

Cyber-resilient Industry 4.0: digital system failures, algorithmic bias and governance challenges for sustainable smart cities

Dr. Sonia GARCÍA-MORENO
University of Castilla-La Mancha, Spain
Sonia.garcia10@alu.uclm.es

Abstract

Objectives: This paper examines a central tension within Industry 4.0. Digitalisation can improve efficiency, monitoring capacity and decision making, but it also creates vulnerabilities linked to digital system failures, cyber incidents and algorithmic bias. The study analyses how these risks condition sustainable and resilient digital transformation in industrial environments and why they are relevant to smart cities. **Prior work:** The paper builds on research on Industry 4.0, digital transformation and productivity, together with studies on cyber physical systems, smart city security, Internet of Things (IoT) vulnerabilities, artificial intelligence (AI) enabled cyber defence and algorithmic governance. It connects these strands to move beyond technical performance and cybersecurity alone. **Approach:** The study follows a conceptual and analytical approach, supported by selected secondary evidence from Spain and the European context. Spain provides a contextual empirical anchor for illustrating the operational and social relevance of cyber risk, while the European framework supports the analysis of cybersecurity, algorithmic transparency, human oversight and AI governance. **Results:** The analysis shows that Industry 4.0 should not be assessed only through automation, efficiency or productivity gains. Its sustainability also depends on resilient digital infrastructures and on the governance conditions under which algorithmic systems are deployed. Under opacity, scale and limited accountability, such systems may amplify hidden biases and generate persistent harms in workforce management, supplier selection and investment allocation. **Implications:** The findings are relevant for researchers, policy makers and practitioners involved in smart industry and smart city strategies, as they stress the need to align innovation with resilience, accountability and organisational trust. **Value:** The paper contributes an integrated framework linking cyber resilience, algorithmic governance and social sustainability, showing that sustainable digital transformation requires both resilient infrastructures and accountable automated decisions.

Keywords: trustworthy AI; algorithmic accountability; regulatory compliance; organisational trust; digital governance.

1. Introduction

Industry 4.0 has become one of the central paradigms for explaining the digital transformation of manufacturing. Since the early formulation of the German strategic initiative, the concept has been associated with the integration of cyber physical systems, industrial Internet of Things, data intensive platforms, cloud computing, artificial intelligence and new forms of automation in production environments [1], [2]. The literature has emphasised that these technologies allow firms to improve monitoring capacity, process coordination, flexibility and productivity, while also reshaping industrial value creation through connected machines, real time data and autonomous or semi-autonomous decision making [3], [4]. At the same time, smart city research has developed around a similar technological logic, in which information and communication technologies are embedded in urban infrastructures to improve service delivery, mobility, energy management, environmental monitoring and quality of life [5], [6].

This connection confirms that smart factories and smart cities should not be treated as isolated fields. Both rely on interconnected infrastructures, continuous data flows, digital platforms and automated decisions, and both face common challenges concerning reliability, security, transparency and governance [7], [8]. Recent research on smart city initiatives and smart governance has also shown that urban digital transformation depends on institutional capacity, citizen services, public administration and legal frameworks for algorithmic transformation [9], [10]. This governance orientation is reinforced by work that frames the transition from smart systems to smart governance as a shift in which the value of digital transformation depends on public value, institutional learning and coordinated decision making [11].

The expansion of digital infrastructures has often been presented as a condition for more efficient and sustainable production and urban management. In industrial environments, sensors, cyber physical systems and predictive analytics can reduce downtime, optimise energy consumption and improve the allocation of resources. In smart cities, similar technologies can support traffic management, public safety, water and energy systems, waste collection and environmental planning. However, the same interconnection that enables efficiency also increases exposure to new forms of fragility. Cyber physical systems combine software, hardware, communication networks and physical processes, which means that failures may emerge from technical malfunction, poor integration, data errors, insecure devices or hostile attacks [12], [13]. Industrial Internet of Things environments are particularly exposed because their architecture frequently involves heterogeneous devices, legacy systems, external connectivity and several layers of interaction between operational technology and information technology [14], [15]. Consequently, digital transformation cannot be assessed only through technological performance or productivity gains. It must also be examined through the resilience of the infrastructures that sustain automated and data driven operations.

Cybersecurity research has shown that smart industrial and urban systems are vulnerable to confidentiality, integrity and availability risks. These risks affect not only information assets, but also physical processes, service continuity and organisational trust [12], [13]. In the European context, the European Union Agency for Cybersecurity (ENISA) has repeatedly identified ransomware, distributed denial of service attacks, supply chain compromise, social engineering and data related threats as major components of the contemporary threat landscape [16]. The Spanish National Cybersecurity Institute (INCIBE) reported that its Computer Emergency Response Team (CERT) managed 83,517 cybersecurity incidents in 2023, including incidents affecting citizens, private companies and essential or critical operators [17].

The Spanish Ministry of the Interior also reported 472,125 recorded cybercrime offences in 2023, with computer fraud representing 90.50% of the total [18]. These data do not refer only to Industry 4.0 environments, but they illustrate the wider context in which digital infrastructures have become a central vector of operational, economic and social risk. For smart industry and smart cities, this reinforces the need to move from a narrow

cybersecurity approach towards a broader understanding of cyber resilience, understood as the capacity to anticipate, absorb, respond to and recover from digital disruptions.

A second dimension of vulnerability is algorithmic. Industry 4.0 and smart city systems increasingly use data-based models to classify, predict, recommend or prioritise decisions. These systems may support maintenance, workforce planning, supplier evaluation, investment allocation, logistics, mobility management or public service distribution. Yet algorithmic systems are not neutral simply because they are mathematical or automated. Prior research has shown that data driven decision making can reproduce disparate impacts when historical data, proxy variables, target definitions or model design choices incorporate previous inequalities or institutional biases [19], [20]. Algorithmic bias may also arise when models optimise global accuracy, cost reduction or efficiency while ignoring fairness, contestability or the unequal distribution of errors across groups [21]. In this sense, digital transformation can create new forms of organisational asymmetry, especially when affected individuals do not understand how decisions are made or lack effective channels to challenge them.

This problem is especially relevant when algorithmic systems operate under conditions of opacity, scale and potential harm. O'Neil's concept of "*weapons of math destruction*" is useful here because it highlights how opaque models can generate large scale and persistent damage when they are applied to sensitive decisions without adequate accountability [22]. In industrial organisations, similar risks may appear in algorithmic systems used for recruitment, performance assessment, task allocation, promotion, workforce monitoring or supplier scoring. Research on artificial intelligence in human resource management has already warned about the gap between the promise of AI and the practical problems associated with fairness, small data, accountability and employee reactions [23]. Studies on algorithms at work also show that algorithmic management may reshape organisational control, intensify surveillance and generate new forms of worker resistance when transparency and participation are weak [24]. Trust in automation is therefore a key condition for adoption. If workers and managers distrust automated systems, they may underuse them. If they trust them blindly, systematic errors or biases may be transferred mechanically into organisational decisions [25].

The governance of digital and algorithmic risks has therefore become a central issue. At European level, the Ethics Guidelines for Trustworthy AI established key requirements such as human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity and non-discrimination, societal and environmental well-being, and accountability [26]. The Artificial Intelligence Act later transformed part of this debate into a binding risk based legal framework, with specific requirements for high risk AI systems, including data governance, transparency, human oversight, accuracy, robustness and cybersecurity [27]. In parallel, the NIST AI Risk Management Framework proposes a structured approach to AI risk through the functions of govern, map, measure and manage [28], while algorithmic auditing literature has developed practical tools to examine AI systems throughout their lifecycle [29]. In Spain, the practical guide on algorithmic information in the workplace has also linked algorithmic governance with

labour rights, information duties, impact assessment and human review [30]. These developments confirm that digital transformation is no longer only an engineering challenge, but also a question of institutional design, organisational responsibility and social legitimacy.

Against this background, this paper examines the relationship between digital system failures, algorithmic bias and governance challenges in Industry 4.0, with particular attention to their relevance for sustainable smart cities. The main argument is that digitally enabled industrial and urban systems cannot be considered sustainable if they are efficient but fragile, automated but opaque, or innovative but weakly accountable. The paper follows a conceptual and analytical approach supported by selected secondary evidence from Spain and the European context. Its contribution is to integrate three debates that are often treated separately: cyber resilience, algorithmic governance and social sustainability. By doing so, it proposes a framework for understanding how digital fragility and algorithmic opacity may condition the long-term resilience of smart industrial and urban ecosystems.

