancements in hardware technologies that merge human and machine interfaces, such as neural link technologies and augmented reality implants, is relevant to understanding their potential for cognitive security. Some technologies to be explored include haptic feedback, augmented reality interfaces, and voice-command systems tailored for cognitive security contexts (Dix, 2017; Zhen et al., 2023).

As one of the critical steps linked to human–computer interactions, autonomous decision-making is particularly challenging. In addition to the potential mistakes, over-reliance on cognitive systems may cause cognitive bias risks (Gilbert, Kather, & Hogan, 2024). Thus, AI-based outcomes require judgements assessed by humans. Nevertheless, as cognitive systems grow in sophistication, interpreting their responses, especially in complex environments, can become challenging. In other words, in the future, we should consider that the complexity of these systems may grow in such a way that it may conflict with their ability to generate easy-to-understand explanations to humans, and thus measures such as modular explanations, the use of intermediate technologies or other strategies may be put in place to enhance user's learning processes (Andrade et al., 2022).

There is an open debate regarding open-sourcing powerful AI tools that are not compliant with current regulations. Stanford's Center for Research on Foundation Models (CRFM) recently evaluated the major AI companies on their transparency (Bommasani et al., 2023). The findings revealed a significant gap in the AI industry's transparency, with the highest-scoring model, Meta's Llama 2, achieving only 54 out of 100 according to their benchmarks. One of the main concerns was the unclear origins of training data in most models and the opacity surrounding refining these models, a step that usually requires human intervention. Consequently, there is a need for greater transparency and ethical considerations in AI development. In this line, efforts should be devoted towards analysing existing AI and cognitive systems focusing on their potential gaps, and ambiguities through evaluation processes (Chang et al., 2023; Feldstein, 2023).

Moreover, adopting a structured, sociotechnical approach to AI red-teaming, involving interdisciplinary teams and diverse threat models, can significantly enhance generative AI's security (Feffer et al., 2024). Such information could be used as input to design models and frameworks that enable cognitive security systems to remain compliant with the relevant regulations automatically, including provisions for periodic audits and assessments. The latter should facilitate the integration of new requirements, such as novel human-enhancing technologies like implantable devices (Fiani et al., 2021), gradually embracing transhumanism practices (Huxley, 2015).

The successful implementation and governance of cognitive security systems require a collaborative effort from multiple stakeholders. Key actors include government regulatory bodies, law enforcement agencies, cybersecurity research institutions, private sector innovators, and civil society organisations. Policy-driven regulations, such as the EU AI Act and GDPR, play a fundamental role in shaping the ethical deployment of cognitive security. Additionally, cross-sector partnerships can facilitate the adoption of standardised cognitive security protocols, ensuring interoperability between public and private security infrastructures. Human oversight remains crucial, with experts responsible for auditing AI-based decisions and establishing accountability measures to mitigate potential algorithmic biases and decision-making opacity.

To fully realise cognitive security, we need to design architectures capable of binding together the different dimensions of cognitive security into a cohesive, adaptable, and efficient solution whose quality can be evaluated. Cognitive security mechanisms should be strategically deployed based on risk indicators, real-time threat intelligence, and environmental conditions. Deployment should be prioritised in high-risk cybersecurity environments, including national security infrastructures, digital forensics investigations, and AI-driven cyber operations. Key activation triggers include suspicious behavioural anomalies, zero-day exploit detections, and predictive threat modelling outcomes. Moreover, deployment must follow a tiered strategy, wherein low-risk environments rely on semi-automated AI assistance, while high-risk sectors demand fully autonomous cognitive security interventions. The deployment timing is crucial in minimising attack dwell time and reducing response latency in cyber incidents. While integrating cognitive security into current cybersecurity frameworks is an open research topic, a first step is to explore technologies capable of realising such architectures and their implementation. For instance, in addition to generative AI, developing neuromorphic computing technologies is an important research line that has revolutionised computer architecture principles since Von Neumann's architecture (Schuman et al., 2022). Other enabling technologies include multi-agent systems (MAS), which allow individual modules (agents) to be devoted to different dimensions of cognitive security and communicate, collaborate, and learn collectively (Savaglio et al., 2020). MAS can mimic the intricate interactions inherent in cognitive security, offering layered and collaborative approaches to decision-making (Lu, Han, Hu, & Zhang, 2016), realising agentic AI. For instance, one agent could continuously learn and adapt to new threats, another could analyse big data to detect patterns indicative of security threats, and another could ensure that the system's actions adhere to current regulations and ethical standards. The agents would collaborate, sharing insights and data, to form a comprehensive, resilient, and adaptive cognitive security system (Talebirad & Nadiri, 2023). Agents can monitor a whole system, provide its corresponding security analysis, and incorporate human-in-the-loop methodologies (Wu et al., 2022), thus transforming outcomes into readable recommendations or orientations for humans (Mosqueira-Rey et al., 2023). This decision-making process can be automated depending on the context by using decision support systems and models, game theory, cognitive map models, and other frameworks such as MAPE-K and OODA (Andrade & Yoo, 2019), ensuring timely and effective responses.

