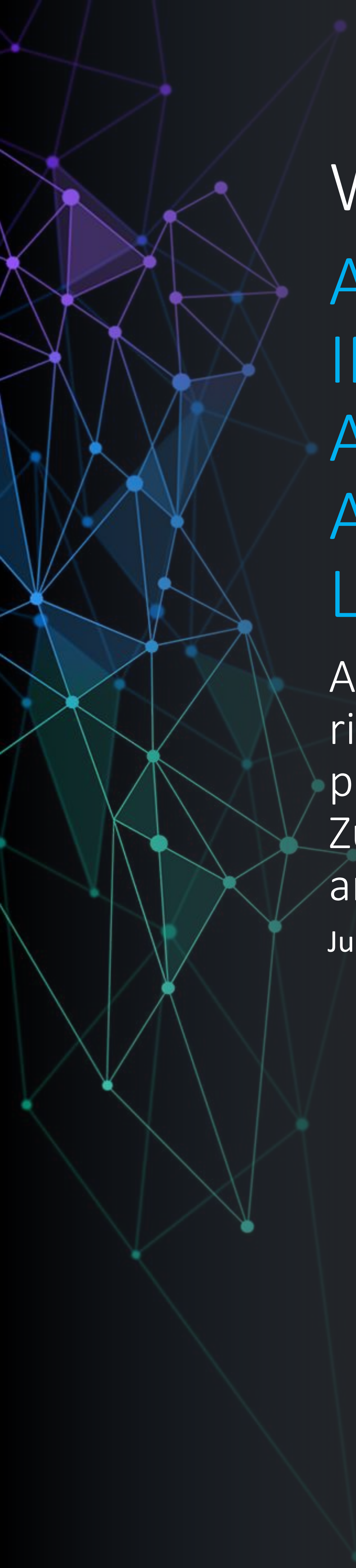


White Paper

ARTIFICIAL INTELLIGENCE AND ALGORITHMIC LIABILITY

A technology and risk engineering perspective from Zurich Insurance Group and Microsoft Corp.

July 2021

TABLE OF CONTENTS

This paper introduces the growing notion of AI algorithmic risk, explores the drivers and implications of algorithmic liability, and provides practical guidance as to the successful mitigation of such risk to enable the ethical and responsible use of AI.

Authors:

Zurich Insurance Group

Elisabeth Bechtold

Rui Manuel Melo Da Silva Ferreira

Microsoft Corp.

Rachel Azafrani

Christian Bucher

Franziska-Juliette Klebôn

Srikanth Chander Madani

Executive summary

A focused, multidimensional, and forward-looking analysis of algorithmic risk to enable the ethical and responsible use of AI

Artificial Intelligence (AI)[1] is changing our world for the better. Unleashing the power of data and AI creates endless business opportunities to ultimately improve the quality of our lives[2].

In 2018, McKinsey suggested that AI could deliver economic activity of $13 trillion by 2030, equivalent to *1.2% additional annual global GDP growth*[3]. In 2020, a paper in Nature asserted that AI could "help address some of the world's most pressing challenges[1] and deliver positive social impact in accordance with the priorities outlined in the United Nations' 17 Sustainable Development Goals[4]".

As business models become increasingly digitized, human lives are impacted significantly by design choices of algorithm creators. AI applications carry a broad spectrum of risks encompassing not only regulatory compliance, but also liability and reputational risk if algorithmic decision-making triggers unintended and potentially harmful consequences, examples of which are given in Chapter 3.

This white paper brings together expertise and insights from both Zurich as a global insurer and risk manager and from Microsoft as a security and technology platform provider to illustrate the growing spectrum of AI algorithmic risk, and to present suggestions for mitigating such rapidly developing risk. Including a foundational overview of the triggers, relevant legal and regulatory aspects, and potential complications of algorithmic liability, this white paper provides practical guidance on the successful mitigation of such risk and presents a focused, multidimensional, and forward-looking analysis of governance principles and assessment tools to enable the ethical and responsible use of AI in line with best practices and emerging regulation.

The analysis includes an in-depth discussion of highly relevant use cases across industries in the areas of product liability, professional indemnity, and medical malpractice, to help companies avoid customer harm, minimize liability exposure and reputational damage resulting from AI solutions, and provide support to leverage advanced technologies for the benefit of customers and society at large.

Defining AI algorithmic risk

While there is a variety of definitions of AI or algorithmic risk, for the purpose of this paper, algorithmic risk is defined as risk arising from the use of data analytics and cognitive technology-based software algorithms in various automated and semi-automated decision-making environments, originating in input data, algorithmic design, and output decisions, and caused by human biases, technical flaws, usage flaws, or security flaws[5].

In consideration of specific challenges of AI such as its high complexity, interconnectivity, opacity, and self-learning, inaccurate, biased, or otherwise flawed model output are among the most prominent failures of AI systems and are discussed in this white paper[6]. In Chapter 3, a closer look is provided on intended and unintended externalities triggering the risk of (potentially unlawful) discriminatory outcomes. It is worth noting here that the terms "AI risk" and "algorithmic risk" are used interchangeably in this paper[7].

Understanding liability risk

If AI-induced risks materialize and cause harm to individuals, or damage to companies or other stakeholders[8], the question of liability arises.

For providers or users of AI solutions, it is key to understand the own-liability exposure as well as potential interconnectivities along the value chain.

Who is ultimately responsible for an AI system's fault and how should fault be identified and apportioned?

> *“Technology can improve the quality of our lives in many ways. Data and AI further accelerate change. Meanwhile, we need to be diligent on managing the algorithmic risk and ethical challenge they bring by ensuring fairness and privacy, with transparency and clear accountability.”*
>
> Ericson Chan
> Group Chief Information and Digital Officer
> Zurich Insurance Group

What sort of remedy should be imposed and what type of losses should be recoverable? Identifying the potential triggers for such algorithmic liability as well as potential complications is essential to mitigate the risks along an AI system's life cycle, as we describe in Chapter 4.

Allocating legal liability for AI systems faces three key challenges, illustrated in Chapter 5.

First, complications of causality may arise due to various contributors that are typically involved in the creation and operation of an AI system, including data providers, developers, programmers, users, and the AI system itself.

Second, nature and cause of damage created by an AI system are important indicators for establishing and allocating liability among these contributors.

Third, legal liability will often be determined on the basis of a patchwork of generic legal concepts expressed in existing legislation but also influenced by best practices and industry standards. However, the increasingly complex use of AI can be expected to test the boundaries of current laws, and regulators have demonstrated interest in expanding legal liability regimes to address AI-specific risks.

In order to mitigate such liability exposure, there are suggested governance principles of responsible AI that organizations would do well to adopt. For developers, there is a multitude of technical control tools across the AI development lifecycle. Both perspectives are covered in Chapter 6.

The insurance industry's effort to understand AI algorithmic risk is in its early stages due to the lack of loss experience data and models that estimate the potential frequency and severity of AI risks. Insurers are starting to assess the impact of AI risks on major lines of business such as product liability, professional indemnity, and medical malpractice, and to design specific risk management services to address potential issues in the AI system development lifecycle. Chapter 7 explores implications for insurers.

Recent advances in enterprise audit software for trusted and responsible AI systems will enable the development of liability risk mitigation strategies, which can be applied to enhance data and algorithm privacy, security, fairness, explainability, transparency, performance robustness, and safety. The technology and insurance industries can combine their strengths to better assess the state of algorithmic risk and find novel solutions.

Introduction

Diving into the world of algorithmic risk and its complexities

This paper illustrates the notion of algorithmic risk along with examples of two kinds of negative externalities, the complication through the increasing use of artificial intelligence, potential triggers of algorithmic liability, and key legal and regulatory considerations to understand the exposure to algorithmic liability. Building on such foundational overview, this paper offers practical guidance, governance principles for the ethical and responsible use of AI, and tools to manage algorithmic risk – for firms in general, and insurance carriers, in particular. At the end, we attempt to anticipate future developments in this fast-moving field.

