Foundations of Management, Vol. 13 (2021)
eISSN 2300-5661, pISSN 2080-7279, DOI: 10.2478/fman-2021-0008

DIGITAL INNOVATIONS AND SMART SOLUTIONS FOR SOCIETY AND ECONOMY: PROS AND CONS

Marcin SIKORSKI

Gdańsk University of Technology, Faculty of Management and Economics, POLAND
e-mail: marcin.sikorski@zie.pg.edu.pl

Abstract: Recent developments in artificial intelligence (AI) may involve significant potential threats to personal data privacy, national security, and social and economic stability. AI-based solutions are often promoted as "intelligent" or "smart" because they are autonomous in optimizing various processes. Because they can modify their behavior without human supervision by analyzing data from the environment, AI-based systems may be more prone to malfunctions and malicious activities than conventional software systems. Moreover, due to existing regulatory gaps, development and operation of AI-based products are not yet subject to adequate risk management and administrative supervision. Resonating to recent reports about potential threats resulting from AI-based systems, this paper presents an outline of a prospective risk assessment for adaptive and autonomous products. This research resulted in extensive catalogs of possible damages, initiating events, and preventive policies that can be useful for risk managers involved in conducting risk assessment procedures for AI-based systems. The paper concludes with the analysis and discussion of changes in business, legal, and institutional environments required to ensure the public that AI-based solutions can be trusted, are transparent and safe, and can improve the quality of life.

Keywords: digital innovations, digital services, artificial intelligence, smart services, risk assessment, risk management.

JEL Classification: M10.

1 Introduction[1]

Businesses and organizations constantly search for new development opportunities by investing in digital technologies. Information Technology (IT) systems, business intelligence software and cloud services have become a backbone of contemporary economics, administration, and social life.

We are witnessing a shift from product- to service-oriented businesses, which has been largely enabled by advancements in digital innovations (Lasi, et al., 2014). The original term "digital innovation" proposed by Yoo, et al. (2010) relates to novel IT solutions, for instance, regarding autonomous robots, vehicles, and software-based services, to which terms such as "intelligent," "adaptive," or "smart" are now often used to address their cognitive-like capabilities.

Artificial intelligence (AI) is the key software technology that enables computer-controlled devices to learn and behave in an adaptive, "intelligent" manner. AI is gaining popularity in many application areas, for instance, adaptive control systems in engineering and industry, automated decision-making in business and health diagnostics, or fraud detection in financial services.

The concept of "smartness" was thoroughly explored by Romero, et al. (2020), who analyzed its many notions depending on the type of a specific solution, object, or technology. After that, we deal with the smart systems defined as specific IT solutions, which have the capability for autonomous, self-controlled learning, decision-making, and adopting their behavior to a specific context. Because what was once a digital innovation soon becomes a commonly used smart service or solution, the generic term Smart Digital Solutions (SDS) will be used in this paper to describe a broad class of "intelligent" solutions increasingly present in our everyday lives.

[1] This article was submitted at the "Digital Economy – Management, Innovation, Society and Technology" Conference 2020 (DEMIST'20) held on November 17, 2020 (http://demist.eu/).

AI-based "intelligent" solutions, essential for SDS, fundamentally differ from conventional software applications. They perform operations in the cloud, exchange data with other systems, process reasoning without a human operator, and remain invisible to their users unaware of how exactly a specific SDS works. Most importantly, SDS so far operate beyond sufficient regulatory supervision, basing largely on software designers' belief that the AI would learn and self-adapt as expected, making decisions adequate to the context.

Nevertheless, each technology sometimes fails. In large-scale application areas, such as traffic control, monitoring anomalies in global financial markets, or automated image recognition for public safety, the costs of suboptimal machine-based decisions can be very high. Similarly, a machine-based misdiagnosis in heath treatment or erroneous actions performed by AI algorithms on financial assets may cause severe damages for individuals, businesses, and organizations.

Most recently, in the light of reported incidents about malfunctioning AI, such as accidents caused by self-driving vehicles, human trust in AI-based SDS cannot be taken for granted. In many countries, experts and agencies attempt to draw public attention to potential threats, including the opportunity for unauthorized re-programming, hacking, sabotage, or using AI-based solutions as a tool for crime.

The doubts regarding SDS and other AI-based solutions include, for instance:

- insufficient human control on the adaptive process of machine learning,
- lack of transparency and explainability in why specific decisions were made,
- limited users' trust as to the validity of machine-based diagnosis or decisions, and
- insufficient regulatory framework for assuring the public that AI-based components and systems are secure, reliable, and trusted.

As a result, forecasting the possible impacts of SDS on business, administration, or society is still a much indefinite area.