The paper is structured as follows. Section 2 provides the literature background and conceptual framework, linking Industry 4.0 technologies with smart city infrastructures and identifying their main sustainability promises and vulnerabilities. Section 3 presents the research approach and explains the conceptual, analytical and documentary basis of the study. Section 4 analyses digital system failures and cyber resilience in interconnected industrial and urban environments. Section 5 examines algorithmic bias, opacity and accountability in data driven decision making. Section 6 proposes an integrated governance framework for cyber resilient and trustworthy digital transformation. Section 7 discusses the main implications for researchers, policy makers and practitioners, and Section 8 concludes.

2. Literature background and conceptual framework

Industry 4.0 and smart cities are often presented as separate areas of digital transformation. The first is mainly associated with manufacturing, production systems and industrial value chains, whereas the second refers to the digital management of urban infrastructures, public services and territorial resources. However, both paradigms share a common technological basis. They rely on connected devices, cyber physical systems, Internet of Things infrastructures, cloud and edge computing, large scale data processing, artificial intelligence and platform-based coordination. For this reason, the smart factory and the smart city can be understood as two expressions of a broader transition towards data intensive and interconnected systems [1], [3], [4], [5], [6].

In the industrial field, the original formulation of Industrie 4.0 emphasised the integration of physical production with digital technologies, allowing machines, products, workers and management systems to communicate across the production process [1]. Later contributions have described Industry 4.0 as a technological and organisational transformation based on cyber physical production systems, industrial connectivity, real time data, automation and decentralised decision making [2], [3], [4]. This does not mean

that production becomes fully autonomous or that human intervention disappears. Rather, it means that industrial decisions increasingly depend on the continuous interaction between physical assets and digital infrastructures. Sensors collect information from machines and processes, communication networks transmit these data, analytical systems transform them into knowledge and automated tools support decisions on maintenance, quality control, logistics, energy use and production planning.

A similar logic can be observed in the smart city literature. Smart cities are not defined only by the presence of digital technologies, but by the integration of those technologies into urban systems to improve service delivery, environmental management, mobility, safety and quality of life [5], [6]. In this context, the city becomes a complex digital ecosystem where traffic systems, energy networks, public buildings, waste services, water infrastructures and administrative platforms can be monitored and coordinated through data. Batty et al. describe future smart cities as environments increasingly supported by sensors, digital intelligence and modelling capabilities [5]. Albino et al. also stress that the smart city concept involves several dimensions, including technology, people, governance, mobility, environment and living conditions [6]. Therefore, the smart city should not be reduced to an urban technological device. It is a sociotechnical system in which infrastructure, institutions, citizens and data-based services interact.

The link between Industry 4.0 and smart cities becomes clearer when attention is placed on their enabling infrastructures. Software platforms for smart cities are commonly built around cyber physical systems, Internet of Things, Big Data and cloud computing [7]. These are also central components of advanced industrial environments. In both cases, digital transformation requires the capture, transmission, storage and interpretation of data across distributed systems. Industrial data management research shows that networked manufacturing generates large volumes of data with various levels of volume, variety, traffic and criticality, from the field level to cloud-based applications [31]. Smart cities face a comparable problem, because urban systems also generate heterogeneous data from sensors, platforms, users and public services. Research on local smart city initiatives and smart governance confirms that digital transformation in public environments depends not only on technological availability, but also on institutional capacity, interoperability, legal conditions and coordination between public services [9], [10]. The challenge is therefore to ensure dependable, interoperable and secure data flows for decision making.

This shared technological base explains why risks affecting Industry 4.0 are also relevant to smart cities. When infrastructures are connected, failures are no longer confined to isolated devices. A malfunction in a sensor, a communication error, a vulnerable platform, or an incorrect model output may affect wider chains of decisions. In a factory, this may lead to production stoppages, defective products, unsafe operations, or inefficient resource use. In a city, comparable failures may affect transport coordination, energy distribution, emergency response, public service allocation, or environmental monitoring. Security, safety, and privacy are central challenges in smart city environments, where digital infrastructures connect people, servers and physical objects [8]. This concern is not exclusive to urban systems. It is also relevant to Industry 4.0, where information technology

and operational technology increasingly converge through cyber physical systems, industrial networks, and Industrial Internet of Things infrastructures [12], [14], [15]. A balanced reading is therefore needed. Smart factories and smart cities can support efficiency, sustainability, and coordination, but they also create dependencies on data quality, connectivity, software reliability, cybersecurity and institutional capacity. Their sustainability depends not only on advanced technologies, but also on the resilience, transparency and governance of the infrastructures through which those technologies operate.

The literature on Industry 4.0 has frequently highlighted the positive effects of digitalisation on operational efficiency, flexibility and productivity. Connected production systems can improve process visibility, reduce downtime, support predictive maintenance and enable a more precise allocation of resources [3], [4]. These possibilities are also relevant from a sustainability perspective. If data are dependable and systems are correctly integrated, digital technologies may help reduce energy consumption, optimise material flows, detect inefficiencies and support more informed environmental decisions. In smart city strategies, similar expectations are associated with intelligent mobility, energy monitoring, urban planning, public service management and environmental control [5], [6], [7].

However, these promises are not automatic. They depend on the quality and robustness of the digital systems that support them. A highly connected environment can become more efficient, but also more fragile. Cyber physical systems combine hardware, software, communication networks and physical processes, so failures may emerge from multiple sources, including technical malfunction, poor integration, insecure devices, data errors or hostile attacks [12], [13]. In industrial environments, this fragility is particularly important because a local failure can affect production continuity, safety conditions and the quality of data used for later decisions. In urban environments, the same problem may affect the continuity of essential services, the reliability of public information and the capacity of institutions to respond to disruptions.

The Industrial Internet of Things reinforces this tension. Its value lies in the possibility of connecting machines, sensors, workers, platforms and external systems, but this same connectivity expands the attack surface and increases dependence on interoperable and secure architectures. Research on Industrial Internet of Things security has shown that risks are distributed across different layers, including perception, network and application layers [14], [15]. Vulnerabilities may compromise confidentiality, integrity and availability, but their consequences can go beyond information loss. In a smart factory, they may alter production parameters, interrupt industrial control systems or affect worker safety. In a smart city, they may disturb mobility systems, energy services, surveillance infrastructures or administrative platforms. The technical problem therefore becomes an organisational and social problem.

For this reason, cybersecurity is necessary but not sufficient. A narrow cybersecurity approach tends to focus on preventing unauthorised access, protecting systems and reducing exposure to attacks. These objectives remain essential, but sustainable digital transformation also requires cyber resilience. In cyber physical systems, cyber resilience refers to the capacity to prepare, absorb, recover and adapt to adverse effects associated with cyber threats while maintaining core functions [32]. This idea is particularly useful for Industry 4.0 and smart cities, because both operate through interconnected infrastructures where complete prevention is unrealistic. The relevant question is not only whether an incident can be avoided, but whether the system can continue operating, limit damage, recover functions and learn from the event. Such an approach also changes the meaning of digital sustainability. A digital system cannot be considered sustainable only because it improves efficiency under normal conditions. It must also be able to resist failures, maintain essential operations and avoid transferring digital fragility to workers, firms, citizens or public institutions. Sustainability therefore has a technical dimension, because systems need robust architectures, redundancy, monitoring and recovery capacity. It also has an organisational dimension, because firms and administrations need protocols, responsibilities, training and decision-making procedures.

Finally, it has a social dimension, because failures in automated systems may weaken trust, create perceptions of vulnerability and reduce the legitimacy of digital transformation. This reading is consistent with previous research on Industry 4.0 in Spain, which has shown that digital transformation must be assessed together with its energy, environmental and regulatory implications [33]. It is also coherent with recent empirical evidence on Spanish small and medium sized enterprises (SMEs), where labour productivity under Industry 4.0 is more closely connected with effective technological implementation than with a merely formal or declarative approach to digitalisation [34].

Algorithmic decision making introduces another layer of vulnerability. Industry 4.0 and smart city systems increasingly use data-based models to classify, predict, recommend or prioritise decisions. In industrial contexts, such systems may support maintenance, workforce planning, supplier evaluation, investment allocation, logistics and quality control. In urban contexts, they may influence mobility management, public service allocation, risk detection, energy planning or environmental monitoring. The problem is that algorithmic systems are not neutral simply because they are automated. They are designed, trained and deployed within specific social, organisational and institutional contexts.