A. What is algorithmic risk and why is it so complex? 'Because the computer says so'

Last year, a paper[9] co-authored by Microsoft Research applied to AI the phrase Enchanted Determinism: "discourse that presents deep learning techniques as magical, outside the scope of present scientific knowledge, yet also deterministic, in that deep learning systems can nonetheless detect patterns that give unprecedented access to people's identities, emotions and social character." The paper continues, "The discourse of exceptional, enchanted, otherworldly and superhuman intelligence ... has social and political effects, often serving the interests of their powerful creators. Most important among these is that it situates deep learning applications outside of understanding, outside of regulation, outside of responsibility, even as they sit squarely within systems of capital and profit." AI risk can thus be *perceived* as something that cannot be understood, despite being significant.

Furthermore, AI risk is *topical*. In a multinational survey on trust by Edelman[10], 61% agreed that "*Government does not understand emerging technologies enough to regulate them effectively*."

Regulators too are advocating a risk-based approach to managing AI systems. For instance, the European Commission in April 2021 established four categories of AI risk[11]: unacceptable, high, limited, and minimal risks.

Widespread adoption of AI increases the algorithmic risk

Ten years ago, the Financial Times reported a curious phenomenon. The prominence of a certain actress in the news cycle seemed to trigger growth in a stock with the same name[12]. While causality has not been (cannot be?) proven, Anne Hathaway's movie releases have been correlated with gains in Berkshire Hathaway stock. According to the article, "Trading programs are trained to pick up on key words, and the more sophisticated ones can read for sentiment too."

Another writer on the subject opined[13], "As hedge fund managers' computing resources grow ever more powerful ... they are actually able to correlate everything against everything. Oh, it's raining in Kazakhstan? ... Dump Apple stock! Why? Because the computer says that in 193 of the last 240 times it rained in Kazakhstan ... Apple shares went down."

An added complication to algorithmic risk is indicated by the "Because the computer says so" idea of the previous paragraph.

Business models and processes are not just increasingly digital, they increasingly incorporate AI, democratized via the Cloud: The number of enterprises implementing AI grew 270% in the past four years[14].

This AI adoption, including machine learning (ML) models with potential lack of transparency, is accompanied by risks.

> *“The design and development process itself must prioritize privacy, cybersecurity, digital safety and responsible AI, across everything we do. No one will want technology that rapidly scales but breaks the world around us.”*
>
> Satya Nadella
> Chairman and CEO Microsoft Corp.
> Build Developer Conference, May 2021

According to McKinsey[15], these include privacy violations, discrimination, accidents, manipulation of political systems, loss of human life (if an AI medical algorithm goes wrong), compromise of national security (if an adversary feeds disinformation to a military AI system), reputational damage, revenue losses, regulatory backlash, criminal investigation, and diminished public trust.

The body of work around AI risk is evolving. For instance, the Artificial Intelligence/Machine Learning Risk & Security Working Group (AIRS) was founded as recently as two years ago, and its AI risk framework[16] refers to tools from Microsoft referenced later in this paper.

B. Microsoft and Zurich: Leveraging leading cyber security and risk expertise

In 2017, Microsoft announced that it invests over $1 billion annually on security[17]. In September 2020, Microsoft communicated[18] that it analyzed over 470 billion emails and 630 billion authentication events monthly and blocked more than 5 billion threats: “Our unique position helps us generate a high-fidelity picture of the current state of cybersecurity, including indicators to help us predict what attackers will do next. This picture is informed by over 8 trillion security signals per day”. In January 2021, CEO Satya Nadella reported that Microsoft’s security business revenue had surpassed $10 billion annually[19].

Zurich Insurance relies on expertise and insight gathered over nearly 150 years to help its customers manage their increasingly interconnected and complex risks. As co-author of the World Economic Forum’s Global Risks Report, Zurich is recognized as an insurer that understands the needs of its customers, which include large companies, small enterprises, and individuals in more than 125 countries and territories.

As the insurance market has increasingly demanded digital services, Zurich has developed solutions and risk management services that make businesses more efficient through the use of artificial intelligence, data analytics, and other related technologies. Its customer service approach earned the company the Global Innovator of the Year award[20].

In September 2020, Zurich’s CEO Mario Greco confirmed that “Zurich continues to launch innovative offerings to meet the demands of our customers and partners for a fully digital and streamlined experience[21]”.

Algorithmic risk: Intended or not, AI can foster discrimination

As indicated in the previous chapter, AI can cause negative externalities, defined as costs imposed upon third parties without their consent as a result of a transaction between a provider and a consumer. These negative externalities are of two types: intended and unintended.

A. Creating bias through intended negative externalities

As an example of an intended negative externality, consider the legal action[22] in 2019 against a leading social media company by the U.S. Department of Housing and Urban Development. That social media company is accused of "unlawfully discriminating against people based on race, religion, familial status, disability and other characteristics that closely align with the 1968 Fair Housing Act's protected classes" through the tools it provides to its advertisers, according to National Public Radio.

The charging document[23] cites specific examples, noting that the "Respondent has offered advertisers hundreds of attributes from which to choose, for example to exclude 'women in the workforce,' 'moms of grade school kids… or people interested in 'parenting,' 'accessibility,' 'service animal,' 'Hijab Fashion,' or 'Hispanic Culture.' Respondent also has offered advertisers the ability to limit the audience of an ad by selecting to include only those classified as, for example, 'Christian' or 'Childfree.'" That same document calls out the social media company's alleged use of "machine learning and other prediction techniques to classify and group users so as to project each user's likely response to a given ad." According to that document, certain individuals (prospective tenants) were discriminated against (in that information on available dwellings was withheld from them) through the use of ostensibly benign tools.

B. Bias as a result of unintended negative externalities

As an example of an unintended negative externality, consider an assessment carried out in the U.S. on the risk of convicted criminals carrying out future crimes. i.e. recidivism probability. According to a 2016 study[24], the proprietary assessment algorithm in question appeared to have significant ethnic bias: "Black defendants were… 77% more likely to be pegged as at higher risk of committing a future violent crime and 45% more likely to commit a future crime of any kind."

While this study has been quoted in the Harvard Business Review[25], it has also been challenged – and the challenge in the Harvard Data Science Review[26] opines that the "focus on the question of fairness is misplaced, as these (secret) algorithms (that make important decisions about individuals) fail to meet a more important and yet readily obtainable goal: transparency." In this example, certain individuals ("Black defendants") were set at a disadvantage unintentionally.

Another example of an unintended negative externality is from the U.S. healthcare system. Researchers from the University of California at Berkeley, the Booth School of Business, and Partners HealthCare found evidence of racial bias in a popular algorithm widely used to guide health decisions. Their paper[27] stated, "At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients."

Yet another case is of an employer's recruitment-support algorithm reportedly discriminating on the basis of gender. According to a 2018 report by Reuters[28], "(that employer's) computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry. In effect, (that employer's) system taught itself that male candidates were preferable. It penalized resumes that included the word women's, as in women's chess club captain."

Data and design flaws as key triggers of algorithmic liability

If an AI deployed by a company is found to be malfunctioning and causing damage to an employee, a customer, or any third party, such company may be held liable. To mitigate such liability risk, it is essential to identify the errors or shortcomings that could trigger liability along the life cycle of an algorithmic application, and to account for potential complications[29].

The relevant phases of an algorithm life cycle are sketched out below, combined with a summary of key risks that need to be mitigated from a model provider or user perspective, respectively.

A. Model input phase

The quality of the data fed into an algorithmic model is key to its successful operation according to plan. To ensure reliable and high-quality data, the following potential flaws need to be prevented by the model provider or user, respectively, during the model input phase[30]:

- Inaccurate or otherwise flawed model input due to poor data quality
- Unforeseen inherent bias of data as data is reflective of the biases in society

The prevention of unforeseen inherent bias is one of the biggest challenges in the context of data quality.