Benchmarking and evaluation of cognitive security systems are required for their validation. Recalling each dimension of cognitive security, we should be able to provide evaluation mechanisms that can evaluate the quality of data and its verifiable provenance, the accuracy of the overall system after defining specific key performance indicators and goals, the performance of the system in terms of efficiency and robustness, and its compliance by using, e.g., standards and guidelines (Bommasani et al., 2023). While defining and creating such benchmarks is challenging, these benchmarking strategies can be effective, enforcing mechanisms when tied to standardisation. In a more global perspective, using such systems requires assessment criteria to ease AI red-teaming efforts, ensuring comprehensive evaluations of generative AI models to prevent their misuse. In this regard, robustness is crucial in order to ensure policy-aware capabilities as well as coherent system responses to hallucinations or potential attacks such as jailbreaks (Barberá, 2025). The latter, paired with clear responsibility definitions (Raza et al., 2025), allows for an auditable landscape, fostering the adoption of cognitive systems.

Finally, we consider that adopting cognitive security is tied to using disruptive methodologies and techniques and the potential benefits to society as a whole (Ma & Huo, 2023). Thus, we have summarised such benefits in Table 3 according to different actors and institutions.

Beyond the contributions of the article, some limitations are worth noting. While this study comprehensively examines cognitive security, several limitations should be acknowledged. First, the interdisciplinary nature of cognitive security means that its realisation requires expertise from multiple fields, including AI, cybersecurity, cognitive science, and regulatory compliance. This study, while thorough, may not have fully explored all the nuances within each of these domains. Additionally, the rapidly evolving landscape of AI and cybersecurity presents a

Table 3
Potential benefits of cognitive security according to different actors and institutions.

Actor/Institution	Description
Emergency response centres	Emergency response centres (ERCs), such as national disaster response agencies and first responders, could integrate cognitive security solutions to enhance communication networks and information systems. These solutions could provide real-time analysis of cyber-threats, automate the detection of anomalies in communication traffic, and prioritise threats based on their potential impact on public safety. By incorporating cognitive security, these centres can ensure that their operations are not disrupted by cyber incidents, which is crucial when coordinating responses to natural disasters or other emergencies.
CSIRTs and CERTs	For CSIRTs and CERTs, cognitive security solutions could be integrated into their existing incident-handling workflows. These solutions could assist in rapidly classifying and prioritising incidents, enhancing red-teaming's methodologies, and enabling responders to focus on the most critical issues first. Cognitive systems could also simulate attacker strategies and suggest the most effective countermeasures, enhancing the overall incident response strategy.
Security Agencies and Institutions	Organisations like CISA, ENISA and ECSO could integrate cognitive security solutions to develop a more cohesive and standardised response to cyber-threats across different jurisdictions. Cognitive security could enhance cross-border collaboration by providing a common platform for threat intelligence sharing and analysis with, e.g., self-learning capabilities and advanced human-machine interactions, reducing language and procedural barriers. The latter would enable a coordinated response to cyber-threats, ensuring that resources are efficiently allocated to where they are most needed.
Law Enforcement and Judiciary Authorities	Law enforcement agencies and judiciary authorities can leverage cognitive security tools to streamline cybercrime investigations and enhance digital forensics processes. These solutions enable automated evidence verification, using machine-learning-based forensic analysis to triage vast amounts of digital data, identify patterns, and establish connections between cybercrime activities. By integrating cognitive security, investigators can reduce case processing times, while judiciary authorities gain better insights into the technical aspects of cybercrime, facilitating more accurate interpretation of digital evidence and informed legal judgements.
Private Sector Institutions	In the private sector, cognitive security solutions could be integrated into existing security operations centres (SOCs). Financial institutions, healthcare providers, and critical infrastructure operators could use cognitive systems to monitor suspicious activities and continuously adapt their defences in real-time. This would protect against current threats and provide insights into emerging risks.
Educational Sector and Retail Businesses	Educational institutions could integrate cognitive security to safeguard against threats that target intellectual property and personal data. Retail businesses, particularly those with a significant online presence, could implement cognitive security to protect against fraud and maintain customer trust.
Social Media Platforms	Social media platforms can deploy AI-driven cognitive security solutions to detect and mitigate misinformation and harmful or illegal content in real-time. By analysing user behaviour and content patterns, these systems can proactively flag or remove content that violates platform policies or poses security risks. Misinformation detection, particularly in political disinformation campaigns, has become a key focus, with platforms like Twitter/X and Facebook integrating AI classifiers to identify deepfake videos and manipulated media. However, despite these advancements, real-time adaptation to evolving threat tactics remains a major challenge, highlighting the need for more robust cognitive security frameworks.

challenge in maintaining up-to-date insights as new threats and technological advancements continue to emerge. Another limitation lies in the theoretical nature of some proposed strategies, which this article could not discuss in-depth. More concretely, while cognitive security frameworks present promising solutions, further empirical validation and real-world implementation studies are necessary to assess their efficacy. Finally, regulatory and ethical concerns remain dynamic, thus requiring assessment schemes (i.e., beyond research articles) that adapt to such frequent changes to ensure cognitive security frameworks comply with evolving global standards.

6. Conclusion

In an era characterised by continuous digitisation and increasing AI integration, the potential impact of novel security threats is worse than ever (Casino et al., 2022a; The European Union Agency for Cybersecurity (ENISA), 2023; United Nations International Computing Centre, 2023). As previously discussed, cognitive security systems designed to leverage AI's self-learning and adaptation capabilities offer a promising solution to counter these threats. At the same time, the emergence of technologies once deemed science fiction, such as implantable devices (Fiani et al., 2021), requires the swift development of guidelines and ontologies to keep pace with these innovations and the evolving landscape of human–computer interactions (Huxley, 2015).