The objective of this paper is to present an outline of a prospective risk assessment process for SDS, focusing on two main aspects:

- identifying the basic risk factors related to SDS: categories of possible damages, initiating events, and risk-preventive policies, and
- specifying the required circumstances and preconditions for successful adoption of risk assessment practices from industry to the area of AI-based applications and SDS in particular.

2 Related research

Many studies on the possible impacts of AI-based adaptive systems (e.g. Bughin, et al., 2017; Castro and New, 2016; Purdy and Daugherty, 2016) show primarily the expected advantages and benefits, for instance:

- reducing human cognitive load by adaptive automation, intelligent robots,
- performance improvements in manufacturing, business, transport, and logistics;
- intelligent decision-making, using big data for behavioral and cognitive predictions,
- automated image recognition for public security, and
- personalized customer experience in e-commerce, chatbots, recommender systems, and generating tailored offers.

Nevertheless, accomplishing benefits from AI-based systems requires collecting a large amount of data, for instance, on individual customers' actual behavioral patterns. Therefore, despite the fact that often enthusiastic messages are received from business and industry circles, the public gets increasingly concerned by the social threats resulting from potential using AI tools for malicious activities (Millar, et al., 2017; Schneiderman, 2016).

Available literature and reports authored by think-tanks, expert groups, or advisory bodies (e.g. EAF, 2015; Bowser, 2017; Müller and Bostrom, 2016; Campolo, et al., 2017; Mehr, 2017; Walsh, 2017; Allianz, 2018; Brundage, et al., 2018; Desouza, 2018; Villani, 2018) specify four main areas where destructive AI may endanger stability and security on a national level:

1) business and engineering;
2) social and economic;
3) legal ethical, and cultural;
4) political, governmental, and defense.

These areas can be targeted by many categories of threats, for instance:

- AI-related side effects: sudden incidents and long-term impacts to infrastructures, organizations, or markets, including unforeseen problems of compliance with the existing law,
- AI-based systems hacked by humans: SDS operation overtaken by hackers to steal data or do any other type of harm,
- AI-related human negligence: allowing self-made modifications by unsupervised learning or any malfunctioning due to a human error in programming or software maintenance, and
- AI as a crime tool: SDS deliberately programmed by a human to be destructive or used for criminal behavior.

Because SDS malfunctions may have severe economic and social impacts, and official reports on specific AI-related incidents are sporadic, there is a noticeable deficit of information for the public about how trusted and reliable SDS actually are. This deficit is obviously incomparable to areas such as the safety of transport, engineering machinery, health equipment, food, or pharmaceuticals, which are subject to legal regulations specifying how customers should get informed by manufacturers and service providers.

Regarding the origin of threats related to AI-based systems, Henfridsson, et al. (2018) and Holmström (2018) point out lack of theoretical fundaments for designing AI-based systems and their quick and agile development process where testing is very limited and usually based on a small set of training data. These weaknesses usually result in neglecting the evaluation of potential risks that a specific AI-based solution may bring to the society if abused or hijacked for any unauthorized use.

Many industries routinely conduct comprehensive risk assessments for various components of their infrastructure and operational activities. However, in the IT business, it is usually limited to identifying the risks endangering a specific project's success and not covering the risks and threats related to a specific IT product, especially to adaptive and autonomous ones, such as SDS.

Based on selected reviews of the available risk assessment methodologies (Rovins, et al. 2015; EC, 2011; Kumar, 2010; Habegger, 2008; Voros, 2003; Rowe and Wright, 1999), the following approaches could be applicable to SDS- and AI-based products:

1) Quantitative, data-based approach

Widely used for risk management in industry and engineering, where probabilistic input data are usually more available than in other fields. For SDS, except tree-based propagation methods borrowed from cybersecurity, this approach is not very feasible; there is no systematic collection of data for AI-related incidents, so their probability distributions remain largely unknown.

2) Qualitative, expert-based approach

The expert-based qualitative approach is advantageous in situations where hard data are lacking, but predicting development scenarios and estimating rough likelihoods are more valuable than producing exact numerical predictions. The most popular qualitative methods include:

- the Delphi method: a moderated, questionnaire-based process of iterative data collection and analysis designed to search for consensus among the anonymous experts, and
- the Foresight method: an iterative process that explores the human capacity to think ahead and envision responses to face future social and technological challenges.

3) Semiquantitative, expert-based approach

In this approach, human experts act as cognitive agents capable of identifying threats, estimating their sources and the scale of possible impacts. Numerical estimations of risk index are calculated using indicator-based methods such as:

- scoring methods: FMEA, Risk Score, HAZOP, or nomograms for computing a specific risk index value,
- graphical methods: risk matrix, maps; or graphs, which identify specific risks and allocate them to categories linked with the required types of managerial actions,

- Analytic Hierarchy Process (AHP): an intuitive decision support technique, based on a series of pairwise comparisons, producing a visual ranking of risk-related alternatives such as actions, policies, or design solutions.