Research on Big Data and algorithmic ethics has shown that automated systems can reproduce or amplify existing inequalities when they rely on biased historical data, non-representative samples, proxy variables or objectives that ignore distributional effects [19], [20]. Bias may also emerge from the design of the model itself. The selection of input variables, target functions, performance metrics and decision thresholds involves choices about what should be optimised and what can be treated as acceptable error. Fairness research has therefore stressed that maximising global accuracy or reducing costs may conflict with equity, contestability and the protection of affected groups [21]. In other

words, an algorithm can be technically efficient and socially problematic at the same time. O’Neil’s concept is useful for understanding why some algorithmic systems are especially risky. According to this approach, problematic models tend to share three characteristics: opacity, scale and potential harm [22]. They are opaque when affected individuals cannot understand how decisions are made or what data and criteria have been used. They operate at scale when the same model is applied repeatedly across many cases, thereby multiplying its effects. They have potential for harm when their outputs influence relevant opportunities, rights or resources without adequate accountability. Although O’Neil’s examples refer to education, credit, insurance and criminal justice, the framework can also be applied to industrial and urban digital systems.

At organisational level, algorithmic bias may affect workforce management, promotion systems, task allocation, performance assessment or supplier scoring. Research on artificial intelligence in human resource management has identified important challenges concerning fairness, data quality, accountability and employee reactions [23]. Studies on algorithms at work also show that algorithmic management may reshape organisational control, intensify surveillance and generate resistance when transparency and participation are weak [24]. Urban systems face a comparable problem. Automated decision systems may influence who receives services first, which areas are prioritised for investment, how risks are classified or how citizens interact with public institutions. When such systems are opaque or difficult to challenge, digital innovation may weaken rather than strengthen trust.

Trust in automation becomes a decisive governance condition in both settings. Lee and See define trust in automation as an attitude that allows a user to accept vulnerability to a system under conditions of uncertainty [25]. This trust needs to be calibrated. If users distrust automated systems excessively, they may underuse them, duplicate manual controls or resist digital transformation. If they trust them blindly, they may transfer systematic errors or biases into organisational decisions without sufficient scrutiny. Appropriate reliance requires information, experience, feedback and the possibility of human intervention. Algorithmic governance, accordingly, cannot be limited to technical validation. It must also include transparency, explainability, accountability, human oversight and mechanisms of contestation.

European and international frameworks have increasingly translated this concern into governance requirements. The Ethics Guidelines for Trustworthy AI identify human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity and non-discrimination, societal and environmental well being, and accountability as key requirements for trustworthy artificial intelligence [26]. The Artificial Intelligence Act adopts a risk-based approach and establishes specific obligations for high risk AI systems, including requirements on data governance, technical documentation, transparency, human oversight, accuracy, robustness and cybersecurity [27]. In parallel, the National Institute of Standards and Technology (NIST) AI Risk Management Framework proposes a structured approach to governing, mapping, measuring and managing AI risks throughout the system lifecycle [28]. Algorithmic auditing literature

also offers practical tools to examine data, models, deployment conditions and impacts before and after implementation [29].

Taken together, these contributions support a broader understanding of digital transformation governance. Industry 4.0 and smart cities require institutional and organisational arrangements capable of addressing both cyber resilience and algorithmic accountability. This is the conceptual basis of the present paper. Digital system failures and algorithmic bias are treated as connected vulnerabilities of data intensive systems. Their effects are technical, because they affect reliability and performance. They are organisational, because they influence decisions, responsibilities and trust. They are also social and ethical, because they may generate unequal impacts and reduce the legitimacy of digital innovation. From this perspective, sustainable smart cities and smart industrial environments require not only smart technologies, but also resilient infrastructures and trustworthy governance.

3. Research approach

This study follows a conceptual, analytical and documentary approach. Its purpose is to examine how digital system failures, cyber risks and algorithmic bias condition the sustainability of Industry 4.0 environments and their connection with smart and resilient cities. The analysis is based on the integration of three bodies of literature: Industry 4.0 and smart city infrastructures, cyber physical systems and cyber resilience, and algorithmic governance and trustworthy artificial intelligence. The documentary basis combines academic literature with institutional and regulatory sources. The academic literature is used to define the main concepts and to frame the relationship between digital infrastructures, cyber resilience, algorithmic bias and organisational trust. Institutional and regulatory sources are used to contextualise the analysis within the European and Spanish digital governance framework, including cybersecurity reports, artificial intelligence governance documents and legal instruments related to transparency, human oversight and accountability [16], [17], [18], [26], [27], [28], [30].

The analysis is organised around three categories. The first category is digital system failure, understood as disruptions, errors, vulnerabilities or cyber incidents affecting the digital infrastructures that support industrial and urban processes. The second category is algorithmic bias, understood as systematic and unjustified disadvantage generated by data-based decision systems. The third category is governance response, which refers to the institutional and organisational mechanisms required to strengthen resilience, transparency, accountability, human oversight and trust. This approach makes it possible to connect technical vulnerabilities with organisational and social consequences. The paper therefore treats cyber resilience and algorithmic governance as complementary conditions for sustainable digital transformation. The analytical focus is not limited to the existence of advanced technologies, but extends to the way these technologies are designed, deployed, supervised and corrected when they affect firms, workers, citizens and public institutions.

4. Digital system failures and cyber resilience in interconnected industrial and urban environments

The common technological base of Industry 4.0 and smart cities creates a shared exposure to digital disruption. Both depend on connected infrastructures, cyber physical systems, Internet of Things devices, data platforms, cloud services, analytical tools and automated decision support. The first major vulnerability considered here is the failure of the digital systems that sustain connected industrial and urban operations. The analysis focuses on how disruptions, vulnerabilities and cyber incidents may affect the continuity, reliability and sustainability of industrial and urban digital systems.

4.1. Digital system failures in interconnected environments

In Industry 4.0 environments, digital system failures should not be understood only as isolated technical errors. They include a broader set of disruptions, vulnerabilities and cyber incidents affecting the digital infrastructures that support production, monitoring, communication and decision making. These infrastructures include cyber physical systems, industrial networks, Internet of Things devices, cloud services, data platforms and advanced analytics. Their value lies in the integration of physical and digital processes, but that same integration also creates new forms of dependence and exposure.



Fig. 1. Components of digital system failures in cyber physical systems.

Source: Own elaboration.

Cyber physical systems are especially relevant in this context because they connect computation, communication and physical processes. Their operation depends on the continuous interaction between sensors, controllers, software, networks and physical assets. As a result, a failure may originate in a hardware component, a software module, a data transmission process, a communication protocol or a cyber-attack [12], [13]. In traditional production systems, many incidents could remain relatively localised. In highly connected environments, however, a local disruption can affect wider chains of operation, because different components depend on each other to maintain continuity, accuracy and control. This problem becomes more complex in Industrial Internet of Things environments. The Industrial Internet of Things expands the capacity to collect data from

machines, workers, products and processes, but it also increases the number of connected devices and the volume of interactions between operational technology and information technology [14], [15]. This generates a wider attack surface and makes system reliability dependent on interoperability, secure communication and data integrity. If a sensor sends incorrect information, if a network is interrupted or if a platform processes altered data, the consequences may affect production planning, quality control, predictive maintenance or energy management. The issue is therefore not limited to the availability of a device. It also concerns the reliability of the decisions made based on that device.

The same reasoning applies to smart city infrastructures. Urban digital systems also rely on sensors, platforms, communication networks and analytical tools to manage mobility, energy, water, waste, safety or public services [5], [6], [7]. A failure in this context may disturb traffic coordination, energy distribution, environmental monitoring, emergency response or administrative services. The link with Industry 4.0 is clear: both smart factories and smart cities transform physical processes into data dependent systems. When data flows are interrupted, manipulated or incorrectly interpreted, the effects may move from the digital layer to the physical, organisational and institutional layers. Digital system failures may therefore arise from several sources. Some are technical, such as software errors, hardware malfunction, poor integration between systems or incompatibility between devices supplied by different providers. Others are related to data, such as incomplete, inaccurate, outdated or manipulated information.

A third group derives from cybersecurity threats, including unauthorised access, ransomware, denial of service attacks, malicious code, node capture or attacks affecting communication channels [14], [15], [16]. In practice, these categories often overlap. A cyber incident may produce data errors, a data error may trigger an incorrect automated response, and an incorrect response may generate operational, economic or safety consequences. From a sustainability perspective, these failures matter because they can neutralise part of the expected benefits of digital transformation. Industry 4.0 is usually associated with efficiency, monitoring and optimisation, but these benefits depend on the stability and reliability of the digital infrastructure. If the system fails, production may stop, resources may be wasted, energy consumption may increase or safety conditions may deteriorate. In smart cities, similar failures may weaken service continuity, reduce institutional response capacity and affect citizens' trust in digital public systems. Digital fragility is therefore not only a technical limitation. It is also a condition that affects economic continuity, environmental efficiency and social legitimacy.