The research questions posed in Table 1 summarise the main objective of our research, namely providing a comprehensive analysis of the state of practice in cognitive security and its main challenges, and elaborating a fruitful discussion on this particular matter. We discuss them in order as follows:

***RQ1**: What are the fundamental components and defining characteristics of cognitive security?*

To provide enough background to discuss the current state of the art, we provide an extensive analysis of related work on cognitive security in Section 2. Moreover, as cognitive security is in its infancy, we particularly focus on the concept's definition and its multidisciplinary nature, which is essential to establishing sound methodologies for integrating cognitive systems into cybersecurity pipelines, as seen in Section 3. Cognitive security encompasses a multi-dimensional approach, integrating AI, human cognition, regulatory compliance, and cybersecurity principles. The definition of characteristics includes adaptive learning mechanisms, real-time decision-making, and the ability to process and respond to complex cybersecurity threats dynamically. Section 3 explores these dimensions in-depth, highlighting the role of learning theories, AI technologies, big data, dynamic contexts, regulatory frameworks, and human–computer interactions. These elements work together to realise cognitive security systems capable of detecting, mitigating, and preventing security threats in real-time, surpassing traditional security paradigms.

***RQ2**: How do cognitive security systems interact with emerging cybersecurity technologies to mitigate advanced threats?*

Cognitive security augments cybersecurity frameworks by incorporating self-learning AI models, federated learning approaches, and cognitive decision-support systems. As discussed in Sections 3 and 4, cognitive security solutions integrate with technologies such as blockchain for data integrity, adaptive AI models and agents, and human-in-the-loop to ensure explainability and accountability in decision-making. Furthermore, these technologies allow cybersecurity systems to function proactively rather than reactively, dynamically adjusting their strategies in response to evolving threats (Casino et al., 2022a; Huang et al., 2023). The interconnections between cognitive security and current cybersecurity landscapes are discussed in Section 4.2, which maps these emerging technologies to existing cybersecurity challenges and potential mitigation strategies.

***RQ3**: Why is the current cybersecurity landscape insufficient in addressing the growing sophistication of cyber threats, and how can cognitive security address these gaps?*

Current cybersecurity solutions struggle to keep up with evolving threats due to their reliance on static defence models, fragmented threat intelligence, and limited adaptability to novel attack vectors. Section 4 outlines the limitations of current cybersecurity frameworks, identifying issues such as jurisdictional constraints, insufficient forensic readiness, and a lack of coordination in cybercrime investigations (Casino et al., 2022b; The European Union Agency for Cybersecurity (ENISA), 2023). Cognitive security addresses these gaps by introducing resilience through continuous threat analysis, adaptive learning, and predictive analytics. Unlike traditional approaches, cognitive security solutions evolve dynamically, leveraging behavioural anomaly detection and real-time response mechanisms to mitigate sophisticated cyber threats before they escalate. Additionally, Section 5 discusses how emerging trends, such as AI-driven malware and cyber–physical system vulnerabilities, further highlight the need for cognitive security frameworks capable of autonomously detecting and responding to cyber threats. The latter, however, has to come with the pertinent regulatory frameworks so that standardised procedures are in place, avoiding bottlenecks, as discussed in RQ4.

***RQ4**: Who are the key stakeholders in cognitive security, and what roles do they play in its implementation and governance?*

The realisation of cognitive security requires collaboration among multiple stakeholders, including researchers, policymakers, cybersecurity practitioners, regulatory bodies, and industry leaders. Section 5 details the distinct roles these stakeholders play, emphasising how researchers contribute to advancing AI-driven security models, policymakers ensure compliance with ethical and legal frameworks, cybersecurity professionals deploy cognitive security tools in operational environments, and industry stakeholders drive the adoption and integration of these technologies into existing security infrastructures. The discussion also highlights the importance of cross-sector collaboration to establish standardised evaluation metrics, enhance interoperability, and develop regulatory guidelines that balance security concerns with privacy and ethical considerations (Stuurman & Lachaud, 2022).

***RQ5**: Where are cognitive security frameworks most applicable, and how do different environments impact their design and effectiveness?*

Cognitive security is particularly relevant in critical infrastructure, financial systems, law enforcement, and national security domains. Each environment presents unique challenges, necessitating tailored cognitive security strategies. Section 4.2 outlines the application of cognitive security frameworks in various domains, emphasising their impact on incident response, fraud detection, cybercrime investigations, and regulatory compliance. For instance, cognitive security helps detect fraud and insider threats through real-time behavioural analysis in financial systems. At the same time, in law enforcement, it aids in digital forensics by automating evidence classification and prioritisation. The adaptability of cognitive security frameworks to different operational contexts is crucial to their effectiveness, and this adaptability is further explored in Section 5, which discusses emerging technological advancements and deployment strategies.

***RQ6**: When should cognitive security mechanisms be deployed in cybersecurity workflows, and which strategies can be used to address cybercrime in the near future?*

Cognitive security mechanisms should be deployed as part of a proactive cybersecurity strategy, integrated in multiple layers of digital ecosystems. As highlighted in Section 5, cognitive security should not be an afterthought but a core component of cybersecurity workflows, providing real-time threat intelligence, automated decision-making, and continuous learning capabilities. Strategies such as predictive analytics, automated threat intelligence processing, and adversarial AI defence mechanisms ensure real-time protection against sophisticated cyber threats. Additionally, strengthening regulatory compliance and ethical considerations will be vital for the widespread adoption and long-term sustainability of cognitive security solutions. Organisations can ensure transparency and accountability by integrating AI-driven cybersecurity models with legal and ethical oversight.