Among the above, the semiquantitative approach seems to be the most suitable option for prospective risk assessment for SDS. For AI-based systems, probabilistic data are lacking, but experts' experiences from related areas often can be used for semiquantitative estimation of values of risk-scoring indicators. Sikorski (2020) presented a pilot study in which the semiquantitative approach combined an expert panel with the AHP-based procedure for risk assessment of AI-based solutions and linking results with potential risk-mitigating strategies.

Currently, the following gaps seem to be significantly limiting opportunities for adopting systematic risk assessment for SDS- and AI-based systems:

- shortage of empirical probabilistic data about AI-related incidents,
- lack of risk assessment procedures dedicated to autonomous IT products such as SDS, and
- unreadiness of IT business and local regulatory institutions for monitoring AI-related challenges.

The remaining parts of this paper aim to present an outline of a prospective risk assessment process for SDS, build upon a basic catalog of risk-related factors, and specify the required circumstances to enable established risk assessment practices to be transferred from traditional industries' AI-related businesses.

3 Methodology

A prospective risk assessment process for SDS should follow leading security frameworks such as ISO/IEC 27005 (2018) and NIST 80-300 (2012), which define typical stages of risk assessment for engineering and business continuity management:

- context establishment: identification of assets and threats,
- risk modeling: identification and estimation,
- evaluation: risk analysis and treatment, and
- implementation: risk monitoring and review.

This study is aimed to cover only selected aspects of this process, namely:

- identifying specific risk factors related to SDS: categories of possible damages, initiating events, and risk-preventive policies, and
- specifying the required circumstances and preconditions for successfully adopting risk assessment practices from industry to the area of AI-based applications and SDS in particular.

The straightforward research procedure applied for this research covers:

Step 1: Identification of damage areas and possible initiating events ("triggers");

Step 2: Identification of possible risk-preventing policies; and

Step 3: Linking specific risk factors to adequate risk-preventing policies (programs, actions) according to the score values of risk impact factors.[2]

Steps 1 and 2, critical for commencing the prospective risk assessment process for SDS, resulted in producing three catalogs for (1) possible damage areas, (2) possible initiating triggers (triggers), and (3) risk-preventing policies.

Steps 1 and 2 were performed as a desk research procedure, covering the following activities:

1) A systematic review of published sources such as:
   - academic literature: authored research papers, journals, and books, and
   - gray literature: analytic reports published by business, government, and academic organizations and relatively sparse media reports on incidents related to autonomous systems.
2) Collecting data items (such as examples, actions, or events) in three categories: possible damage areas, possible triggers, and risk-preventing policies.
3) Clearing and refining the contents of categories by rephrasing the items and removing redundancies or unmeaningful elements.
4) Clustering and classification of data items using affinity diagrams and sticky cards, which resulted

[2] Step 3 remains beyond the scope of this paper; it has been largely addressed in Sikorski (2020).

in a hierarchical, two-level structure serving as the basic model for the three categories mentioned above: possible harm, triggers, and prevention policies.

In this study, only a two-level, simple hierarchic model was applied (level 1: Category; level 2: Items), neglecting the possible internal and cross-category relationships among items.

The method used in this procedure can be described as a fully manual bottom-up modeling, from collecting single data items (events, incidents, and malfunctions reported in available literature) to clustering them into specified three categories.

The manual method was deliberately applied for data exploration, clustering, classification, and synthesis of available textual materials. Although initially, the use of dedicated software for text analysis was also considered, this option was eventually excluded for the following reasons:

- A thorough search over all available texts was not the primary aspiration of this study; instead, it was rather identification of items to be classified into categories suitable for further use by experts. After all, in each evaluation with incomplete data, human expertise remains subjective, but it adds a unique predictive value lacking in computer-based procedures.
- A set of relevant data available online is unlimited and constantly enlarging by newly appearing publications; for this reason, extracting appropriate information will always be incomplete. So, manual techniques, although less efficient than routinized machine-based procedures, are more helpful in selecting important aspects by utilizing expert's experience and intuition.
- Last but not least, because this study intended to perform a viable job useful for initial exploration of the problem, considering the size of this study and the workload needed for categories coding with software-supported analysis, choosing the manual mode seemed to be a reasonable decision.

For a similar study with a larger scope, available software tools could be surely used for qualitative text analysis. Nevertheless, it is hard to estimate their impact on the validity of results; for instance, manual selection of category coding remains a significant subjectivity factor in computer-supported qualitative analysis also.