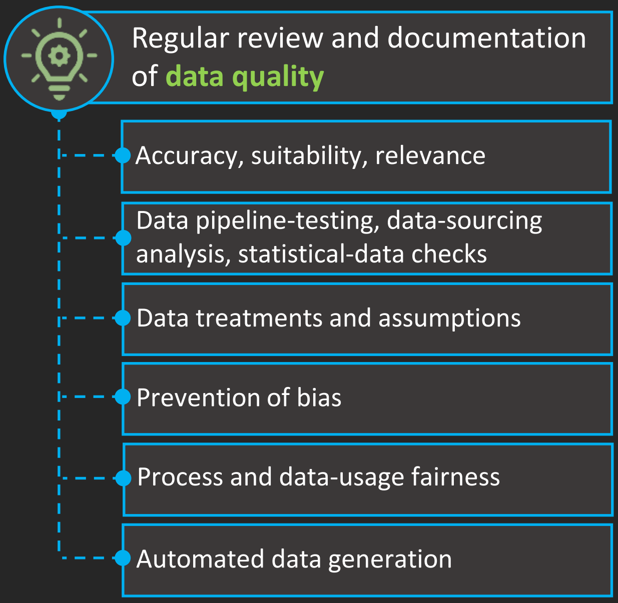
Even in instances where the data appears perfect to the human eye, AI can pick up patterns in the data that were not anticipated during the training process. This can cause an AI system to draw inaccurate conclusions and thus, ultimately, generate incorrect or undesired outcomes.

If advanced AI systems collect data from multiple sources and use non-interpretative models with little to no human interaction, then they might pick up on micro-signals that may not be detectable by humans. As a result, ***particularly complex and interconnected AI solutions bear the risk of even further amplifying any existing biases*** *(and potentially making such biases systemic)*.

The following practical guidance can serve as a reference point to mitigate algorithmic model risk.

PRACTICAL GUIDANCE

Model input phase



Regular review and documentation of **data quality**

- Accuracy, suitability, relevance
- Data pipeline-testing, data-sourcing analysis, statistical-data checks
- Data treatments and assumptions
- Prevention of bias
- Process and data-usage fairness
- Automated data generation

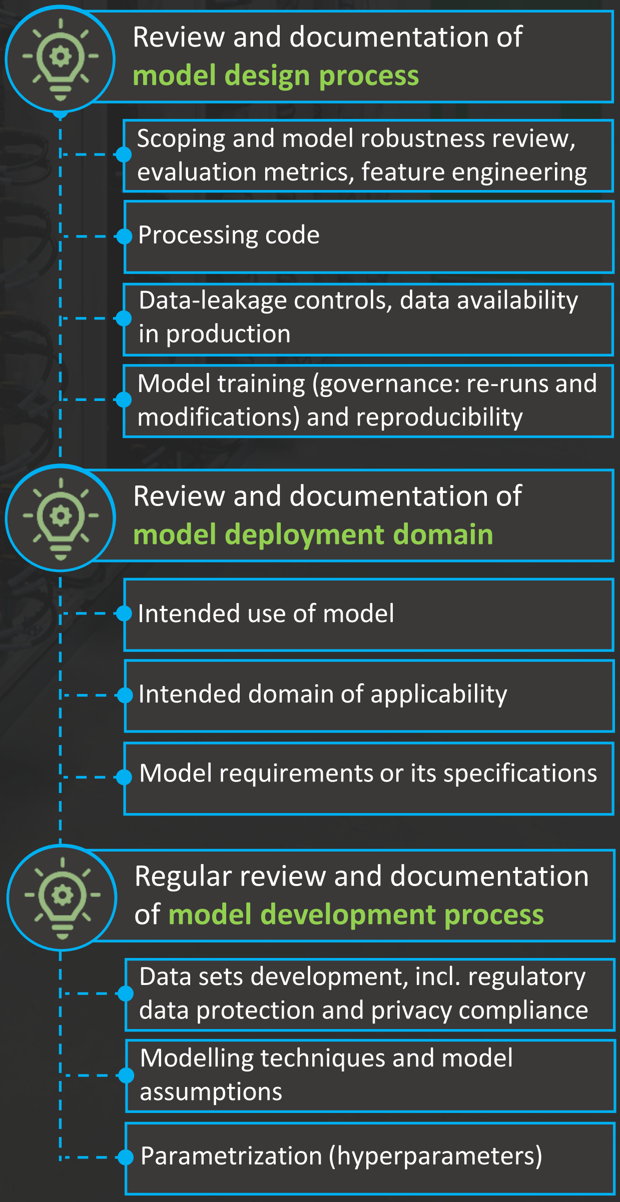
B. Model design and development phase

If a model is designed or further developed to solve a problem specified during ideation, the following potential flaws need to be prevented by the model provider[31]:

- Deficiencies in AI model design (e.g., scoping, model robustness, business-context metrics)
- Appropriateness of deployment domain
- Deficiencies in model development process
- Inaccurate or otherwise flawed model training data

PRACTICAL GUIDANCE

Model design and development phase



Review and documentation of **model design process**

- Scoping and model robustness review, evaluation metrics, feature engineering
- Processing code
- Data-leakage controls, data availability in production
- Model training (governance: re-runs and modifications) and reproducibility

Review and documentation of **model deployment domain**

- Intended use of model
- Intended domain of applicability
- Model requirements or its specifications

Regular review and documentation of **model development process**

- Data sets development, incl. regulatory data protection and privacy compliance
- Modelling techniques and model assumptions
- Parametrization (hyperparameters)

C. Model operation and output phase

AI models may not perform or deliver outcomes according to plan.

It is therefore crucial to understand and manage the potential key errors or shortcomings during the model operation and output phase:

- Malfunctioning of model
- Deficiencies in model interpretability (transparency / explainability)
- Insufficient elimination of risk of inappropriately biased / unfair outcomes
- Inaccurate or flawed model outcomes

This applies to both the model provider and the user perspective.

Once an AI model is in production, it needs to be regularly verified that the algorithm continues to work as intended and its use continues to be appropriate in the current business environment.

PRACTICAL GUIDANCE

Model operation and output phase

Regular review and documentation of model operation

- Performance testing, feature-set review, rule-based threshold setting
- Overall model robustness (overfitting)
- Accuracy and precision of model operation according to plan
- Relevant business requirements and potential restrictions
- Model interpretability (transparency/ explainability

Review of model output monitoring

- Continuous and effective monitoring of plan coverage by model operation
- Metrics and acceptance criteria
- Accuracy, appropriateness and correctness of model output in line with external commitments (fair and unbiased outcomes) and ethical values
- Dynamic model calibration
- Model governance, reporting and escalation process

> *"Despite being significant, AI risk is sometimes perceived as something that cannot be understood."*
>
> Srikanth Madani
> Industry Advocate
> Worldwide Financial Services | Microsoft

D. Potential complications can cloud liability, make remedies difficult

If flaws and failures along an algorithm's life cycle cause harm or damages, questions of damage attribution, apportionment, and possible remedies will arise. Who is ultimately responsible for such failure and how should fault be identified and apportioned? What sort of remedy should be imposed and what type of losses should be recoverable?

Such questions, which are fundamental for the assessment of algorithmic liability, can be further complicated by a variety of circumstances from the perspective of both providers and users of algorithmic models.

First of all, regardless of the programming language, all algorithms require specialized knowledge to be understood and there are no requirements for judges to understand them. In addition, the complex and often opaque nature of algorithms, specifically "black box" algorithms or deep learning applications, means that they lack transparency and can sometimes hardly be understood by experts (inherent challenge of explainable AI). Potential modifications through updates or self-learning during operation and limited predictability are additional factors adding to the complexity and opacity of AI systems. Also, hidden errors are likely to go undetected for a long time (often until it is too late) which again complicates the traceability of relevant failures.

Second, complications may arise due to intricate origins, as algorithms are frequently made up of different - not necessarily coordinated - contributions. For example, algorithms might have potential interdependencies with other sophisticated systems in such a way that the reliability of these algorithms might depend upon conditions in those other systems, making it potentially difficult to identify who is responsible for a specific result.