Future research on cognitive security should explore one or several of the identified dimensions and the corresponding challenges, with a particular focus on the topics discussed in Section 5. Thus, we aim to study integration techniques, adaptive and collaborative approaches such as agentic AI and agentic workflows, and enriched human–computer interaction strategies, all considering ethical and legal collaborative frameworks. In addition, we will explore the creation of benchmarks and standardised methodologies that can be used to verify the robustness of cognitive security systems. Finally, researchers, practitioners, and policymakers should provide strategies to support the realisation of novel paradigms such as cognitive security, for instance, by providing interdisciplinary insights to address gaps in ethical compliance and scalability, enhancing auditability and explainability, and integrating regulatory frameworks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abie, H. (2019). Cognitive cybersecurity for cps-iot enabled healthcare ecosystems. In *2019 13th international symposium on medical information and communication technology* (pp. 1–6).

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals–a comprehensive survey. *IEEE Access*.

Aggarwal, N., Albert, L. J., Hill, T. R., & Rodan, S. A. (2020). Risk knowledge and concern as influences of purchase intention for internet of things devices. *Technology in Society*, *62*, Article 101311.

Aggarwal, P., Venkatesan, S., Youzwak, J., Chadha, R., & Gonzalez, C. (2024). Discovering cognitive biases in cyber attackers' network exploitation activities: A case study. In Abbas Moallem (Ed.), *Human factors in cybersecurity. AHFE (2024) International conference*.

Akhtar, N., & Mian, A. (2018). Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access*, *6*, 14410–14430.

Al Amin, M. (2022). Supervised learning for detecting cognitive security anomalies in real-time log data. In *2022 IEEE world conference on applied intelligence and computing* (pp. 840–845). IEEE.

Al-Emran, M., & Deveci, M. (2024). Unlocking the potential of cybersecurity behavior in the metaverse: Overview, opportunities, challenges, and future research agendas. *Technology in Society*, Article 102498.

Al-Kadhimi, A. A., Singh, M. M., & Khalid, M. N. A. (2023). A systematic literature review and a conceptual framework proposition for advanced persistent threats (APT) detection for mobile devices using artificial intelligence techniques. *Applied Sciences*, *13*(8056).

Al-Sharif, Z. A., Al-Saleh, M. I., Alawneh, L. M., Jararweh, Y. I., & Gupta, B. (2020). Live forensics of software attacks on cyber–physical systems. *Future Generation Computer Systems*, *108*, 1217–1229.

Alsharida, R. A., Al-rimy, B. A. S., Al-Emran, M., & Zainal, A. (2023). A systematic review of multi perspectives on human cybersecurity behavior. *Technology in Society*, Article 102258.

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review*, *36*, 247–271.

Alwarafy, A., Al-Thelaya, K. A., Abdallah, M., Schneider, J., & Hamdi, M. (2020). A survey on security and privacy issues in edge-computing-assisted internet of things. *IEEE Internet of Things Journal*, *8*, 4004–4022.

Amann, P., & James, J. I. (2015). Designing robustness and resilience in digital investigation laboratories. *Digital Investigation*, *12*, S111–S120.

Andrade, R. O., Fuertes, W., Cazares, M., Ortiz-Garcés, I., & Navas, G. (2022). An exploratory study of cognitive sciences applied to cybersecurity. *Electronics*, *11*(1692).

Andrade, R., & Torres, J. (2018). Self-awareness as an enabler of cognitive security. In *2018 IEEE 9th annual information technology, electronics and mobile communication conference* (pp. 701–708). IEEE.

Andrade, R., Torres, J., & Cadena, S. (2019). Cognitive security for incident management process. In Á. Rocha, C. Ferrás, & M. Paredes (Eds.), *Information Technology and Systems* (pp. 612–621). Cham: Springer International Publishing.

Andrade, R., Torres, J., & Flores, P. (2018). Management of information security indicators under a cognitive security model. In *2018 IEEE 8th annual computing and communication workshop and conference* (pp. 478–483).

Andrade, R., Torres, J., & Tello-Oquendo, L. (2018). Cognitive security tasks using big data tools. In *2018 international conference on computational science and computational intelligence* (pp. 100–105). IEEE.

Andrade, R. O., & Yoo, S. G. (2019). Cognitive security: A comprehensive study of cognitive science in cybersecurity. *Journal of Information Security and Applications*, *48*, Article 102352.

Bagdasaryan, E., & Shmatikov, V. (2022). Spinning language models: Risks of propaganda-as-a-service and countermeasures. In *2022 IEEE symposium on security and privacy* (pp. 769–786). IEEE.

Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N., & Jararweh, Y. (2022). Federated learning review: Fundamentals, enabling technologies, and future applications. *Information Processing & Management*, *59*, Article 103061.

Barberá, I. (2025). AI privacy risks and mitigations large language models (llms).

Bayne, T., et al. (2019). What is cognition? *Current Biology*, *29*, R608–R615.

Beegum, T. R., et al. (2023). Optimized routing of uavs using bio-inspired algorithm in fanet: A systematic review. *IEEE Access*, *11*, 15588–15622.