4 Results

4.1 Catalogs of risk-related factors

For preparing the foundation of a risk assessment process for SDS, the following deliverables were developed using a procedure described in the section "Methodology":

1) Catalog of damages

The category term "Damages" describes possible damage areas (level 1 – 5 groups), aggregated from examples of possible losses or destructions (level 2 – 36 items) specified in the right-hand column of Table 1.

Table 1. Categories of damages (*Source:* Own elaboration)

Damages (level 1)	Description (level 2)
Social and political	– Undermining public order and trust to state, businesses, and society – Affecting AI-based governments, justice, etc. – Generating false recommendations, judgments, and decisions – State abusing the use of automated electronic surveillance – Automated AI-based censorship online – Social manipulation for rebel or pro-government campaigns – Social trust put on fabricated entities interacting online like humans – Malicious hijacking online campaigns – Impersonalized, anonymous, distant relation to state or institutions

Table 1. Categories of damages, cont. (*Source:* Own elaboration)

Damages (level 1)	Description (level 2)
Physical and material	– IT-initiated crashes and disruptions (caused or accidental) – Generating false alarms and panic – Remote or delayed attack operations – Robots disabling or entering security zones and damaging infrastructures – Machine-based false judgments and decisions leading to material loss – Human sabotage and damage of automated surveillance equipment
Business and economic	– Disruption of markets or regional economies – Paralyzing important institutions – Manipulations in social media for discrediting business brands – Business-oriented manipulations aimed at affecting conjuncture – Reputational damages, erosion of trust – Financial losses and damages due to malicious activities online – Criminal, legal, or insurance problems
Individual and private	– AI used for streamlining users from/to specific content – AI-propelled emotional scam (dating, financial, etc.) – Privacy violations, data breach – AI-based medical misdiagnosis, physical/health damages – AI-based abusive profiling of users, patients, or consumers – Undermined personal trust to state, businesses, and society – Self-imposed auto-censorship due to ubiquitous online surveillance – Fabricated evidences (videos) in media or in judicial cases – Personal addiction to digital platforms (social, entertainment, etc.)
Defense and security	– Using AI to accessing classified information – Using AI to attack critical infrastructure, command centers – Overtaking control, mimicking human operators – Creating a panic, provoking conflicts affecting national security – AI-controlled robots disabling national security

2) Catalog of triggers

The term “Triggers” describes a category of events, actions, or agents, whose activity may lead to specific damages. Categories of triggers (level 1 – 5 groups) were aggregated from examples (level 2 – 38 items) specified in the right-hand column of Table 2.

Table 2. Categories of triggers (*Source:* Own elaboration)

Triggers (level 1)	Description (level 2)
System malfunction	– Allowing AI to use incorrect or incomplete input data – Technological flaws resulting in suboptimal decisions or control actions – Poor quality of AI: faulty machine learning, inadequate supervision – Attacks self-initiated by AI, self-initiated modification of software – Lack of explainability, transparency, and traceability of AI software – Learning and adaptation of AI software is beyond human control
Hacking and hijacking	– Dual use of AI software: for terrorism, hijacking, overtaking control – Automated fabricating of data, news for blackmailing or discrediting – Swamping information channels with noise – AI-based prioritizing of attack targets, automated vulnerability discovery – Open code, open algorithms, destructive tools easier to develop – Human reprogramming AI for malicious use – Corrupting algorithms by disgusted employees or external foes – Hijacking autonomous vehicles or software robots (overtaking control) – Building and deploying malicious bots or robots – Nanobots for deploying toxins to the environment or living bodies
Social manipulation	– Fake news for destabilizing, manipulating elections – Automated social engineering attacks – Malicious chatbots mimicking humans, chatbots pretending as friends – Automated influence campaigning (elections, shopping, etc.) – Automated scam and targeted blackmail – Social bots propagating or draw-in to extreme/hysteric groups – Malicious streamlining of users to/from a specific content
Business greed	– Greed, rush, releasing untested, unvalidated software – Ignorance or recklessness of business leaders or companies – No governance, no supervision, no ethics related to AI – No AI-related risk management activities – No recovery plans for AI-related damages/impacts – No forecasting/assessment of social effects caused by AI
Regulatory gaps	– No dedicated consumer protection from AI (smart) products – No control/registry of AI software applications – Lack of coordinated supervision or one responsible body on a national level – Leaders unaware of or ignoring the opinions of experts – Poor awareness of customers with regard to AI-caused harms – No systematic risk analysis, no forecasting, no foresights – No lessons learned from reported incidents – No risk identification performed as to the social impact of AI – AI-related gaps in the legal system, lacking standards and procedures

3) Catalog of preventive policies

The term "Policies" (level 1 – 6 groups) describes a category of possible interventions – risk-related preventive or mitigating actions, projects, programs, or strategies (level 2 – 42 items) specified in the right-hand column of Table 3. If adequately selected and correctly executed, these policies should reduce the impact of known risks to an acceptable level.