This is why cyber resilience becomes a necessary complement to cybersecurity. Cybersecurity focuses on protecting systems against threats, reducing exposure and preventing unauthorised actions. Cyber resilience adds a further requirement, the capacity to prepare for disruption, absorb its effects, recover critical functions and adapt after the event [32]. In interconnected industrial and urban environments, this broader perspective is essential because complete prevention is unlikely. The relevant challenge is to ensure that digital systems can continue operating under stress, limit cascading effects and restore their functions without transferring excessive costs to firms, workers, citizens or public

institutions. For this reason, digital system failures should be treated as a structural issue in the governance of Industry 4.0 and smart cities. The question is not only whether a technology is advanced, efficient or automated, but whether the infrastructure that supports it is robust, auditable, recoverable and institutionally governed. In this sense, cyber resilience is not an additional technical feature. It is a condition for sustainable digital transformation.

4.2. Cyber risks and selected evidence from Spain and the European context

Recent cybersecurity evidence helps to place these failures in a more concrete context. Although aggregated cybersecurity indicators do not refer exclusively to Industry 4.0 or smart city environments, they provide a useful context for understanding the exposure of firms, public institutions and citizens to digital disruption. In highly connected systems, cyber incidents are not only a matter of information loss. They may affect service continuity, operational reliability, economic activity and institutional trust. At European level, ENISA identifies ransomware, distributed denial of service attacks, social engineering, data related threats and supply chain compromise as major components of the current threat landscape [16]. This distribution matters here because smart cities depend heavily on public digital infrastructures, while Industry 4.0 depends on the resilience of private and industrial systems. In Industry 4.0, this means that cyber incidents can interfere with production planning, industrial control, predictive maintenance or energy optimisation. In smart cities, similar incidents may affect mobility systems, public services, energy networks, administrative platforms or emergency response.

In Spain, INCIBE reported that its CERT managed 83,517 cybersecurity incidents in 2023, representing a 24.00% increase compared with 2022. More than 58,000 incidents affected citizens and more than 22,000 affected private companies, including SMEs and self-employed workers. INCIBE also identified 183,077 vulnerable systems exposed to possible attacks and managed 237 incidents affecting essential and critical operators, including sectors such as energy, transport and the financial system [17]. These figures show that cyber risk is not confined to large technological infrastructures. It also affects productive organisations, public services and everyday digital interactions. The Spanish Ministry of the Interior reported 472,125 recorded cybercrime offences in 2023, with computer fraud representing 90.50% of the total [18]. From another institutional perspective, the Observatory of Risks for Companies in Spain estimated 192,774 cyber incidents managed in 2023 by three Computer Emergency Response Teams: CCN CERT, INCIBE CERT and ESDEF CERT. According to this source, 55.90% corresponded to the public sector and 43.30% to the private sector, while the remaining share affected the defence sector [35]. This distribution is relevant for the present analysis because smart cities depend heavily on public digital infrastructures, while Industry 4.0 depends on the resilience of private and industrial systems. Beyond incident measurement, research on cybercrime prevention campaigns documents the role of public communication and governmental digital awareness initiatives in cyber risk reduction, especially when citizens and institutions interact through digital channels [36].

Table 1. Selected cybersecurity evidence relevant to cyber resilient Industry 4.0 and smart cities

Indicator	Value	Main relevance
INCIBE CERT cybersecurity incidents	83,517 incidents in 2023	Shows the scale of digital exposure affecting citizens, firms and critical operators
Vulnerable systems identified by INCIBE	183,077 vulnerable systems	Indicates the presence of exposed devices and services that may affect connected infrastructures
Incidents affecting essential and critical operators	237 incidents	Connects cybersecurity with sectors that sustain industrial and urban continuity
Cyber incidents managed by CCN CERT, INCIBE CERT and ESDEF CERT	192,774 incidents in 2023	Provides an aggregated view of public, private and defence related exposure
Recorded cybercrime offences in Spain	472,125 offences in 2023	Shows the wider social and economic relevance of digital risk
Computer fraud within cybercrime	90.50% of recorded cybercrime offences	Highlights the predominance of fraud in the cybercrime structure

Source: Own elaboration based on INCIBE [17], Ministerio del Interior [18] and Institut Cerdà [35].

Fig. 2 summarises two complementary dimensions of the Spanish evidence. First, it shows the distribution of cyber incidents managed by public CERTs, where the public sector represents the largest share, followed by the private sector. Second, it shows the structure of recorded cybercrime, where computer fraud accounts for most offences. These two patterns reinforce the same interpretation: digital risk affects both institutional infrastructures and market-based activities and therefore has direct relevance for smart city governance and Industry 4.0 resilience.

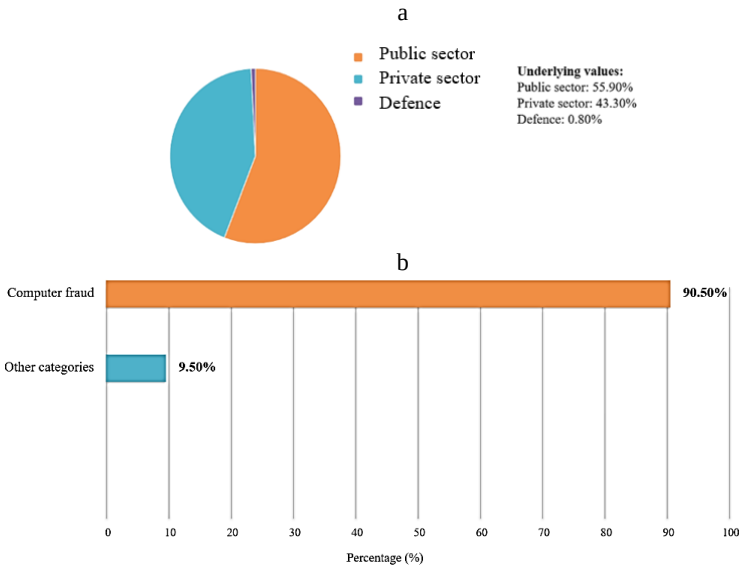


Fig. 2. (a) Distribution of cyber incidents managed by CCN CERT, INCIBE CERT and ESDEF CERT by sector. Spain, 2023 (%). (b) Distribution of recorded cybercrime offences by main category. Spain, 2023 (%).

Source: Own elaboration based on Ministerio del Interior [18] and Institut Cerdà [35].

Note: "Other categories" in Fig. 2(b) is calculated as the remaining share after computer fraud.

The evidence reinforces the need to treat cyber resilience as a sustainability condition. A digitally enabled system may be efficient under normal operating conditions, but it becomes vulnerable if it cannot maintain essential functions under disruption. In industrial environments, this vulnerability may generate production stoppages, additional costs, resource waste, safety risks or loss of trust in automation. In urban environments, it may affect essential services, public response capacity and confidence in digital governance. For this reason, resilience cannot be reduced to technical protection. It requires preparedness, redundancy, monitoring, recovery capacity and clear institutional responsibility. Research on resilient information security management links Business Continuity Management in public institutions with the capacity to maintain essential services under disruption [37]. Overall, the Spanish and European evidence points to systemic rather than occasional cyber risk. These risks affect citizens, firms, public administrations and critical operators simultaneously. This is especially important for sustainable smart cities, because urban resilience depends on the continuity of services that are increasingly mediated by digital infrastructures. It is also important for Industry 4.0, because smart factories depend on the availability, integrity and reliability of data flows. In both cases, digital failure may become operational failure, and operational failure may become economic, environmental and social damage.

5. Algorithmic bias, opacity and accountability in data driven decision making

After digital system failures, the second major vulnerability considered in the paper is algorithmic bias. In Industry 4.0 and smart city systems, decisions are increasingly supported by models that classify, predict, recommend or prioritise actions. These models can improve coordination and efficiency, but they may also reproduce hidden inequalities, generate unfair outcomes or make decisions difficult to understand and contest. For this reason, algorithmic bias is not treated here as an isolated technical error, but as a governance problem with organisational, ethical and social consequences.

5.1. Algorithmic bias in industrial and urban data driven decisions

Algorithmic bias can be understood as a systematic and unjustified distortion in the results produced by data-based decision systems. It appears when an automated or semi-automated model generates disadvantageous outcomes for certain individuals, groups, firms or territories because of biased data, design choices or deployment conditions [19], [20]. The problem is especially relevant in Industry 4.0 and smart city environments because both rely on data driven systems to organise complex processes. Within firms, algorithms may support workforce planning, performance assessment, supplier evaluation, predictive maintenance, logistics, investment allocation or quality control. In urban governance, they may influence mobility management, public service distribution, energy planning, risk detection, administrative prioritisation or environmental monitoring.