Benet, J. (2014). Ipfs-content addressed, versioned, p2p file system. arXiv preprint arXiv:1407.3561.

Benraouane, S. A. (2024). *AI management system certification according to the ISO/IEC 42001 standard: How to audit, certify, and build responsible AI systems*. CRC press.

Benzaid, C., Taleb, T., & Song, J. (2022). AI-based autonomic and scalable security management architecture for secure network slicing in b5g. *IEEE Network*, *36*, 165–174.

Biever, C. (2023). Chatgpt broke the turing test-the race is on for new ways to assess AI. *Nature*, *619*, 686–689.

Blažič, B. J., & Klobučar, T. (2020). Removing the barriers in cross-border crime investigation by gathering e-evidence in an interconnected society. *Information & Communications Technology Law*, *29*, 66–81.

Bommasani, R., Klyman, K., Zhang, D., & Liang, P. (2023). Do foundation model providers comply with the eu AI act?

Bommasani, R., et al. (2023). The foundation model transparency index.

Burton, J. (2023). Algorithmic extremism? the securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence. *Technology in Society*, *75*, Article 102262.

Butakova, M. A., Chernov, A. V., & Shevchuk, P. S. (2019). An approach for distributed reasoning on security incidents in critical information infrastructure with intelligent awareness systems. In *Proceedings of the computational methods in systems and software* (pp. 248–255). Springer.

Casino, F., et al. (2022a). Research trends, challenges, and emerging topics in digital forensics: A review of reviews. *IEEE Access*, *10*, 25464–25493.

Casino, F., et al. (2022b). Sok: cross-border criminal investigations and digital evidence. *Journal of Cybersecurity*, *8*, tyac014.

Chang, Y., et al. (2023). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*.

Chen, H.-C., Yang, W.-J., & Chou, C.-L. (2021). An online cognitive authentication and trust evaluation application programming interface for cognitive security gateway based on distributed massive internet of things network. *Concurrency and Computation: Practice and Experience*, *33*, Article e6128.

Chouhan, P. K., Chen, L., Hussain, T., & Beard, A. (2021). A situation calculus based approach to cognitive modelling for responding to iot cyberattacks. In *2021 IEEE smartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, internet of people and smart city innovation* (pp. 219–225). IEEE.

Cinque, M., Cotroneo, D., & Pecchia, A. (2019). Towards cognitive security defense from data. In *2019 49th annual IEEE/iFIP international conference on dependable systems and networks–supplemental volume* (pp. 11–12). IEEE.

CISA (2024). Cybersecurity best practices.

Claverie, B., & du Cluzel, F. (2022). The cognitive warfare concept. In *Innovation hub sponsored by NATO allied command transformation* (pp. 1–11).

Crowdstrike (2024). Crowdstrike 2024 global threat report.

Dautenhahn, K. (1998). The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied Artificial Intelligence*, *12*, 573–617.

Demertzis, K., & Iliadis, L. (2023). An autonomous self-learning and self-adversarial training neural architecture for intelligent and resilient cyber security systems. In *International conference on engineering applications of neural networks* (pp. 461–478). Springer.

Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In *The Sage handbook of organizational research methods* (pp. 671–689).

Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). Qlora: Efficient finetuning of quantized llms. arXiv preprint arXiv:2305.14314.

Devadarshini, P., Chandrashekar, B., Pundir, S., Tiwari, M., Madala, R., & Indhuma, E. (2023). Cognitive defense cyber attack prediction and security design in machine learning model. In *2023 6th international conference on contemporary computing and informatics*: *Vol. 6*, (pp. 1361–1366). IEEE.

Dix, A. (2017). Human–computer interaction, foundations and new paradigms. *Journal of Visual Languages & Computing*, *42*, 122–134.

Doherty, G. (2023). Cognitive security: An architecture informed approach from cognitive science. In *International conference on human-computer interaction* (pp. 395–415). Springer.

Dominijanni, G., et al. (2021). The neural resource allocation problem when enhancing human bodies with extra robotic limbs. *Nature Machine Intelligence*, *3*, 850–860.

Dupoux, E. (2018). Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, *173*, 43–59.

Echterhoff, J., Liu, Y., Alessa, A., McAuley, J., & He, Z. (2024). Cognitive bias in high-stakes decision-making with llms. arXiv preprint arXiv:2403.00811.

El-Sayed, A., Toony, A. A., Alqahtani, F., Alginahi, Y., & Said, W. (2025). Co-stop: A robust p4-powered adaptive framework for comprehensive detection and mitigation of coordinated and multi-faceted attacks in sd-iot networks. *Computers & Security*, Article 104349.

Engler, A. (2023). The EU and US diverge on AI regulation: A transatlantic comparison and steps to alignment.

ENISA (2024). Best practices for cyber crisis management.

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, *6*, 50–72.

Feffer, M., Sinha, A., Lipton, Z. C., & Heidari, H. (2024). Red-teaming for generative ai: Silver bullet or security theater? arXiv preprint arXiv:2401.15897.

Feldstein, S. (2023). Evaluating Europe's push to enact ai regulations: how will this influence global norms? *Democratization*, 1–18.

Fiani, B., Reardon, T., Ayres, B., Cline, D., Sitto, S. R., Reardon, T. K., et al. (2021). An examination of prospective uses and future directions of neuralink: the brain-machine interface. *Cureus*, *13*.