Table 3. Categories of preventive policies (*Source:* Own elaboration)

Policies (level 1)	Description (level 2)
Fixing technology	– Monitoring systems and behaviors, early detection of hackers – Adapting cybersecurity techniques to smart systems – Compromising attackers (buy-in) – AI tools used reversely – for security and defense – "Red-teams" forecasting malicious activities for security, fraud, or abuse
Public awareness	– Educating consumers about threats from "smart" products – Expert bodies to be heard louder than now – Publishing case studies on incidents and threats affecting real life – Presenting AI with a balanced view, objective tone, and no hype – Expert bodies answering questions from consumers – Promoting consumer rights to have smart systems safe and validated – Educating consumers in critical thinking as to biased or fake news – Providing free tools for validating credibility of news and media sources
Social approach	– Promoting ethical AI to engineers and prospective developers (students) – Interdisciplinary design teams able to assess social impact – Including new (public and social) stakeholders into design process – Feeding from social sciences, not only from tech domains – Promoting mandatory assessment of the social impact of AI applications
Business governance	– Rewarding ethical and sustainable governance in AI business companies – Implementing supervised design, deployment and operation of AI – Assuring AI compliance to regulations (auditing, certificates) – Assigning process owners and leadership in AI business governance – Company monitoring assessments of the social impact of AI – Promoting explainability and traceability of AI algorithms
Regulatory framework	– Improving the regulatory framework for technological solutions – Establishing a repository of AI-related incidents and damages – Assigning one major AI-regulatory institution on the national level – Formalizing communication: regulators, governments, and AI business – Legal requirements for auditing, certification, and verification of AI – Intelligence involved in monitoring AI-related incidents and damages – Protecting AI against unauthorized reverse engineering and decoding

Table 3. Categories of preventive policies, cont. (*Source:* Own elaboration)

Policies (level 1)	Description (level 2)
Controls and measures	– Hardware supply chain control: hardware manufacturers and distributors – Software supply chain control for critical AI components – Mandatory registration and insurance for robots/drones/vehicles – Regulatory institutions make pressure on governments to update the law – Standardized security barriers to airspace and other open spaces – Assigning one major AI regulatory institution on the national level – Automated detections and automated interventions – Surveillance of and moderating social media, public health discourse – Banning specific AI technologies from authoritarian governments – Pervasive use of total encryption – Technical tools for detecting malicious bots, fake news, and forgeries

While Tables 1–3 show a considerably expanded version of the catalogs presented in Sikorski (2020), they are fundamental for commencing the prospective risk assessment process for SDS outlined hereafter.

4.2 The outline of a risk assessment process for SDS

A prospective risk assessment process for SDS follows leading industrial security frameworks such as ISO 27005 (2018) and NIST 80-300 (2012), with following steps:

1) Identification of risk factors: damages, triggers, and preventive policies;
2) Evaluation: analysis and estimation of consequences, likelihoods, and other risk-related parameters;
3) Presentation of assessment results: visualization and interpretation;
4) Operationalization: formulating adequate risk-preventive policies, strategies, actions;
5) Implementing selected policies and monitoring their results.

These activities should be performed in a repetitive cycle and in a systematic manner using regularly updated catalogs of risk factors. This process should be conducted by the teamwork of experts also representing areas beyond AI and IT engineering. Although to some extent formalized, the whole process should be moderated by one of the experts, also representing the team to external stakeholders or customers.

Resources that need to be provided to the expert team include:

- on-site and remote teamwork environment,
- catalogs of risk factors (predefined or elaborated ad hoc),
- methods for risk scoring, agreed beforehand and familiar to all experts,
- tools for visualization, presentation, and interpretation of results, and
- administrative support as needed.

After evaluation, the operationalization phase starts with a presentation of results to the internal or external customer; then follows the collaborative work with the customer or a relevant committee to formulate adequate risk-preventive policies, strategies, and actions. It is also possible that the external customer may carry out this part internally, without the participation of external experts. Subsequently, implementing these policies should be conducted by respective organizations or institutions (like system owners or regulators) and subject to administrative supervision.

The proposed risk assessment process for SDS needs to be performed with SDS developers. It should follow up the guidelines aimed at introducing AI governance principles as proposed by EU (2020), EU (2021), and OECD (2021), covering the entire SDS lifecycle from the initial design to deployment and operation.