The first source of algorithmic bias is usually found in the data used to train, validate or operate the model. Historical data are not neutral records of reality. They may reflect previous organisational practices, unequal access to opportunities, underrepresentation of

some groups or errors in the way information has been collected. Research on Big Data and discrimination has shown that automated systems may generate disparate impacts even when they do not explicitly use sensitive variables, because proxy variables can indirectly reproduce previous inequalities [19]. In Industry 4.0, this may occur when historical productivity, absenteeism, promotion or supplier performance data are used without considering the organisational conditions that produced them. In smart cities, similar risks may arise when public service, mobility or security data reflect uneven territorial coverage, unequal reporting patterns or previous administrative priorities.

A second source of bias is linked to model design. The selection of input variables, target functions, performance metrics and decision thresholds involves choices about what the system should optimise. A model designed only to maximise efficiency, accuracy or cost reduction may produce results that are technically strong but socially problematic. Fairness research has shown that different definitions of fairness may lead to different outcomes, and that global accuracy can hide unequal error distribution across groups [21]. In industrial environments, this is relevant when algorithms rank workers, suppliers or investment alternatives. In urban environments, it matters when models prioritise areas, classify risks or allocate resources among neighbourhoods. In both cases, the model does not merely describe reality. It participates in the organisation of decisions. A third source of bias appears during use and interpretation. Even when the data and model are technically acceptable, the way the system is introduced into an organisation can shape its effects. Algorithmic tools are used by managers, public officials, technical teams and platform providers, each with different levels of knowledge, responsibility and power. If the model is opaque, if affected people cannot understand its criteria, or if there are no effective channels for review, biased outputs may become routine decisions. This is why algorithmic bias should be analysed as part of a wider sociotechnical system, where code, data, institutions and organisational practices interact [20], [24].

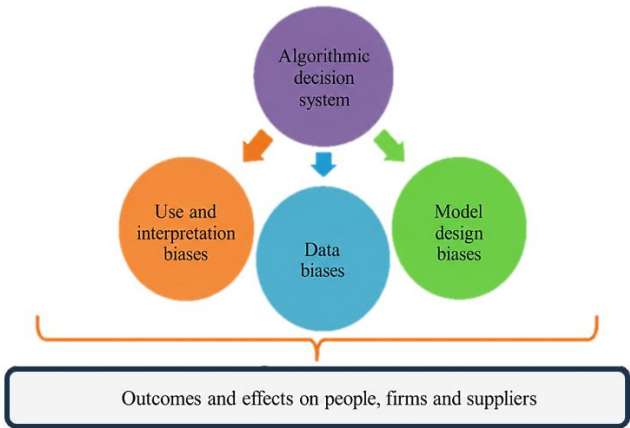


Fig. 3. Sources of algorithmic bias in data-based decision systems.
Source: Own elaboration.

Algorithmic bias matters because it links automated decision making with sustainability and resilience. If algorithmic systems produce systematic disadvantage, weaken trust or

limit the capacity of affected people to challenge decisions, they may undermine the social legitimacy of digital transformation. The connection with digital system failures is direct: poor data quality, cyber incidents or system errors may feed biased models, while opaque models may make it harder to detect altered or misinterpreted data. Algorithmic governance and cyber resilience should therefore be treated as complementary conditions for trustworthy digital transformation.

5.2. *Opacity, scale and potential harm*

The risks described above become more serious when algorithmic systems operate under conditions of opacity, scale and potential harm. O’Neil’s notion of “weapons of math destruction” is useful for this paper because it identifies a specific type of algorithmic system whose effects are not only technical, but also institutional and social [22]. These systems are problematic because affected people cannot easily understand how decisions are made, because the same model can be applied repeatedly across many cases, and because its results may influence relevant opportunities, rights or resources without sufficient accountability. In Industry 4.0, opacity may appear when algorithmic tools are supplied by external providers and integrated into production planning, predictive maintenance, workforce management or supplier evaluation. In these cases, the firm may not have full visibility over the data used, the variables selected, the optimisation criteria or the internal logic of the model. For workers, suppliers or smaller firms affected by these systems, opacity may be even greater. Decisions may appear as technical outputs rather than contestable organisational choices. Scale is also relevant. A single algorithmic logic can be deployed across several plants, production lines, warehouses, suppliers or territorial units.

In smart city environments, a similar logic may be applied across neighbourhoods, mobility systems, administrative procedures or public service platforms. If the model incorporates biased data or poorly defined criteria, its effects can be repeated and amplified. The problem is therefore not only that a specific decision may be wrong, but that the same decision pattern may be reproduced systematically. Potential harm emerges when algorithmic systems influence decisions that affect employment, working conditions, supplier access, investment priorities or the distribution of public services. In industrial environments, this may include recruitment, performance assessment, task allocation, promotion, supplier scoring or investment selection. In smart cities, it may involve service prioritisation, risk classification, mobility planning or the allocation of public resources. Under these conditions, algorithmic bias may undermine trust in digital transformation and weaken the social legitimacy of smart industrial and urban strategies.

Table 2. Opacity, scale and potential harm in Industry 4.0 and smart city algorithmic systems

Feature	General meaning	Application to Industry 4.0 and smart cities
Opacity	The functioning of the system, the data used and the decision criteria are not transparent to affected people or even fully visible to the organisation using the system	Algorithmic tools for planning, maintenance, human resources, supplier scoring or urban service allocation may be developed by external providers, making their criteria difficult to understand, audit or contest

Feature	General meaning	Application to Industry 4.0 and smart cities
Scale	The system is applied repeatedly across many cases, amplifying its effects	The same model may operate across several plants, production lines, suppliers, neighbourhoods or public service platforms, reproducing the same decision logic across different contexts
Potential harm	The system influences relevant opportunities, rights or resources without clear accountability mechanisms	Algorithms may affect working conditions, promotion, supplier selection, investment allocation, access to services or territorial priorities, creating persistent disadvantages for certain profiles, firms or areas

Source: Own elaboration based on O'Neil [22].

This framework is particularly relevant for the present paper because it connects algorithmic bias with cyber resilience and sustainability. Previous research on Industry 4.0 in Spain has shown that digital transformation cannot be assessed only through the adoption of advanced technologies but must also be interpreted in relation to its economic, energy, environmental and regulatory implications [33]. Recent empirical evidence on Spanish SMEs also links productivity effects more closely to effective technological implementation than to a merely formal or declarative approach to digitalisation [34]. This reinforces the idea that the value of Industry 4.0 depends not only on technological availability, but also on the quality of its organisational and governance conditions.

Under weak oversight, opaque and large-scale algorithmic systems may reduce the benefits of digital transformation. They may generate distrust among workers, increase asymmetries between large firms and smaller suppliers, or prioritise short term efficiency over social and environmental objectives. In smart cities, comparable risks may arise when automated systems classify risks, allocate resources or prioritise services without sufficient transparency or participation. Therefore, opacity, scale and potential harm should be treated as governance risks that affect both smart industry and smart city strategies.

5.3. Accountability, human oversight and organisational trust

The risks associated with algorithmic bias cannot be addressed only through technical improvement. Better data, more accurate models and stronger validation procedures are necessary, but they are not sufficient when algorithmic systems affect workers, suppliers, citizens or public institutions. In data driven environments, governance must also define who is responsible for the system, who can understand its logic, who can question its outputs and how errors or unfair effects can be corrected. Accountability, human oversight and organisational trust therefore become central conditions for the sustainable deployment of Industry 4.0 and smart city technologies. Trust in automation is especially important in this context. Lee and See argue that appropriate reliance on automated systems depends on calibrated trust, neither excessive distrust nor blind acceptance [25]. This idea is directly relevant for Industry 4.0 and smart cities. If workers, managers or public officials do not trust algorithmic systems, they may avoid using them, duplicate manual controls or resist digital transformation. If they trust them without sufficient scrutiny, errors and biases may be transferred mechanically into organisational or institutional decisions. Sustainable

digital transformation therefore requires transparency, feedback, human review and clear channels for contesting decisions.