Forum, W. E. (2023). The cybersecurity skills gap is a real threat — here's how to address it.

Furdek, M., Natalino, C., Lipp, F., Hock, D., Giglio, A. D., & Schiano, M. (2020). Machine learning for optical network security monitoring: A practical perspective. *Journal of Lightwave Technology*, *38*, 2860–2871.

Garcés, I. O., Cazares, M. F., & Andrade, R. O. (2019). Detection of phishing attacks with machine learning techniques in cognitive security architecture. In *2019 international conference on computational science and computational intelligence* (pp. 366–370). IEEE.

Gautam, R., Kaur, P., & Sharma, M. (2019). A comprehensive review on nature inspired computing algorithms for the diagnosis of chronic disorders in human beings. *Progress in Artificial Intelligence*, *8*, 401–424.

George, S. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, *2*, 3–10.

Georgoulias, D., Pedersen, J. M., Falch, M., & Vasilomanolakis, E. (2023). Botnet business models, takedown attempts, and the darkweb market: A survey. *ACM Computing Surveys*, *55*, 1–39.

Gilbert, S., Kather, J. N., & Hogan, A. (2024). Augmented non-hallucinating large language models as medical information curators. *NPJ Digital Medicine*, *7*(100).

Gill, S. S., Xu, M., Ottaviani, C., Patros, P., Bahsoon, R., Shaghaghi, A., et al. (2022). Ai for next generation computing: Emerging trends and future directions. *Internet of Things*, *19*, Article 100514.

Girin, L., Leglaive, S., Bie, X., Diard, J., Hueber, T., & Alameda-Pineda, X. (2021). Dynamical variational autoencoders: A comprehensive review. *Foundations and Trends in Machine Learning*, *15*, 1–175.

Greenstadt, R., & Beal, J. (2008). Cognitive security for personal devices. In *Proceedings of the 1st ACM workshop on workshop on AISec* (pp. 27–30).

Guerrero, L. E., Castillo, L. F., Arango-Lopez, J., & Moreira, F. (2025). A systematic review of integrated information theory: a perspective from artificial intelligence and the cognitive sciences. *Neural Computing and Applications*, *37*, 7575–7607.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). Xai—explainable artificial intelligence. *Science Robotics*, *4*, eaay7120.

Guo, Y. (2023). A review of machine learning-based zero-day attack detection: Challenges and future directions. *Computer Communications*, *198*, 175–185.

Gutierrez-Garcia, J. O., & López-Neri, E. (2015). Cognitive computing: a brief survey and open research challenges. In *2015 3rd international conference on applied computing and information technology/2nd international conference on computational science and intelligence* (pp. 328–333). IEEE.

Gutzwiller, R. S., Fugate, S., Sawyer, B. D., & Hancock, P. (2015). The human factors of cyber network defense. In *Proceedings of the human factors and ergonomics society annual meeting*: *Vol. 59*, (pp. 322–326). Los Angeles, CA: SAGE publications Sage CA.

Hacker, P., Engel, A., & Mauer, M. (2023). Regulating chatgpt and other large generative AI models. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency* (pp. 1112–1123).

Hasani, R. M., Lechner, M., Amini, A., Rus, D., & Grosu, R. (2018). Liquid time-constant recurrent neural networks as universal approximators.

Huang, R., Zheng, X., Shang, Y., & Xue, X. (2023). On challenges of AI to cognitive security and safety. *Security and Safety*, *2*, Article 2023012.

Huang, L., & Zhu, Q. (2023). *Cognitive security: A system-scientific approach*. Springer Nature.

Huxley, J. (2015). Transhumanism. *Ethics in Progress*, *6*, 12–16.

Jackson, E. A. (2024). *The evolution of artificial intelligence: A theoretical review of its impact on teaching and learning in the digital age*. Kiel, Hamburg: ZBW–Leibniz Information Centre for Economics.

Jagadeesan, L., Mc Bride, A., Gurbani, V. K., & Yang, J. (2015). Cognitive security: Security analytics and autonomics for virtualized networks. In *Proceedings of the principles, systems and applications on IP telecommunications* (pp. 43–50).

Jaramillo, J. A. (1996). Vygotsky's sociocultural theory and contributions to the development of constructivist curricula. *Education*, *117*, 133–141.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*, 577–586.

Javed, A. R., et al. (2022). A comprehensive survey on computer forensics: State-of-the-art, tools, techniques, challenges, and future directions. *IEEE Access*, *10*, 11065–11089.

Jayaganesh, J., & Parvees, M. M. (2022). A study of identity threat using cyberthreat defense network. In *2022 4th international conference on smart systems and inventive technology* (pp. 737–740). IEEE.

Jiang, Y., & Atif, Y. (2021). A selective ensemble model for cognitive cybersecurity analysis. *Journal of Network and Computer Applications*, *193*, Article 103210.

Kavitha, R., Priya, N., & India, T. (2019). Cognitive security in software define network layer. *Journal of Mechanics of Continua and Mathematical Sciences*.

Khodabandehloo, E., Riboni, D., & Alimohammadi, A. (2021). Healthxai: Collaborative and explainable ai for supporting early diagnosis of cognitive decline. *Future Generation Computer Systems*, *116*, 168–189.