4.3 Implications for businesses and institutions

For the successful implementation of a systematic risk assessment process for SDS- and other AI-based solutions, several factors are essential:

1) Establishing a consistent legal framework

According to van Berkel, et al. (2020), in Europe, there are too many fragmented and localized perspectives to AI. A synchronized European framework, accompanied by relevant national regulations, is badly needed to coordinate the security and trustworthiness of SDS- and other AI-based solutions. Such a framework should include mandatory and voluntary audits, reviews, and assessments, located in coordinated national strategies and policies.

2) Specifying the role of local telecommunication regulators

National regulatory institutions, cooperating with local government agencies, should take the supervisory role upon SDS and AI applications' entire lifecycle. Risk assessment and monitoring activities should be performed within an existing legal framework and in compliance with appropriate cybersecurity procedures and practices established at the international level (EU, 2020).

3) Creating a jointly recognized liability framework

This framework, addressing the issues of insurance, liability and accountability, should cover the entire AI-based lifecycle, from the concept to deployment and operation (Allianz, 2018). It is important to specify an adequate scope of liability for bodies responsible for the design, operation, and maintenance and for informing the public. In this framework, differing liability regulations will be required for autonomous vehicles and transport systems, industrial machinery (like autonomous robots) and SDS systems processing consumer and personal data. Moreover, the established liability framework should be jointly recognized beyond a national level as to how current insurance regulations should work for businesses and individual customers.

4) Reframing the evaluation practices in AI-related IT projects

Leveraging AI-related guidelines and principles to the operational level in IT projects requires gaining acceptance and support of IT businesses. Regarding AI components and systems, requirement specifications, testing and evaluation procedures will need to be expanded with the issues specified by the EU White Paper on AI (EU, 2020) and the EU AI Strategy (EU, 2021), for instance, assuring robustness throughout the lifecycle, assuring reproducibility of behavior, and providing transparency and resilience against malicious attacks or data manipulations. Eventually, SDS suppliers are expected to bear responsibility not only for the quality of design, but also for the quality in use, which is fundamentally different from the current customary responsibilities of software manufacturers.

5) Educating the public for AI presence in social life

The public (citizens, consumers, business owners, or employers) should get prepared for AI-related changes expected in social life and primarily in the labor market (Ahmed, 2018). Relevant activities should be shared among specific bodies, such as regulators and educational institutions, as well as media and consulting agencies cooperating with IT and AI businesses.

Educating young people (students, teenagers, and children) early is essential to comprehend AI capabilities long before they enter the AI-intensive labor market. Specialists familiar with AI will be needed for design and development, deployment and integration, and legal and security issues in many application areas. AI-related education should be included not only in institutional teaching programs, but also in AI-related programming contests organized for young innovators and inventors.

Availability of funding for AI-related research projects, business initiatives, start-ups, and cooperation networks is also an important element for attracting young people to the AI field and for stimulating successful innovative entrepreneurship beyond local or national markets.

5 Discussion

Through exploration of selected research literature, analytic reports, and national AI policy documents, this paper:

- elaborated extensive catalogs of risk-related factors, fundamental for conducting risk assessments for SDS- and other AI-based systems, and
- highlighted the need for incorporating a systematic risk assessment process into AI development processes for SDS, based on the above catalogs and localized in a specific legal and institutional environment.

Certainly, this study was not free from limitations, some of which are as follows:

- The content of catalogs was extracted relying on subjective human expertise and remains subject to changes due to newly arriving knowledge, ongoing regulatory activities, and human creativity in inventing malicious deeds; moreover, data items contained in each catalog may be interrelated, which was neglected in this study.
- Projecting the general outline for a prospective process of risk assessment without the possibility of validating it in industrial practice; only a part of this process was validated in a pilot study described by Sikorski (2020).
- Generality of the projected outline, resulting from the fact that IT industry, legal framework, and institutional environment are not yet prepared for conducting risk assessment for SDS in a way similar to cybersecurity procedures; furthermore, existing national and supranational policies for AI development and oversight remain purely postulative so far, without being effectively used in practical regulatory procedures applicable to AI projects and enterprises.

Nevertheless, in addition to providing extended catalogs of risk factors, this article appears to benefit researchers and practitioners by analyzing the current challenges faced by business, legal, institutional, and educational environments in reassuring the public that an SDS will function as safe, controlled, and trustworthy.

The problem of how to convince the public that AI-based solutions are free from excessive risk was not the subject of this paper. However, it can be hypothesized that the competitive advantage in profiting from the lucrative market of "intelligent" solutions will be accomplished firstly by strong industrial and high-tech brands, already recognized by consumers for their:

- long experience in supplying reliable and trustworthy products to demanding industries such as healthcare, automotive, aviation, military, cybersecurity, or critical infrastructures,
- recognizable corporate governance, including no involvement in abusing consumer rights or conducting unethical campaigns, and
- supportive online communities, advocating the brand and active in recommending specific SDS as the ones proved to be safe, transparent, and improving the quality of life.