Similarly, the integration of algorithms into products and services (if algorithms are only components of the whole) complicates the search for the actual error and the respective responsibility.

This is of particular relevance in case of mass consumer products and services where algorithms may pass through the hands of a variety of people other than their developers, such as designers, manufacturers of physical products, providers of services, distributors, licensees, etc.

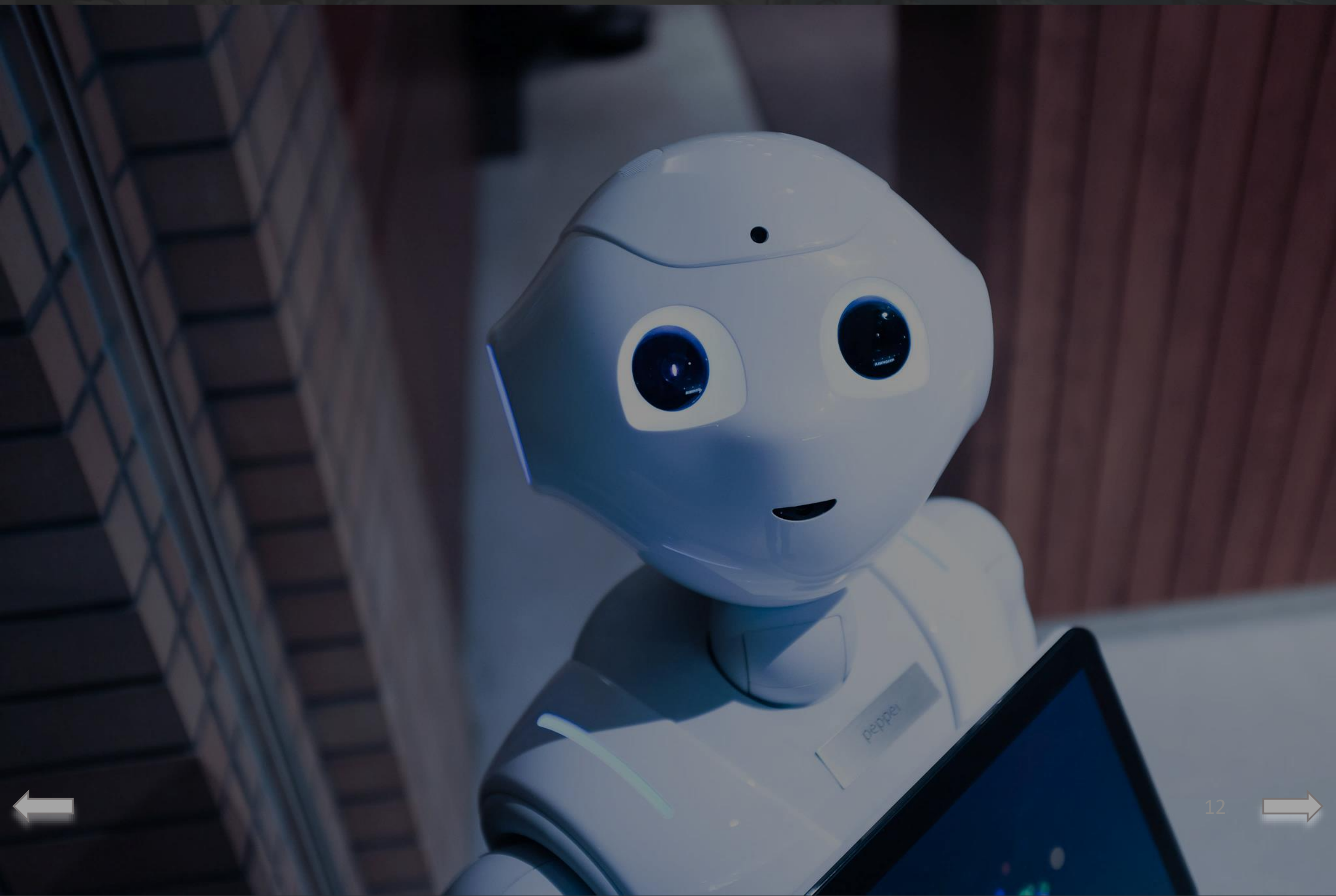
If there are multiple contributors to multi-layer algorithms, identifying who is responsible and needs to be held legally accountable from a liability perspective often presents a major challenge. In this context, the use of external code repositories and data such as GitHub, GitLab, Amazon ML, etc. play an important role (see Chapter 6.C).

Third, as illustrated in Chapter 3, AI failures can be much more impactful than traditional model failures due to the increasing use of AI in sensitive areas such as autonomous driving or medical advice (with AI taking or contributing to decisions over life or death), and its greater scale and reach than comparable decisions or actions taken by humans (AI-based manipulation of elections, pricing on e-commerce platforms, etc.).

Such high impact is particularly relevant for the potential spread or institutionalization of bias and discriminatory outcomes.

From a liability perspective, such large-scale impact can significantly expand the range of potential claimants, the volume and severity of damages.

In addition, as algorithms flow across borders and potential damage can be caused across various jurisdictions, AI failures can give rise to transboundary claims, forum shopping practices, and, possibly, regulatory competition.



How to determine algorithmic liability?

Allocating legal liability for AI systems faces three key challenges[32]. First, various contributors are typically involved in the creation and operation of an AI system, including data providers, developers, programmers, users, and the AI system itself which leads to complications of causality. Second, the nature and cause of damage created by an AI system can be decisive for establishing and allocating liability among these contributors, as the following table[33] from CMS indicates:

Nature of damage	Question of liability
Was damage caused when in use and were the instructions followed? Was the AI system provided with any general or specific limitations and were they communicated to the purchaser?	User vs. Owner?
Was the damage caused while the AI system was still learning?	AI Developer vs. Data Provider?
Was the AI system provided with open-source software?	Programmer vs. Inventor?
Can the damage be traced back to the design or production of the AI system, or was there an error in the implementation by its user?	AI Developer vs. User?

Third, building on the analysis in Chapter 4 of potential liability triggers along an AI's life cycle, a closer look at existing liability regimes reveals that very generic legal concepts can be applied to algorithmic liability in most countries. In addition, there is a slowly emerging set of specific rules which some countries have adopted or are planning to implement over the next years.

Putting the spotlight on the U.S., European Union and China, this chapter will provide an overview of both existing liability regimes as applicable to algorithmic liability, as well as of emerging regulatory regimes specific to AI systems[34].

The analysis will also point to current limitations of the existing legal frameworks and indicate the potential for new areas of liability to emerge.

A. General legal approaches: Caution in a fast-changing field

In general, the civil law provides ample opportunities for parties damaged by AI-powered decisions to seek compensation, damages, or other redress. For example, products with inbuilt AI may be covered by existing machine safety laws, and general norms of product liability law may be applied to determine liability, causation, and damages[35]. Furthermore, general principles from contract and tort law may provide the basis for holding contributors to AI-systems liable. At the same time, firms may equip their AI-powered products and services with contractual limitations on liability, terms of use, warnings and notices, exclusions, and indemnities in a similar way as if their product or service relied on human intelligence.

Against the fast rise of algorithm-related activities, the responses of the U.S., Chinese and European legal systems have so far been rather cautious.

In a field so diverse, turbulent, and fast-changing, where much of the benefits and risks are unknown, legislatures in all three regions have not yet intervened with additional general measures[36]. Besides ad hoc interventions for some high-profile products and services incorporating algorithms (primarily financial robo-advisors, algorithms designed for risk segregation in insurance, and driverless cars), algorithm-related activities are captured by existing laws and regulations only in case of conflicts with general legal concepts that are not specifically designed to address algorithmic risk and liability[37].

Relevant regimes are, in particular, product liability, machine safety, tort law in general as well as anti-discrimination, data protection and privacy, and antitrust laws.