Table 3. Governance requirements for trustworthy algorithmic systems in Industry 4.0 and smart cities

Governance source	Main requirement	Relevance for Industry 4.0 and smart cities
Ethics Guidelines for Trustworthy AI [26]	Human agency, technical robustness, privacy, transparency, non-discrimination, societal well-being and accountability	Provides the ethical basis for trustworthy AI systems in industrial and urban decision making
Artificial Intelligence Act [27]	Risk based regulation, data governance, transparency, human oversight, robustness and cybersecurity for high-risk AI systems	Connects algorithmic governance with binding legal obligations, especially where AI affects employment, safety or access to services
NIST AI Risk Management Framework [28]	Govern, map, measure and manage AI risks across the lifecycle	Supports a practical risk management approach from design to deployment and monitoring
Algorithmic auditing literature [29]	Review of data, models, documentation, deployment conditions and impacts	Provides tools to detect bias, errors and accountability gaps before and after implementation
Spanish guide on algorithmic information in the workplace [30]	Information duties regarding parameters, rules and instructions of algorithms affecting employment conditions	Links algorithmic governance with labour rights, transparency and organisational trust

Source: Own elaboration based on [26], [27], [28], [29], [30].

Taken together, these sources show that algorithmic governance is no longer a purely technical issue. It involves ethical principles, legal obligations, risk management procedures, audit mechanisms and labour related information duties. This is particularly relevant for Industry 4.0, where algorithmic tools may influence task allocation, performance assessment, promotion, workforce planning, supplier scoring or monitoring. It is also relevant for smart cities, where automated systems may affect service prioritisation, risk classification, territorial allocation of resources or interaction with public institutions. Accountability and human oversight should therefore be understood as practical governance mechanisms. They require organisations to document the purpose of algorithmic systems, define responsibilities, assess risks, monitor outcomes and provide review procedures when decisions are contested.

They also require a human capacity to intervene when automated outputs are inconsistent, unfair or insufficiently justified. This does not mean rejecting automation. It means ensuring that automation remains explainable, supervised and aligned with social, organisational and sustainability objectives. Research on AI integration in smart cities shows that organisational culture and institutional readiness influence whether artificial intelligence can be implemented successfully in complex public sector environments [38]. This governance dimension is reinforced by previous research on Industry 4.0 in Spain, where digital transformation has been analysed not only as a technological process, but also

in relation to its economic, energy, environmental and regulatory implications [33]. It is also confirmed by empirical evidence showing that the productivity effects of Industry 4.0 depend on effective technological implementation, rather than on a merely formal approach to digitalisation [34]. In this sense, algorithmic governance should be seen as part of the conditions that allow digital technologies to generate value without undermining trust, fairness or social legitimacy.

The analysis developed in this section shows that algorithmic bias, opacity and accountability are not secondary issues. They affect the way digital systems distribute opportunities, risks and responsibilities. In Industry 4.0, they may shape labour relations, supplier networks and investment decisions. In smart cities, they may affect public services, territorial priorities and citizens' trust in digital institutions. For this reason, trustworthy digital transformation requires more than cybersecurity and technical efficiency. It also requires transparent, accountable and human supervised algorithmic systems.

6. Integrated governance framework for cyber resilient and trustworthy digital transformation

Digital transformation in Industry 4.0 and smart cities cannot be governed through separate responses to separate risks. Cyber incidents, system failures, biased data, opaque models and weak accountability mechanisms interact within the same digital ecosystem. A disruption in data infrastructures may affect the quality of automated decisions, while an opaque algorithmic system may make it more difficult to detect errors, contest outcomes or assign responsibility. For this reason, cyber resilience and algorithmic governance should be understood as complementary conditions for sustainable digital transformation. The framework proposed in this section integrates four dimensions: digital vulnerabilities, algorithmic vulnerabilities, sustainability impacts and governance responses. Its purpose is not to replace existing cybersecurity or artificial intelligence governance models, but to organise them within a common analytical structure adapted to interconnected industrial and urban environments.

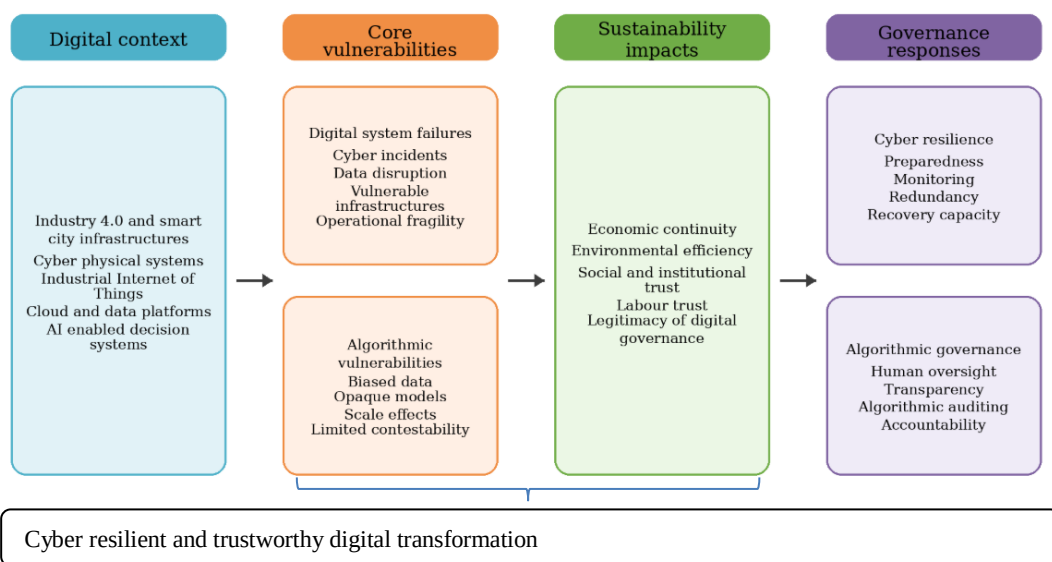


Fig. 4. Integrated governance framework for cyber resilient and trustworthy digital transformation.
Source: Own elaboration based on the analytical framework developed from [12]–[30], [33] and [34].

The first dimension refers to digital vulnerabilities. These include system failures, cyber incidents, vulnerable infrastructures, data disruption and the fragility of connected environments. In Industry 4.0, these risks may affect production planning, predictive maintenance, industrial control, energy management or quality assurance. In smart cities, they may affect mobility systems, public services, emergency response, administrative platforms or environmental monitoring. The common issue is that both industrial and urban systems increasingly depend on continuous and reliable data flows.

The second dimension concerns algorithmic vulnerabilities. These include biased data, model design choices, opaque decision rules, limited contestability and the amplification of errors through scale. In industrial contexts, algorithmic systems may affect workers, suppliers, investment priorities or the distribution of tasks and opportunities. In urban contexts, they may influence service prioritisation, risk classification, territorial allocation of resources or interaction with public institutions. These risks show that the governance of digital transformation cannot be reduced to infrastructure protection. It must also address the way automated decisions are produced, explained and reviewed.

The third dimension captures sustainability impacts. Digital vulnerabilities may affect economic sustainability when they interrupt production, increase costs or reduce operational reliability. They may affect environmental sustainability when failures generate inefficient resource use, additional energy consumption or material losses. They may also affect social sustainability when they reduce trust, increase perceptions of unfairness or weaken the ability of workers, firms or citizens to question decisions. This is why cyber resilience and algorithmic governance should be interpreted as sustainability conditions, not only as technical or legal safeguards.

The fourth dimension refers to governance responses. A cyber resilient and trustworthy digital transformation requires technical robustness, but also institutional and organisational mechanisms. These include cybersecurity preparedness, redundancy, recovery capacity, data governance, algorithmic auditing, human oversight, transparency, accountability and channels for contestation. The evidence and literature reviewed in this paper show that these mechanisms should not be applied separately. Cyber resilience is weakened if algorithmic systems remain opaque, and algorithmic governance is incomplete if the infrastructures that feed the models are insecure, unreliable or poorly monitored.

At its core, the framework rests on a central proposition: sustainable digital transformation requires alignment between infrastructure resilience and decision accountability. Industry 4.0 and smart cities may generate efficiency and productivity gains, but these gains are fragile if systems cannot resist disruption or if automated decisions undermine trust. Previous research on Industry 4.0 in Spain has already shown that digital transformation must be interpreted together with its energy, environmental and regulatory implications [33]. Recent empirical evidence on Spanish SMEs also indicates that productivity gains are associated with effective technological implementation, rather than with a merely formal approach to digitalisation [34]. The framework proposed here extends this reasoning by adding cyber resilience and algorithmic accountability as necessary governance conditions. This is particularly relevant in the adoption of advanced AI systems, where research on agentic artificial intelligence shows that readiness, resistance and security sensitivity influence deployment conditions across different domains [39].