6 Conclusions

The risks related to SDS are still perceived as high due to the black-box nature of AI-based solutions. Uncertainty is also prevalent in the public about their possible negative impact on human privacy, public security, business, and social life.

This paper attempts to emphasize that providing IT companies with simple-to-use risk assessment methods is essential for assuring the public that SDS- and other AI-based products are trustworthy and can be safely deployed to daily operations. Transparency and explainability (Arrieta, et al., 2020) are crucial for the SDS owners from the viewpoint of liabilities resulting from faulty operations caused by AI algorithms.

Furthermore, facilitating broad acceptance of AI-based products largely depends on consumers' trust in institutions accountable for screening and auditing AI development. Derisking AI, as defined by Baquero, et al. (2020), is a shared responsibility of business organizations and regulatory institutions. Business executives are expected to redefine their strategies and governance concepts for ethical use of AI (Albinson, et al., 2019) and implement them in their projects and processes.

Subsequently, relevant regulatory institutions are urged to convert national AI policies and regulations (OECD, 2021) into operational frameworks acceptable by the business. Last but not least, properly balanced regulatory actions should not only benefit business, but also help to limit the spread of common misconceptions about AI.

7 References

[1] Ahmed, K., 2018. *Bank of England Chief Economist Warns on AI Jobs Threat*. [online] Available at: https://www.bbc.com/news/business-45240758 [Accessed 12 February, 2021].

[2] Albinson, N., Balaji, S., and Chu, Y., 2019. *Building Digital Trust: Technology Can Lead the Way*. Deloitte Insights, https://www2.deloitte.com/lu/en/pages/innovation/articles/building-long-term-trust-in-digital-technology.html.

[3] Allianz, 2018. *The Rise of Artificial Intelligence: Future Outline and Emerging Risks.* Allianz Global. https://www.agcs.allianz.com/news-and-insights/reports/the-rise-of-artificial-intelligence.html.

[4] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F., 2020. *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI*. Information Fusion, Volume 58, pp.82-115, https://doi.org/10.1016/j.inffus.2019.12.012

[5] Baquero, J.A., Burkhardt, R., Govindarajan, A., and Wallace, T., 2020. *Derisking AI by Design: How to Build Risk Management into AI Development*. McKinsey Analytics. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development.

[6] Berkel van, N., Papachristos, E., Giachanou, A., Hosio, S. and Skov, M.B., 2020. *A Systematic Assessment of National Artificial Intelligence Policies: Perspectives from the Nordics and Beyond*. In Proceedings of the 11th Nordic Conference on Human-Computer Interaction: (NordiCHI '20). Association for Computing Machinery, New York, NY, USA, Article 10, pp.1-12. https://doi.org/10.1145/3419249.3420106.

[7] Bowser, A., Sloan, M., Michelucci, P. and Pauwels, E., 2017. *Artificial Intelligence: A Policy-oriented Introduction*. Wilson Center Technology and Innovation Program. https://wilsoncenter.org/publication/artificial-intelligence-policy-oriented-introduction [Accessed 1 June 2021].

[8] Brundage M., et al. (26 others), 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Future of Humanity Institute, Oxford, UK. https://maliciousaireport.com/.

[9] Bughin, J. Hazan, E., Ramaswamy, S., Chui, M., Alla, T., Dahlstrom, P, Henke, M. and Trench, M., 2017. *Artificial Intelligence. The Next Digital Frontier?* McKinsey Global Institute. https://doi.org/APO-210501.

[10] Campolo A., Sanfilippo M., Whittaker M. and Crawford K., 2017. *AI Now 2017 Report*. AI Now Institute. https://ainowinstitute.org/AI_Now_2017_Report.pdf.

[11] Castro, D., New, J., 2016. *The Promise of Artificial Intelligence*. Center for Data Innovation. https://datainnovation.org/2016/10/the-promise-of-artificial-intelligence/.

[12] Desouza K.C., 2018. *Delivering Artificial Intelligence in Government: Challenges and Opportunities*. IBM Center for The Business of Government. http://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf.

[13] EAF (Effective Altruism Foundation), 2015. *Artificial Intelligence: Opportunities and Risks*. https://ea-foundation.org/artificial-intelligence/ [Accessed 24 May 2021].

[14] EC, 2011. *Risk Assessment and Mapping Guidelines for Disaster Management.* European Commission. Commission staff working paper, European Union. https://ec.europa.eu/jrc/en/publication/recommendations-national-risk-assessment-disaster-risk-management-eu.

[15] EC, 2020. *White Paper On Artificial Intelligence - A European Approach to Excellence and Trust*. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[16] EU, 2021. *A European Approach to Artificial Intelligence*. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

[17] Habegger, B. (ed.), 2008. *The International Handbook on Risk Analysis and Management*. Center for Security Studies at ETH Zurich. https://www.files.ethz.ch/isn/47029/hb_riskanalysis&management.pdf.