> *"The increasingly complex use of AI can be expected to test the boundaries of existing legal approaches to liability. A proactive governance approach to ensure trustworthy AI will mitigate algorithmic risk and inspire confidence in a digital society."*
>
> Elisabeth Bechtold
> Global Lead Data Governance & Oversight
> Zurich Insurance Group

Similarly, judicial cases on damages or personal injuries arising out of algorithmic products have been largely focusing so far on the damages caused by robots, criminal predictive systems and smart devices, and on overall privacy protection from algorithmic invasion[38]. Several recent disputes illustrate how principles of tort liability and employer liability for workplace injuries can be applied to cases involving AI[39].

Overall, decisions on algorithmic liability have so far been adopted mostly by courts or specialized agencies competent to track down violations of rules in the aforementioned fields as well as of finance, insurance and consumer protection laws[40].

As far as the application of existing doctrines from contract, tort, product liability, and other areas of the law are concerned, the increasingly complex use of AI can be expected to test the boundaries of such laws with respect to appropriate mechanisms for identification of fault and causation, damage attribution and apportionment, type and amount of potentially recoverable losses and appropriate remedies.

B. Challenges and limitations of existing legal approaches

The areas of **tort** and **product liability law** represent important mechanisms to **mitigate algorithm-induced harms** and are also highly illustrative of the challenges and limitations existing legal approaches are facing. **Complications may arise** due to AI-specific features such as an **AI's autonomy**, its frequent appearance **as a "service"** (thus not subject to product liability laws), potentially **multi-layered** third-party **involvement** and the **interface between human and AI**.

Challenge 1: Finding fault when algorithms are wrong

Its ability to learn is a key characteristic of an AI system. Going beyond the implementation of human-designed algorithms, AI systems create their own algorithmic models and data structures, sometimes by revising algorithms originally designed by humans, and sometimes completely from scratch. This raises complex issues in relation to product liability, which is centered on the issue of attributing responsibility for products that cause harm[41].

So, if an algorithm designed largely or completely by computers fails, who should be held responsible?

Challenge 2: Assigning responsibility within the supply chain is difficult

Along the life cycle of an AI system, there will typically be a multitude of suppliers upstream of a consumer. At the very beginning, there is the company that sold the product directly to the consumer. Software components of such products may have been purchased from another company. While some portions of this software may have been built in-house and licensed, other portions may have been acquired from yet another company.

Apportioning responsibility within the supply chain will involve not only technical analysis regarding the sources of various aspects of the AI algorithm, but also the legal agreements among the companies involved, including any associated indemnification agreements[42].

From a liability perspective, where AI systems communicate and engage with one another, the potential culpability for anything that could go wrong shifts from one component to another as easily and quickly as the data transferred between them[43].

If an AI system fails, allocating responsibility for such failure requires a multi-layered assessment if several parties are involved.

There are AI developers, algorithm trainers, data collectors, controllers, and processors, manufacturers of the devices or designers of the services incorporating the AI software; owners of the software (which are not necessarily the developers); and the final users of the devices or services (and perhaps other third parties involved)[44].

Challenge 3: Failing to consider human AI interface may lead to harm

Understanding the human-AI interface is a key challenge, crucial to mitigating product liability risk. Developing a successful AI solution requires not only a proper understanding of the interactions among the system's different software components, but also an understanding of how humans will interact with those components.

Companies that fail to anticipate the assumptions and decisions that will shape human/AI interactions risk releasing products that might behave in unintended and potentially harmful ways[45].

C. AI-specific best practice standards and emerging regulation

To date, a broad range of standards and regulatory best practices for the responsible and ethical use of artificial intelligence developed across the world[46] provide industry guidance on how algorithms should be written and deployed. While such non-binding standards provide valuable and often highly detailed guidance[47], in practice, it is still unclear to which extent such best practice recommendations will impact the legal consequences for the possible failure of ex-ante incentives (e.g., whether it might serve as a "safe harbor" in certain instances), until AI-specific legislation, regulation, or jurisprudence will be issued[48].

Conscious of the limitations and challenges of existing legal regimes, regulators and legislators across the globe are increasingly adopting regulatory frameworks that specifically target issues related to AI liability, complementary to the general legal principles sketched out at the beginning of this chapter. In the following, selected highlights of the most recent liability-relevant regulatory action in the United States, Europe and China are provided.

> *Algorithmic risk is not solely a technical issue and requires holistic governance practices to address."*
>
> Rachel Azafrani
> Security Strategist | Microsoft Corp.

United States

Until recently, U.S. regulatory efforts on AI systems focused on regulating autonomous driving. However, broader legislative action can now be observed in the areas of AI, automated systems, in particular concerning algorithmic accountability, automated decision-making, facial recognition technology and transparency[49]. Several U.S. states have also become active in this respect.

The California Consumer Privacy Act of 2018, as amended by the California Privacy Rights Act of 2020 (CPRA), empowers the future California Privacy Protection Agency to issue regulations whether consumers may opt out of automated decision-making systems and how consumers may receive meaningful information about the logic involved in such systems[50].

The Virginia Consumer Data Protection Act of 2021 has created a related opt-out right concerning automated data profiling systems[51]. Also in this context, the rise of the surveillance industry is a growing concern for privacy advocates due to the widespread collection of individuals' images and biometric information without notice or consent for the extensive use of facial recognition.

In addition to such new AI-specific legislation, the *Federal Trade Commission (FTC)* has specified established existing laws such as the *Fair Credit Reporting Act (FCRA)* and the *Equal Credit Opportunity Act (ECOA)* in light of technological advances, now prohibiting unfair and deceptive practices in the context of AI-powered automated decision making[52].

European Union

Building on its influential General Data Protection Regulation (GDPR), the European Union has taken a progressive stance on trustworthy AI and its implications for algorithmic liability by launching specific (non-binding) resolutions such as *the Civil Law Rules on Robotics (2017)*[53], the *Civil Liability Regime for AI (2020)*[54], and on April 21, 2021, a proposal of a binding comprehensive legal framework for AI (Artificial Intelligence Act)[55]. The proposed *Artificial Intelligence Act* focuses on the regulation of high-risk AI applications[56], while certain practices should be prohibited for all AI systems as a violation of fundamental human rights[57].

Connecting the key pillars of the European Commission's newly proposed *Artificial Intelligence Act* with the European Parliament's resolution on a *Civil Liability Regime for AI*, it can be expected that a risk-based approach will be introduced, with a common strict liability regime for high-risk AI systems[58].

For harms or damages caused by AI systems not classified as high-risk, as a general rule, fault-based liability rules will likely apply.

However, an affected person might generally benefit from a presumption of fault on the part of the operator, who should be able to exculpate itself by proving it has abided by its duty of care[59]. Importantly, *AI liability insurance is considered essential for ensuring the public's trust in AI technology and might become mandatory for operators of high-risk AI systems*. These regulatory efforts complement the AI-relevant elements of the GDPR prohibits algorithmic automated decision-making if such a decision "produces legal effects concerning [the data subject] or similarly significantly affects [the data subject]"[60].

Where automated decision-making, including profiling, or automated processing occurs, the data subject needs to be informed and equipped with the right to obtain human intervention, an explanation of how the decision had been reached, and the right to challenge such decision[61]. Examples where this central GDPR provision applies include decisions about recruitment, work assessments, credit applications, insurance policies, commercial nudging, behavioral advertisements, and neuro-marketing.

China

Being at the forefront of algorithm development with highly powerful AI systems, China is planning a comprehensive legal, ethical and policy regulatory framework for AI by 2030 under the State Council's *New Generation AI Development Plan*[62]. Until such AI-specific legislation is issued, existing legal regimes, such as tort law and product quality law apply, to AI liability claims[63].