From a practical perspective, the framework implies that firms and public institutions should evaluate digital transformation through three questions. First, whether the infrastructure is resilient enough to maintain essential functions under disruption. Second, whether the data and algorithms used for decision making are transparent, auditable and subject to human oversight. Third, whether the expected efficiency gains are compatible with social trust, environmental responsibility and institutional accountability. If one of these dimensions is absent, digital transformation may remain technologically advanced but weakly sustainable. The integrated framework also helps to connect Industry 4.0 and smart cities. Both depend on connected infrastructures and automated decisions. Both promise efficiency, sustainability and better coordination. Both are exposed to digital failures, cyber incidents and biased algorithmic systems. For this reason, the governance of smart factories and smart cities should be developed through a shared logic of resilience, transparency and accountability. In this sense, cyber resilient Industry 4.0 is not only an industrial requirement. It is also a condition for sustainable and trustworthy smart cities.

7. Discussion and implications

The analysis developed in this paper contributes to the debate on Industry 4.0 and smart cities by shifting the focus from technological deployment alone to the governance conditions under which digital transformation becomes sustainable. Much of the literature has emphasised the capacity of cyber physical systems, industrial Internet of Things, data platforms and artificial intelligence to improve efficiency, flexibility, productivity and

coordination [1], [3], [4]. These contributions remain essential, but they do not fully address the vulnerabilities that emerge when industrial and urban systems become increasingly dependent on connected infrastructures and automated decisions. The evidence and framework presented here show that digital transformation should be assessed not only through its expected performance gains, but also through its resilience, transparency and accountability.

First implication concerns research on Industry 4.0. Previous studies have shown that advanced technologies can reshape industrial value creation and support more data intensive forms of production [3], [4]. Recent empirical evidence on Spanish SMEs also indicates that productivity effects are linked more closely to effective technological implementation than to a merely formal approach to digitalisation [34]. The present paper extends this argument by showing that implementation quality should not be understood only in technological or organisational terms. It also requires resilient infrastructures, reliable data flows, accountable algorithmic systems and mechanisms for human oversight. In this sense, cyber resilience and algorithmic governance should be incorporated more explicitly into future research on digital transformation, productivity and industrial sustainability.

For smart city research, the paper reinforces the idea that urban digitalisation is not only a matter of sensors, platforms or service optimisation. Smart cities depend on the same kind of interconnected infrastructures that sustain smart factories: data collection systems, communication networks, cloud services, analytical tools and automated decision support [5], [6], [7]. This creates opportunities for better coordination, but also shared exposure to disruption, cyber incidents, data errors and opaque decision making. The contribution of the paper is therefore to connect the industrial and urban dimensions of digital risk. Smart industry and smart cities should not be analysed as separate digital worlds, because both are part of a wider sociotechnical ecosystem. This argument is reinforced by studies that treat smart city initiatives, smart governance and the transition from smart systems to smart governance as processes shaped by institutional capacity, public administration, legal frameworks and the quality of digital governance [9], [10], [11].

A second implication concerns policy makers. The European and Spanish institutional context already shows a clear movement towards cybersecurity, trustworthy artificial intelligence, human oversight and algorithmic accountability [16], [26], [27], [28], [30]. However, these areas are often governed through separate policy instruments. Cybersecurity strategies tend to focus on incident prevention, detection and response, while artificial intelligence governance focuses on transparency, risk classification, data quality, human supervision and accountability. The framework proposed in this paper supports a more integrated connection between these two agendas. Digital infrastructures feed algorithmic systems, and algorithmic systems depend on the quality, reliability and security of those infrastructures. Weakness in one dimension may therefore compromise the other. From a regulatory perspective, the Artificial Intelligence Act provides an important basis for addressing high risk AI systems through requirements related to data governance,

transparency, human oversight, robustness and cybersecurity [27]. The Ethics Guidelines for Trustworthy AI and the NIST AI Risk Management Framework also offer useful principles and procedures for managing AI related risks across the lifecycle [26], [28]. In the Spanish context, the guide on algorithmic information in the workplace links algorithmic governance with labour rights and information duties [30]. These developments point in the right direction, but their practical effectiveness will depend on how firms and public institutions translate them into internal procedures, documentation, audit practices and channels for contestation.

A third implication concerns practitioners, including firms, public administrations, technology providers and organisations involved in digital transformation projects. The analysis shows that technological adoption cannot be treated as an end in itself. Organisations implementing Industry 4.0 technologies or smart city systems need to ask whether the infrastructure can maintain essential functions under disruption, whether data quality is sufficient, whether algorithmic outputs can be explained and reviewed, and whether responsibility is clearly assigned when failures or unfair effects occur. These questions are practical rather than abstract. They affect production continuity, worker trust, supplier relations, public service reliability and institutional legitimacy.

At organisational level, the paper highlights the importance of avoiding both blind trust and excessive distrust in automation. Appropriate reliance requires users to understand the purpose, limits and expected behaviour of automated systems [25]. When managers, workers or public officials lack information about algorithmic criteria, automated outputs may be accepted without sufficient scrutiny or rejected without adequate assessment. Both situations reduce the value of digital transformation. For this reason, training, documentation, feedback and review procedures should be treated as part of the digital infrastructure, not as secondary administrative tasks.

The discussion also has implications for sustainability. Industry 4.0 and smart city strategies are often justified through their potential to optimise resources, reduce inefficiencies and improve decision making. Previous research on Industry 4.0 in Spain has already stressed the need to connect digital transformation with energy, environmental and regulatory considerations [33]. Recent research on the working environment under Industry 4.0 also shows that occupational safety and training remain core conditions for a balanced, resilient and sustainable industrial transition [40]. The present paper adds that sustainability also depends on resilience and accountability. A system that is efficient under normal conditions but fragile under disruption cannot be considered fully sustainable. Similarly, an algorithmic system that improves coordination but creates opaque or unfair outcomes may weaken social legitimacy and organisational trust.

The main limitation of the paper is its conceptual and documentary nature. The study does not estimate causal effects, build a quantitative index or test the impact of specific cyber incidents or algorithmic systems on firms, workers or citizens. Its contribution lies instead in integrating evidence and literature that are often treated separately. Future research could build on this framework through empirical studies, comparative indicators, sectoral case

studies or firm level analyses of cyber resilience and algorithmic governance. Particularly useful would be studies examining how firms and public institutions document algorithmic systems, audit data quality, respond to cyber incidents and maintain trust in automated decision making.

Overall, the discussion confirms the central argument of the paper: digital transformation cannot be considered sustainable only because it is technologically advanced. In Industry 4.0 and smart city environments, sustainability requires infrastructures that can resist and recover from disruption, and algorithmic systems that can be explained, supervised and contested. Cyber resilience and algorithmic accountability should therefore be treated as complementary pillars of trustworthy digital transformation.

8. Conclusions

This paper has examined a central tension in Industry 4.0 and smart city development. Digital transformation can improve efficiency, monitoring capacity, productivity and coordination, but it also introduces vulnerabilities that may affect the continuity, fairness and legitimacy of digitally enabled systems. The analysis has focused on two connected risks: digital system failures and algorithmic bias. Both are relevant for Industry 4.0 environments, but they also extend to smart cities because industrial and urban systems increasingly depend on connected infrastructures, continuous data flows, automated decisions and platform-based coordination.

The findings show that digital system failures should not be treated as isolated technical incidents. In interconnected environments, failures may affect production continuity, data reliability, energy use, service provision, safety conditions and institutional response capacity. The evidence from Spain and the European context confirms that cyber risks are not occasional or marginal. They affect citizens, firms, public administrations and critical operators, and therefore have direct implications for smart industry and smart city governance. Cyber resilience consequently becomes a condition for sustainability, because a system that is efficient under normal conditions but fragile under disruption cannot be considered fully sustainable.

The paper has also shown that algorithmic bias is not merely a technical problem of model accuracy. It is a governance challenge. Bias may emerge from historical data, model design choices, proxy variables, opaque decision rules or weak review mechanisms. When algorithmic systems operate at scale and affect employment, supplier selection, investment allocation, public services or territorial priorities, their consequences may become persistent and difficult to contest. Under these conditions, automated decision making can weaken trust, reproduce organisational asymmetries and reduce the social legitimacy of digital innovation.

The main contribution of the paper is the integrated governance framework proposed in Section 6. This framework connects digital vulnerabilities, algorithmic vulnerabilities, sustainability impacts and governance responses within a single analytical structure. Its

core argument is that sustainable digital transformation requires alignment between resilient infrastructures and accountable automated decisions. This conclusion has practical implications. Firms, public administrations and technology providers should not assess digital transformation only through adoption rates, automation levels or expected efficiency gains. They should also examine whether digital infrastructures can maintain essential functions under disruption, whether data and models are auditable, whether human oversight is effective, and whether affected workers, firms or citizens have meaningful channels to understand and contest decisions. In this sense, transparency, accountability, human supervision and recovery capacity are not external constraints on innovation. They are conditions that allow innovation to generate durable value.