Kim, B.-J., & Kim, M.-J. (2024). The influence of work overload on cybersecurity behavior: A moderated mediation model of psychological contract breach, burnout, and self-efficacy in AI learning such as chatgpt. *Technology in Society*, *77*, Article 102543.

Køien, G. M. (2021). Initial reflections on the use of augmented cognition in derailing the kill chain. In *International conference on human-computer interaction* (pp. 433–451). Springer.

Lakhdhar, Y., Rekhis, S., & Sabir, E. (2020). A game theoretic approach for deploying forensic ready systems. In *2020 international conference on software, telecommunications and computer networks* (pp. 1–6). IEEE.

Lilhore, U. K., Dalal, S., & Simaiya, S. (2024). A cognitive security framework for detecting intrusions in iot and 5G utilizing deep learning. *Computers & Security*, *136*, Article 103560.

Liwång, H. (2022). Defense development: The role of co-creation in filling the gap between policy-makers and technology development. *Technology in Society*, *68*, Article 101913.

Lu, J., Han, J., Hu, Y., & Zhang, G. (2016). Multilevel decision-making: A survey. *Information Sciences*, *346*, 463–487.

Lykousas, N., Argyropoulos, V., & Casino, F. (2024). The potential of llm-generated reports in devsecops. In *International conference on AI-empowered software engineering*. Springer.

Ma, X., & Huo, Y. (2023). Are users willing to embrace chatgpt? exploring the factors on the acceptance of chatbots from the perspective of aidua framework. *Technology in Society*, *75*, Article 102362.

Manky, D. (2013). Cybercrime as a service: a very modern business. *Computer Fraud & Security*, *2013*, 9–13.

Manzocco, R. (2019). Transhumanism. In *Engineering the human condition*. Suiza: Springer.

Martínez Santander, C., Yoo, S. G., & Moreno, H. O. (2018). Analysis of traditional web security solutions and proposal of a web attacks cognitive patterns classifier architecture. In *International conference on technologies and innovation* (pp. 186–198). Springer.

maxqda (2024). https://www.maxqda.com/. (Accessed 04 April 2024).

Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digital Medicine*, *6*(120).

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38.

Milosevic, N., Jakovetic, D., Skrbic, S., Savic, M., Stamenkovic, D., Mascolo, J., et al. (2022). Bacs: A comprehensive tool for deep learning-based anomaly detection in edge-fog-cloud systems. In *2022 30th European signal processing conference* (pp. 1097–1101). IEEE.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: a state of the art. *Artificial Intelligence Review*, *56*, 3005–3054.

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, *115*, 619–640.

Mozes, M., et al. (2023). Use of llms for illicit purposes: Threats, prevention measures, and vulnerabilities.

Muraleedharan, R., & Osadciw, L. A. (2009). Cognitive security protocol for sensor based vanet using swarm intelligence. In *2009 conference record of the forty-third asilomar conference on signals, systems and computers* (pp. 288–290). IEEE.

NATO (2023). Cognitive warfare: Strengthening and defending the mind.

Nespoli, P., Albaladejo-González, M., Valera, J. A. P., Ruipérez-Valiente, J. A., Garcia-Alfaro, J., & Mármol, F. G. (2024). Scorpion cyber range: Fully customizable cyberexercises, gamification and learning analytics to train cybersecurity competencies. arXiv preprint arXiv:2401.12594.

Ogiela, U. (2021). Transformative and cognitive approaches to information retrieval and security procedures. *Concurrency and Computation: Practice and Experience*, *33*, Article e5890.

OWASP Foundation (2023). Owasp top 10 for large language model applications. https://owasp.org/www-project-top-10-for-large-language-model-applications/.

Paredes, J. N., Teze, J. C. L., Simari, G. I., & Martinez, M. V. (2021). *On the importance of domain-specific explanations in AI-based cybersecurity systems*: *Technical report*, arXiv preprint arXiv:2108.02006.

Patsakis, C., Arroyo, D., & Casino, F. (2025). The malware as a service ecosystem. In D. Gritzalis, K.-K. R. Choo, & C. Patsakis (Eds.), *Malware: Handbook of Prevention and Detection* (pp. 371–394). Cham: Springer Nature Switzerland.

Prabavathy, S., & Supriya, V. (2021). Sdn based cognitive security system for large-scale internet of things using fog computing. In *2021 international conference on emerging techniques in computational intelligence* (pp. 129–134). IEEE.

Pranckute, R. (2021). Web of science (wos) and scopus: The titans of bibliographic information in today's academic world. *Publications*, *9*(12).

Rajtmajer, S., & Susser, D. (2020). Automated influence and the challenge of cognitive security. In *Proceedings of the 7th symposium on hot topics in the science of security* (pp. 1–9).

Raza, S., et al. (2025). Who is responsible? the data, models, users or regulations? responsible generative AI for a sustainable future. arXiv preprint arXiv:2502.08650.

Ren, K., Zheng, T., Qin, Z., & Liu, X. (2020). Adversarial attacks and defenses in deep learning. *Engineering*, *6*, 346–360.

Rjoub, G., et al. (2023). A survey on explainable artificial intelligence for cybersecurity. *IEEE Transactions on Network and Service Management*, *20*, 5115–5140.

Rowe, Frantz (2014). What literature review is not: diversity, boundaries and recommendations. *European Journal of Information Systems*, *23*(3), 241–255.

Ruiz, C. R., et al. (2023). Scalable transfer learning with expert models. US Patent App. 18/008, 293.