To date, liability questions are particularly relevant in AI-powered areas such as medical malpractice and autonomous driving[64]. In the field of medical malpractice, the burden of proof for the malfunctioning of medical software initially rests with the injured party, but if the medical institution or software provider claim that they shall not be held liable, each of them shall bear the burden of proof for their own defense that the medical software used is free from any flaws[65]. In the field of autonomous driving, the Ministry of Public Security has recently issued specific draft legislation to revise the *Road Traffic Safety Act* that addresses the complex liability challenges in accidents caused by self-driving cars.

For highly autonomous self-driving cars, however, it is expected that the legislation will provide further clarity on liability allocation as the technology develops[66].

In response to growing concerns for privacy, China has been strengthening its data and privacy protection measures in recent years. At the heart of this effort is the *Personal Information Security Specification*, a privacy standard initially released in 2018 and revised in 2020 to elaborate on the broader privacy rules of the *2017 Cyber Security Law*[67]. The focus of the new standard is on protecting personal data and ensuring that people are empowered to control their own information[68].

The voluntary standard is complemented by a certification standard for privacy measures and is considered to be even more onerous than the GDPR in certain instances[69].

Another important recent development is the issuance of the e-commerce law in 2018 which defines a duty of care for e-commerce platform providers for both intellectual property and personal data protection, and thus challenges the traditional argument that platform operations cannot be liable for algorithmic, automatically generated results[70].

D. New approaches to tackle algorithmic liability risk?

The AI-specific challenges and limitations outlined above lead to questions around whether AI merits a new approach to liability[71]. To date, courts have applied traditional legal approaches to complex and sometimes hardly explainable systems across a wide range of fields such as product liability, intellectual property, fraud, criminal, antitrust, anti-discrimination, data protection and privacy laws.

To strike the balance between encouraging innovation, incentivizing investment and effectively protecting individual rights and public safety, policymakers will need to consider whether (and to what extent) existing regulatory structures and tools need to be modified.

With a shifting focus of the public policy debate now increasingly looking at negative externalities of algorithms, far reaching reforms are being suggested, such as endowing AI with legal personhood associated with mandatory insurance, thereby making algorithms capable of both owning assets and being sued in court[72]. Another proposal suggests some form of systemic oversight and ex-ante expert guidance on the development and use of algorithms. In line with the newly proposed Artificial Intelligence Act[73], this could eventually be paired with a certification system run by a federal agency that would penalize algorithms not following the approved standards[74].

As many elements of algorithmic liability are yet unresolved, new AI-specific liability regulation may be emerging. Companies need to be aware of this legal uncertainty and integrate this regulatory and/or liability risk into overall risk considerations and their risk profile.

Principles and tools

Managing algorithmic risk requires a holistic approach

Based on market research[75] on AI adoption conducted with the Altimeter Group and the University of Saint Gallen, Microsoft[76] concluded, "Business leaders are often stalled by questions about how and where to begin implementing AI across their companies; the cultural changes that AI requires companies to make; and *how to build and use AI in ways that are responsible, protect privacy and security, and comply with government rules and regulations*" (emphasis added).

It was to help answer such questions that the company launched the AI Business School in 2019. The AI Business School[77] covers AI strategy, AI culture, responsible use of AI, the scaling of AI, AI for business users, and AI for leaders.

We look here at some of the technical and governance considerations for organizations as they mitigate algorithmic risk.

A. Tools & methodologies for responsible AI

Microsoft's perspective[78] on responsible AI in financial services provides a comprehensive list of tools and methodologies for mitigating AI risks. The following pillars are adapted from that publication.

1. Dataset and model inventory

DevOps in Azure Machine Learning (MLOps)[79] makes it easier to track and reproduce models and their version histories.

MLOps offers centralized management throughout the entire model development process (data preparation, experimentation, model training, model management, deployment, and monitoring) while providing mechanisms for automating, sharing, and reproducing models.

2. Transparency

Transparency is a key part of the so-called impact assessment which can be found on the following page. The tools below can be used for interpreting AI models:

- **InterpretML**[80] is an open-source package for training interpretable models and explaining black box systems. It includes several methods for generating explanations of the behavior of models or their individual predictions (including Explainable Boosting Machine (EBM) , enabling developers to compare and contrast explanations and select methods best suited to their needs.

- **Model Interpretability**[81] is a feature in Azure Machine Learning that enables model designers and evaluators to explain why a model makes the predictions it does. These insights can be used to debug the model, validate that its behavior matches objectives, check for bias, and build trust.

- **Datasheets for datasets**[82] is a paper proposing that dataset creators should include a datasheet for their dataset, such as training datasets, model inputs and outputs, and model features. Like a datasheet for electronic components, a datasheet for datasets would help developers understand whether a specific dataset is appropriate for their use case.

- **Local Interpretable Model-agnostics Explanations (LIME)**[83] provides an easily understood description of a machine learning classifier by perturbing the input and seeing how the predictions change.

3. Impact

Performing an impact assessment and getting it approved by the accountable executive is a key control. The tools below can be used for interpreting AI models for the impact assessment:

- Methodology for reducing bias in word embedding helps reduce gender biases by modifying embeddings to reduce gender stereotypes, such as the association between *receptionist* and *female*, while maintaining potentially useful associations, such as the association between the words *queen* and *female.*
- A reductions approach to fair classification[84] provides a method for turning any common classifier AI model into a "fair" classifier model according to any of a wide range of fairness definitions. For example, consider a machine learning system tasked with choosing applicants to interview for a job. This method can turn an AI model that predicts who should be interviewed based on previous hiring decisions into a model that predicts who should be interviewed while also respecting demographic parity (or another fairness definition).
- Fairlearn[85] is an open-source toolkit that empowers data scientists and developers to assess and improve the fairness of their AI systems.

4. Information protection

Protecting sensitive data elements is a key control and we have a range of tools to help with this:

- Microsoft provides guidance[86] on how to protect algorithms, data, and services from new AI-specific security threats. While security is a constantly changing field, this paper outlines emerging engineering challenges and shares initial thoughts on potential remediation.
- Homomorphic encryption is a special type of encryption technique that allows users to compute on encrypted data without decrypting it. The results of the computations are encrypted and can be revealed only by the owner of the decryption key.

  To further the use of this important encryption technique, Microsoft developed the Simple Encrypted Arithmetic Library (SEAL)[87] and made it open source.
- Multi-party computation (MPC)[88] allows a set of parties to share encrypted data and algorithms with each other while preserving input privacy and ensuring that no party sees information about other members. For example, with MPC one can build a system that analyzes data from multiple hospitals without any of them gaining access to each other's health data.
- Data scientists, analysts, scientific researchers, and policy makers often need to analyze data that contains sensitive personal information that must remain private. Commonly used privacy techniques are limiting and can result in leaks in sensitive information. Differential Privacy (DP)[89] is a technique that offers strong privacy assurances, preventing data leaks and re-identification of individuals in a dataset. Microsoft is a key contributor to SmartNoise[90], a toolkit that uses state-of-the-art DP techniques to inject noise into data, to prevent disclosure of sensitive information and manage exposure risk.
- Artificial Intelligence and the GDPR Challenge[91], a whitepaper authored by representatives from Microsoft's Corporate, External, & Legal Affairs (CELA), addresses issues of AI explainability and provides considerations surrounding GDPR requirements for AI fairness in credit scoring and insurance underwriting.

5. Model monitoring

AI model monitoring is an ongoing control to check for model performance degradation. We have capabilities for this in Azure Machine Learning. DevOps in Azure Machine Learning (MLOps), mentioned above, helps teams monitor model performance by collecting application and model telemetry. These features can help for instance banks to audit changes to their AI models, automate testing, and reproduce model outputs.

> *"Microsoft's perspective on responsible AI in financial services provides a comprehensive list of tools and methodologies for mitigating AI risks."*
>
> Christian Bucher
> Global Data & AI Principal | Microsoft

B. Governance and principles for responsible AI and data use

Algorithmic risk is not solely a technical issue and requires holistic governance practices to address. AI governance can be defined as "the structure of rules, practices, and processes used to ensure that the organization's AI technology sustains and extends the organization's strategies and objectives[92]."