[18] Henfridsson, O., Nandhakumar, J., Scarbrough, H. and Panourgias, N., 2018. Recombination in the Open-ended Value Landscape of Digital Innovation. *Information and Organization*, 28(2), pp.89-100.

[19] Holmström, J., 2018. Recombination in Digital Innovation: Challenges, Opportunities, and the Importance of a Theoretical Framework. *Information and Organization*, Volume 28(2), pp.107-110. https://doi.org/10.1016/j.infoandorg. 2018.04.002.

[20] ISO/IEC 27005, 2018. *Information Technology − Security Techniques − Information Security Risk Management*. International Standard. International Organization for Standardization. Geneva, Switzerland.

[21] Jobin, A., Ienca, M. and Vayena, E., 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, 1(9), pp.389-399. https://doi.org/10.1038/s42256-019-0088-2.

[22] Kumar, D.P., 2010. Managing Project Risk Using Combined Analytic Hierarchy Process and Risk Map. *Applied Soft Computing*, Volume 10, Issue 4, pp.990-1000. https://doi.org/10.1016/j.asoc.2010.03.010.

[23] Lasi, H., Fettke, P., Kemper, H.-G. , Feld, T. and Hoffmann, M., 2014. Industry 4.0. *Business and Information Systems Engineering*, 6 (4), pp.239-242.

[24] Mehr, H., 2017. *Artificial Intelligence for Citizen Services and Government*. Cambridge, MA: Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf [Accessed 4 June 2021].

[25] Millar, C., Lockett, M. and Ladd, T., 2017. *Disruption: Technology, Innovation and Society*. Technological Forecasting and Social Change. https://doi.org/129.10.1016/j.techfore.2017.10.020.

[26] Müller, V.C. and Bostrom, N., 2016. Fundamental Issues of Artificial Intelligence. A Survey of Experts Opinion. In: Müller (ed.). *Fundamental Issues of Artificial Intelligence*. Berlin: Springer,. pp.553-571. https://nickbostrom.com/papers/survey.pdf.

[27] NIST 800-30, 2012. *Guide for Conducting Risk Assessments*. NIST Special Publication 800-30 Revision 1. September 2012. US National Institute of Standards and Technology. http://dx.doi.org/10.6028/NIST.SP.800-30r1.

[28] OECD, 2021. *National AI Policies & Strategies*. OECD Policy Observatory. https://www.oecd.ai/dashboards.

[29] Purdy, M., and Daugherty, P., 2016. *Why Artificial Intelligence is the Future Growth?* Accenture. https://www.accenture.com/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth-Country-Spotlights.pdf.

[30] Romero, M., Guédria, W., Panetto,H. and Barafort, B., 2020. Towards a Characterisation of Smart Systems: A Systematic Literature Review. *Computers in Industry*, Volume 120, 103224. https://doi.org/10.1016/j.compind.2020.103224.

[31] Rovins, J.E., Wilson, T., Hayes, J., Jensen, S., Dohaney, J., Mitchell, J., Johnston, D. and Davies, A., 2015. *Risk Assessment Handbook*, Massey University, NZ.

[32] Rowe G., Wright G., 1999. The Delphi Technique as a Forecasting Tool: Issues and Analysis. *International Journal of Forecasting*, Vol. 15, pp.353-375.

[33] Schneiderman, B., 2016. The Dangers of Faulty, Biased, or Malicious Algorithms Requires Independent Oversight. *Proceedings of the National Academy of Sciences*, Vol. 113, No. 48, pp.13538-13540, https://doi.org/10.1073/pnas.1618211113.

[34] Sikorski M., 2020. Forecasting Risks and Challenges of Digital Innovations: Towards a Socially Responsible Design Agenda. In: Lechman E., and Popowska M.(eds), 2020. *Society and Technology. Opportunities and Challenges*. London: Routledge, pp.169-191. https://doi.org/10.4324/9780429278945.

[35] Villani, C., 2018. *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*. https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf [Accessed 17 June 2021].

[36] Voros, J., 2003. A Generic Foresight Process Framework. *Foresight*, 5(3), pp.10-21. https://doi.org/10.1108/14636680310698379.

[37] Walsh, T., 2017. *It's alive! Artificial Intelligence from the Logic Piano to Killer Robots*. La Trobe University Press and Black Inc, Carlton, Australia.

[38] Yoo, Y., Henfridsson, O., Lyytinen, K., 2010. The New Organising Logic of Digital Innovation: An Agenda For Information Systems Research. *Information Systems Research*, 21(4), pp.724-735.