Şahin, M. (2012). Pros and cons of connectivism as a learning theory. *International Journal of Physical and Social Sciences*, *2*, 437–454.

Şahin, M. (2024). Approaches to learning: From traditional theories to connnectivism. In *Education & science 2024-IV* (p. 7).

Savaglio, C., Ganzha, M., Paprzycki, M., Bădică, M., & Fortino, G. (2020). Agent-based internet of things: State-of-the-art and research challenges. *Future Generation Computer Systems*, *102*, 1038–1053.

Schroer, A. (2023). AI cybersecurity: 25 companies to know. https://builtin.com/artificial-intelligence/artificial-intelligence-cybersecurity.

Schuman, C. D., et al. (2022). Opportunities for neuromorphic computing algorithms and applications. *Nature Computational Science*, *2*, 10–19.

Sen, Ö., Malskorn, P., Glomb, S., Hacker, I., Henze, M., & Ulbig, A. (2023). An approach to abstract multi-stage cyberattack data generation for ml-based ids in smart grids. In *2023 IEEE Belgrade PowerTech* (pp. 01–10). IEEE.

Shandilya, S. K., Datta, A., Kartik, Y., & Nagar, A. (2024). Role of artificial intelligence and machine learning. In *Digital resilience: Navigating disruption and safeguarding data privacy* (pp. 313–399). Springer.

Silva, J. A. H., & Hernández-Alvarez, M. (2017). Large scale ransomware detection by cognitive security. In *2017 IEEE second Ecuador technical chapters meeting* (pp. 1–4). IEEE.

Singh, A., Ehtesham, A., Kumar, S., & Khoei, T. T. (2024). Enhancing ai systems with agentic workflows patterns in large language model. In *2024 IEEE world AI IoT congress* (pp. 527–532). IEEE.

Sreedevi, A., Harshitha, T. N., Sugumaran, V., & Shankar, P. (2022). Application of cognitive computing in healthcare, cybersecurity, big data and iot: A literature review. *Information Processing & Management*, *59*, Article 102888.

Sreekumaridevi, R. M. (2011). *Cognitive security framework for heterogeneous sensor network using swarm intelligence*. Syracuse University.

Stuurman, K., & Lachaud, E. (2022). Regulating ai. a label to complete the proposed act on artificial intelligence. *Computer Law & Security Review*, *44*, Article 105657.

Talebirad, Y., & Nadiri, A. (2023). Multi-agent collaboration: Harnessing the power of intelligent llm agents. arXiv preprint arXiv:2306.03314.

Tantar, E., Tantar, A.-A., Kantor, M., & Engel, T. (2018). On using cognition for anomaly detection in sdn. In *EVOLVE-a bridge between probability, set oriented numerics, and evolutionary computation VI* (pp. 67–81). Springer.

Tariq, U., Ahanger, T. A., Nusir, M., & Ibrahim, A. (2021). A pervasive computational intelligence based cognitive security co-design framework for hype-connected embedded industrial iot. *International Journal of Computers Communications & Control*, *16*.

The brain research through advancing innovative neurotechnologies BRAIN. (2023).

The European Union Agency for Cybersecurity (ENISA) (2023). Foresight 2030 threats. https://www.enisa.europa.eu/publications/foresight-2030-threats.

The Open Worldwide Application Security Project (2024). Owasp devsecops guideline.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*, 207–222.

Triguero, I., Molina, D., Poyatos, J., Del Ser, J., & Herrera, F. (2024). General purpose artificial intelligence systems (gpais): Properties, definition, taxonomy, societal implications and responsible governance. *Information Fusion*, *103*, Article 102135.

UNESCO, C. (2021). Recommendation on the ethics of artificial intelligence.

United Nations International Computing Centre (2023). Cyber threat landscape report 2022.

Vom Brocke, J., et al. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems*, *37*(9).

Vykopal, J., Seda, P., Švábenskỳ, V., & Čeleda, P. (2022). Smart environment for adaptive learning of cybersecurity skills. *IEEE Transactions on Learning Technologies*, *16*, 443–456.

Wall, D. S. (2024). *Cybercrime: The transformation of crime in the information age*. John Wiley & Sons.

Wang, Y. (2002). On cognitive informatics. In *Proceedings first IEEE international conference on cognitive informatics* (pp. 34–42). IEEE.

Wang, L., et al. (2023). A survey on large language model based autonomous agents. arXiv preprint arXiv:2308.11432.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*, 490–495.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, *135*, 364–381.

Xu, C., Qu, Y., Xiang, Y., & Gao, L. (2023). Asynchronous federated learning on heterogeneous devices: A survey. *Computer Science Review*, *50*, Article 100595.

Yılmaz, M. H., Güvenkaya, E., Furqan, H. M., Köse, S., & Arslan, H. (2017). Cognitive security of wireless communication systems in the physical layer. *Wireless Communications and Mobile Computing*, *2017*, Article 3592792.

Zhao, H., et al. (2024). Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*, *15*, 1–38.

Zhen, R., Song, W., He, Q., Cao, J., Shi, L., & Luo, J. (2023). Human–computer interaction system: A survey of talking-head generation. *Electronics*, *12*(218).

Zheng, Y., Moini, A., Lou, W., Hou, Y. T., & Kawamoto, Y. (2016). Cognitive security: securing the burgeoning landscape of mobile networks. *IEEE Network*, *30*, 66–71.