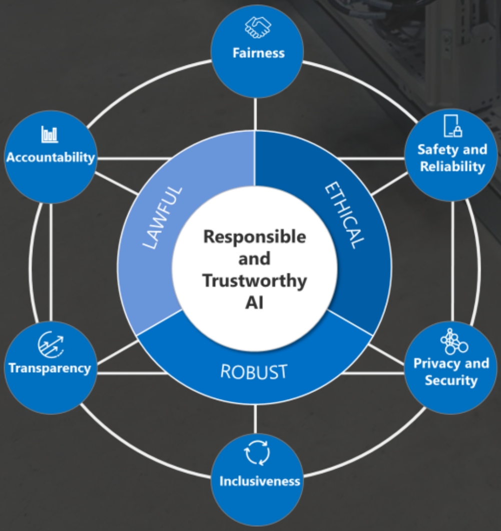
Both insurers and businesses generally must consider how AI systems transform organizations, so as to gain a better understanding of risk posture and, importantly, to implement effective operational and technical governance mechanisms.

The unique characteristics of AI systems and the challenges posed by AI decision-making require new governance mechanisms and, in some cases, new compliance practices, to supplement traditional risk management and compliance.

Best practices for operational and technical AI governance continue to mature in industry and standards bodies, such as the International Standards Organization/International Electrotechnical Commission (ISO/IEC) committee on AI[93]. Corporate adoption of these practices will help propagate better metrics and mitigations for algorithmic risk in every industry.

Due to the horizontal nature of algorithmic risk across business functions, organizational leadership should be responsible for setting AI governance strategy. Organizations can assess their relationship to AI products and services at a strategic and operational level to allow for the identification of potential sources of risk and designate responsible parties to create and implement necessary governance mechanisms. Taking a principled approach can promote an organizational culture conducive to minimizing and confronting algorithmic risk.

Microsoft has adopted six responsible AI principles[94] to guide technology development and operations: Fairness, Reliability and Safety, Privacy and Security, Inclusiveness, Transparency, and Accountability. The *European Commission's high-level expert group on AI*[95] suggested that trustworthy AI has three components: "lawful (complying with all applicable laws and regulations)," "ethical (ensuring adherence to ethical principles and values), and "robust (both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm)." We harmonized both principles in the diagram *pillars of responsible and trustworthy AI*.



Pillars of responsible and trustworthy AI

In the same spirit, Zurich has launched an industry-leading commitment on the responsible use of data (the "Zurich Data Commitment") which also encompasses the ethical use of advanced technologies such as AI[96]. The Data Commitment reflects the objectives of Zurich's sustainability strategy, including the goal to inspire trust in a digital society.

Such trust is founded on core values such as transparency, privacy and security, fairness and customer benefit as expressed in the Data Commitment.

C. Mitigating risks of external code repositories and data

Source code and data repositories and libraries are rich resources and integral tools for AI developers, underpinning millions of software projects globally. Developers can retrieve AI system components from repositories like pretrained models and datasets that can be refreshed and recycled for countless applications. Many repositories are well-recognized and used by developers in small and large companies alike, such as GitHub, GitLab, Bitbucket, Sourceforge, Kaggle, and dozens of others.

AI repositories can be open source or available through paid access, and each type can carry risk. Due to their open nature, publicly contributed AI resources have security vulnerabilities and carry some risk that they may be corrupted intentionally or have unintentional quality issues. Malicious actors could contribute corrupted source code and training data or deliberately target a product in development that is using a certain repository[97].

Even without malicious interference, the quality of information retrieved from public and paid repositories can vary and generate other risks, such as unwanted bias and bugs. Furthermore, developers frequently combine elements from different sources for AI products and services, making it difficult to track vulnerabilities that stem from different origins.

Major AI repositories recognize the challenge of intentional and unintentional risk of public and paid contributions and have implemented security checks and tools for developers to use.

One example is CodeQL on GitHub, a semantic code analysis engine to help detect vulnerabilities in code[98]. Outside of repositories, platforms like Semmle and DeepCode also offer code analysis tools for developers.

It is important that developers making use of AI repositories recognize that every resource will carry different risks and that they leverage the tools available as a regular part of technical AI governance practices.

While it is not possible to eliminate all risks of using AI repositories, organizations must look to implement both technical and organizational governance mechanisms to mitigate these risks and establish courses of action to address them.

> *Microsoft has adopted six responsible AI principles to guide technology developments and operations: Fairness, Reliability and Safety, Privacy and Security, Inclusiveness, Transparency, and Accountability."*
>
> Franziska-Juliette Klebôn
> Data & AI Lead | Microsoft

Managing algorithmic liability risk

Insurance for an emerging exposure

The use of AI systems that fail to function and perform within expected and acceptable business outcomes can lead to claims for property damage, business interruption, personal injury, professional liability, medical malpractice, and cyber risks[99]. However, the findings of a recent McKinsey survey suggest that a *minority of companies recognize many of the risks of AI use*, and only a few are working to reduce the risks[100].

The case for companies to consider insurance of AI risks is indisputable[101]. Yet, the insurance industry is in the early stages of understanding AI algorithmic risks and developing coverages and service propositions that effectively manage those risks, due to the lack of loss experience data as well as models that estimate the potential frequency and severity of AI risks.

Moreover, because of the interconnectedness of businesses, losses associated with AI risks may spread fast across the world, increasing substantially the accumulation of risk and raising insurability issues due to the lack of risk diversification.

As a result, there are few AI insurance policies commercially available and AI losses are not explicitly covered by traditional insurance products. In some cases, the business interruption risks potentially related to unstable AI systems may be covered by existing business interruption policies. Errors and omissions insurance in the field of robotics offers professional liability coverage that can be considered for AI systems and would typically be complemented with specialized risk management services[102].

There is a recent example of an insurance product for companies developing AI systems that guarantees compensation against under-performance[103]. As with Parametric Insurance, the insurer analyzes the AI system and insures its performance against a set of KPIs, so that the users of the AI system can trust it and deploy it at scale with higher confidence across the business. In the event of potential system under-performance (e.g., specific output parameters are out of range of valid and acceptable outcomes), the user of the AI system is indemnified.

AI algorithmic risks vary across insurance lines of business and there are specific risk management services that insurers can potentially offer to large corporates and small to medium-sized enterprises.

***For a given business context, the focus is on potential issues in the AI algorithm development lifecycle from input data to algorithm design, to integration in business processes and use in production to assist with decision making*.**

As indicated in Chapter 4, problems can arise due to:

- **Input data:** Biased training data sets; incomplete, outdated, or irrelevant data sets for the intended purpose; insufficiently representative data sets; inappropriate data collection techniques and use of unreliable data sources; data feature extraction and engineering; mismatch between training data and actual input.
- **Algorithm design:** Lack of business domain specification, misunderstanding of the full ecosystem in which the AI is evolving, undetected out of domain use of the algorithms, algorithms incorporating with biased logic, flawed assumptions or underlying hypotheses, expert judgments in conflict with ethical standards, inappropriate model selection, model overfitting, or ineffective coding practices.’
- **Output-driven decisions:** Users can interpret algorithmic output incorrectly or apply it inappropriately, leading to unintended unfair outcomes, ineffective continuous algorithm monitoring capabilities and intervention workflows.

Risk mitigation strategies can enhance data and algorithm privacy and security, fairness, explainability, transparency and performance robustness and safety[104]. Similar or enhanced risk mitigation controls and governance standards must be applied to AI systems developed by third parties as well as to the use of open source.

Particular emphasis will be given to the use of enterprise audit software for trusted/responsible AI systems[105] [106] [107], such as the examples discussed in Chapter 6.

A. Product Liability

A product or product feature relying on AI algorithms can be defective in multiple ways[108].

In spite of rigorous testing techniques that can be performed to minimize the risks, there might be situations where an autonomous machine encounters an unforeseeable or untested situation. This may cause machine breakdown or malfunction and the design defect, in the worst-case scenario, could be dangerous.