Overall, the paper argues that Industry 4.0 and smart cities should not be evaluated only by their technological sophistication. Their long-term sustainability depends on whether they can remain reliable under disruption, fair under automation and accountable under complexity. Cyber resilient Industry 4.0 is therefore not only a matter of protecting systems. It is a governance condition for building trustworthy, sustainable and socially legitimate smart cities.

References

- [1] H. Kagermann, W. Wahlster and J. Helbig, Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group. Frankfurt am Main: acatech, Joint publications, 2013.
- [2] H. Lasi, P. Fettke, H. G. Kemper, T. Feld and M. Hoffmann, "Industry 4.0," *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014.
- [3] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," *Journal of Industrial Information Integration*, vol. 6, pp. 1–10, 2017.
- [4] L. D. Xu, E. L. Xu and L. Li, "Industry 4.0: State of the art and future trends," *International Journal of Production Research*, vol. 56, no. 8, pp. 2941–2962, 2018.
- [5] M. Batty, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis and Y. Portugali, "Smart cities of the future," *The European Physical Journal Special Topics*, vol. 214, pp. 481–518, 2012.
- [6] V. Albino, U. Berardi and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," *Journal of Urban Technology*, vol. 22, no. 1, pp. 3–21, 2015.
- [7] E. F. Z. Santana, A. P. Chaves, M. A. Gerosa, F. Kon and D. S. Milojevic, "Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture," *Future Generation Computer Systems*, vol. 70, pp. 3–44, 2017.
- [8] A. S. Elmaghraby and M. M. Losavio, "Cyber security challenges in smart cities: Safety, security and privacy," *Journal of Advanced Research*, vol. 5, no. 4, pp. 491–497, 2014.
- [9] R. Y. Maulana, U. Marjamat, D. Subekti and A. Wicaksono, "Catalyzing digital transformation through Smart Cities initiatives in Indonesian local government," *Smart Cities and Regional Development (SCRD) Journal*, vol. 9, no. 1, pp. 35–46, 2025.
- [10] M. Namysłowska and A. Olbryk, "Smart governance in Poland: The case study of the city of Łódź," *Smart Cities and Regional Development (SCRD) Journal*, vol. 9, no. 3, pp. 81–97, 2025.
- [11] C. Vrabie, "From Smart Systems to Smart Governance," *Smart Cities and Regional Development (SCRD) Journal*, vol. 10, no. 2, article C2, 2026.
- [12] A. Humayed, J. Lin, F. Li and B. Luo, "Cyber physical systems security: A survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802–1831, 2017.
- [13] J. P. A. Yaacoub, H. N. Noura, O. Salman and A. Chehab, "Cyber physical systems security: Limitations, issues and future trends," *Microprocessors and Microsystems*, vol. 77, 2020.
- [14] K. P. Tange, M. De Donno, X. Fafoutis and N. Dragoni, "A systematic survey of Industrial Internet of Things security: Requirements and fog computing opportunities," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2489–2520, 2020.

- [15] B. Alotaibi, "A Survey on Industrial Internet of Things Security: Requirements, Attacks, AI-Based Solutions, and Edge Computing Opportunities," *Sensors*, vol. 23, no. 17, article 7470, 2023.
- [16] ENISA, ENISA Threat Landscape 2023. European Union Agency for Cybersecurity, 2023. <https://www.enisa.europa.eu/sites/default/files/publications/ENISA%20Threat%20Landscape%202023.pdf>
- [17] INCIBE, "Los incidentes de ciberseguridad de 2023 gestionados por INCIBE aumentan en un 24%," Instituto Nacional de Ciberseguridad, 2024. <https://www.incibe.es/incibe/sala-de-prensa/los-incidentes-de-ciberseguridad-de-2023-gestionados-por-incibe-aumentan-en>
- [18] Ministerio del Interior, Informe sobre la cibercriminalidad en España 2023. Madrid: Ministerio del Interior, 2024. https://www.interior.gob.es/opencms/pdf/archivos-y-documentacion/documentacion-y-publicaciones/publicaciones-descargables/publicaciones-periodicas/informe-sobre-la-cibercriminalidad-en-Espana/Informe-Cibercriminalidad_2023_126200212_pdfWEB.pdf
- [19] S. Barocas and A. D. Selbst, "Big Data's disparate impact," *California Law Review*, vol. 104, no. 3, pp. 671–732, 2016.
- [20] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, vol. 3, no. 2, 2016.
- [21] S. Verma and J. Rubin, "Fairness Definitions Explained," in *Proceedings of the IEEE/ACM International Workshop on Software Fairness*, 2018, pp. 1–7.
- [22] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.
- [23] P. Tambe, P. Cappelli and V. Yakubovich, "Artificial intelligence in human resources management: Challenges and a path forward," *California Management Review*, vol. 61, no. 4, pp. 15–42, 2019.
- [24] K. C. Kellogg, M. A. Valentine and A. Christin, "Algorithms at work: The new contested terrain of control," *Academy of Management Annals*, vol. 14, no. 1, pp. 366–410, 2020.
- [25] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.
- [26] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission, 2019.
- [27] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, *Official Journal of the European Union*, 2024. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- [28] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework, AI RMF 1.0*. Gaithersburg, MD: NIST, 2023.
- [29] I. D. Raji, A. Smart, R. N. White, M. Mitchell, T. Gebru, B. Hutchinson, J. Smith-Loud, D. Theron and P. Barnes, "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing," in *Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency*, 2020, pp. 33–44.
- [30] Ministerio de Trabajo y Economía Social, *Información algorítmica en el ámbito laboral: Guía práctica y herramienta sobre la obligación empresarial de información sobre el uso de algoritmos en el ámbito laboral*. Madrid: Ministerio de Trabajo y Economía Social, 2022. https://webtrabajouk.empleo.gob.es/ficheros/ministerio/inicio_destacados/Guia_Algoritmos_ES.pdf
- [31] T. P. Raptis, A. Passarella and M. Conti, "Data Management in Industry 4.0: State of the Art and Open Challenges," *IEEE Access*, vol. 7, pp. 97052–97093, 2019.
- [32] M. Segovia-Ferreira, J. Rubio-Hernan, A. R. Cavalli and J. Garcia-Alfaro, "A Survey on Cyber-Resilience Approaches for Cyber-Physical Systems," *ACM Computing Surveys*, vol. 56, no. 8, article 202, pp. 1–37, 2024.
- [33] S. García-Moreno and V.-R. López-Ruiz, "A Review of the Energy Sector as a Key Factor in Industry 4.0: The Case of Spain," *Energies*, vol. 16, no. 11, article 4446, 2023.
- [34] S. García-Moreno and V.-R. López-Ruiz, "Modelling Labour Productivity Under Industry 4.0: Digitalisation, Automation, and Firm Maturity," *Mathematics*, vol. 14, no. 11, article 1855, 2026.
- [35] Institut Cerdà, *Observatorio de riesgos para las empresas en España 2025*. Barcelona: Institut Cerdà, 2025. <https://www.icerda.org/wp-content/uploads/2025/01/Institut-Cerda-Observatorio-de-Riesgos-4-Ed-2025-V.pdf>

- [36] A. F. Rahmat, C. Vrabie and G. B. Soesilo, “Exploring the Cybercrime Prevention Campaign on Twitter: Evidence from the Indonesian Government,” *Smart Cities and Regional Development (SCRD) Journal*, vol. 7, no. 2, pp. 9–24, 2023.
- [37] C. Schachtner, “On the way to resilient information security management: Business Continuity Management (BCM) for public institutions,” *Smart Cities and Regional Development (SCRD) Journal*, vol. 9, no. 1, pp. 7–15, 2025.
- [38] R. M. Benshams, “Cultivating organizational culture for AI integration: A framework for Smart Cities and Regional Development,” *Smart Cities and Regional Development (SCRD) Journal*, vol. 9, no. 4, pp. 125–137, 2025.
- [39] G. Waizel, “The agentic AI security adoption matrix: Understanding readiness and resistance across domains,” *Smart Cities and Regional Development (SCRD) Journal*, vol. 10, no. 2, pp. 75–84, 2026.
- [40] S. García-Moreno and V.-R. López-Ruiz, “Digital Transformation and the Working Environment: International Evidence on Safety and Training in Industry 4.0,” *Journal of Knowledge Dynamics*, vol. 3, no. 1, 2026.