As the use of AI becomes ubiquitous in many industry sectors such as transportation and manufacturing, safety will be paramount for human-machine interactions. Product defects could even result from communication errors between two machines or between machine and infrastructure.

To better understand the risk of AI malfunction, Swiss Re and Hitachi have recently established a partnership to explore insurance products for the rapidly evolving Industrial Internet of Things (IIoT) space with corporate customers[109].

From a liability perspective, allegations of negligent design will be the main concern. Design defects generally arise when the product is manufactured as intended, but the design causes a malfunction or creates an unreasonable risk of harm that could have been reduced or avoided through the adoption of a reasonable and safer alternative design.

USE CASE ILLUSTRATION: Aerospace and autonomous vehicles

Examples of Risk Management Service	Examples of Tools & Techniques
Conduct systematic risk analysis and examine the proposed design and AI system for foreseeable failures	Failure Modes and Effects Analysis (FMEA) is widely used in various industry sectors, such as aerospace, chemical and mechanical engineering and medical devices.
Prior to deployment and use, effective performance metrics need to be defined based on sensitivity and use of AI system to ensure targets are achieved.	DevOps in Azure Machine Learning (MLOps) helps teams to monitor model performance, audit changes to AI models, automate testing, and reproduce model outputs (Chapter 6.A).
Correction mechanisms and/or fallback options should be built in the AI system to detect and correct underperformance or put alternative appropriate processes in place until human intervention can rectify the issue.	▪ Fall-back options should be aligned with complexity of algorithms. ▪ Conduct proper assessment to prevent late-stage failure or poor models going into production. ▪ Apply Formal Verification methods to learning enabled components (LECs) of AI systems[110].
All parties involved in the development of an AI system must maintain an appropriate version control system, documentation of development and history for all components of the AI system including code, training data, trained models and parameters, tools and platforms used for the development with their configurations.	DevOps in Azure Machine Learning (MLOps) offers centralized management of model development process (data preparation, model training, model management, deployment, and monitoring) and provides mechanisms to automate, share, and reproduce models (see Chapter 6.A).
Adhere to best practices for responsible use of technology[111].	▪ Acceptable use policies (AUPs) that specify how customers should and shouldn't use the AI system and establish legal rules for appropriate conduct by the user ▪ White-Listing and Black-Listing define to whom a company will sell an AI system, depending on whether the user of the AI system is able to achieve certain process milestones or minimum quality criteria ▪ Training and guidance on best practices and risk-mitigation priorities, including integration of ethical thinking throughout the lifecycle of the AI system and tailored risk checklists to mitigate improper or harmful use by third-party actors.
Adopt standards and/or certify AI system can give assurance on technical performance and ethical standards[112].	▪ Safety standards: Safety First for Automated Driving (SAFAD) principles[113] (Remark: no specific guidance provided yet on how to design and validate a safety-critical system such as ADAS), ISO 26262, IEC 61508[114] and ARP 4761[115] for civil airborne systems and equipment. ▪ Ethical standards: IEEE P7000/P7009.

B. Professional Indemnity

When individuals act on professional advice provided by chatbots powered by AI algorithms, liability may arise if biased or erroneous information leads to decisions made by individuals that could amount to negligence.

Consequences may be particularly serious if robo-advice is used in healthcare (e.g., incorrect or poor guidance, wrong diagnosis, failure to achieve timely interventions) and in investment management (e.g., inadvertent asset allocation behavior and portfolio rebalancing).

> ***The self-learning capability of AI systems and their ability to make increasingly complex decisions by continuously adapting to new data raise concerns around retention of human control[116].***

Additionally, there is a significant data privacy and security risk if personal sensitive data used for profiling purposes and chatbot conversation history records are compromised by an adversarial attack or a data breach.

USE CASE ILLUSTRATION: Chatbot in Healthcare and Investment Management (I/II)

Examples of Risk Management Service	Examples of Tools & Techniques
1) Data availability, usability, consistency, and integrity are assessed to ensure data is suitable for informing the inferences produced by the algorithm and are sufficiently comprehensive to produce accurate and reliable outcomes. 2) Training and testing data are checked to reflect the diversity of the users (e.g., women, elderly, etc.) to avoid bias in the data. The source and lineage of data within the AI system is known, documented, traceable and auditable.	MS Datasheets for datasets include training datasets, model inputs and outputs, and model features, and help developers understand if a specific dataset is appropriate for a particular use case (Chapter 6.A).
1) The technical features of the algorithms should be documented and designed to enable understanding of how the end-to-end process works, and how it arrives at its outcomes. 2) Clear explanation of the algorithm behavior and the logic behind its design is essential to providing a reasonable explanation when required and ensuring fair treatment and outcomes.	▪ InterpretML open-source package for training interpretable models and explaining black box systems. ▪ Model Interpretability feature in Azure ML enables to explain why a model makes certain predictions, and use these insights to debug the model, validate, and check for bias. ▪ LIME provides an easily understood description of a ML classifier by perturbing the input and seeing how the predictions change. ▪ Fairlearn open-source toolkit empowers data scientists and developers to assess and improve fairness of AI systems. Please, see Chapter 6.A for references.

USE CASE ILLUSTRATION: Chatbot in Healthcare and Investment Management (II/II)

Examples of Risk Management Service	Examples of Tools & Techniques
1) Consider potential adversarial vulnerabilities during design phase and build solutions when developing algorithms (e.g., data poisoning, model leakage, hardware software infrastructure). 2) Robust cyber security measures need to be in place to detect and prevent adversarial ML attacks, hacking and other cyber-attacks that may compromise the performance of the AI system, breach in human and legal rights and result in unfair outcomes.	Rate limitation techniques help defending against AI adversarial attacks. By rate-limiting how quickly individuals or systems can submit a set of inputs to an AI system, companies can increase the effort it takes to train attackers' models[117].
Ensure that the risk of malicious actors re-identifying individuals by combining anonymized data with other sources is effectively identified and managed.	Different algorithm and data privacy mitigation strategies can be applied throughout the AI system development lifecycle: ▪ Pre-processing phase - feature selection, dataset pseudo-anonymization and perturbation. ▪ Processing phase - federated learning, differential privacy, homomorphic encryption, enabled by privacy enhancing technologies (PET), such as Microsoft Simple Encrypted Arithmetic Library (SEAL) open source[118] and OpenDP initiative / platform[119]. ▪ Deployment phase - implementation of rate-limiting and user's queries management.
Conduct a comprehensive data protection impact assessment (DPIA) in the early stages of the AI lifecycle in order to inform the design of the AI-enabled solution.	DPIA is required under the EU GDPR when dealing with data processing activities likely to result in a high risk to the rights and freedoms of natural persons[120].
Adopt industry guidelines and standards to govern responsible use of chatbots.	Examples in healthcare include: Chatbots RESET framework in healthcare[121], Benchmarks for AI-driven medical symptom checkers[122].

> “*Risk management services can be developed to enhance data and algorithm privacy and security, address issues associated with fairness, explainability and transparency, and ensure performance robustness and safety.”*
>
> Rui Manuel Melo Da Silva Ferreira
> Chief Data Governance Officer
> Zurich Insurance Group

C. Medical malpractice

There are several risks associated with the integration of AI-enabled systems into medical devices used to diagnose and treat patients. The major risks include[123]:

- **Harm to patients:** Use of biased algorithms due to biased training data can lead to misdiagnosis and entrenchment of systematic discrimination against certain groups of individuals.
- **Use of personal sensitive data without customer consent:** Lack of transparency and meaningful consent related to collection and use of health data by medical apps and services (e.g., wearables, health-related web searches, online genetic testing services) may impact individuals’ privacy and limit access to health insurance products. The ability to de-anonymize health data with relative ease further impacts individuals’ privacy.
- **Sub-standard collaborative medical decision-making between health professional and patient:** AI-driven diagnostic services lack information that health professionals can take into consideration (e.g., non-quantifiable or non-captured patient data), leading to worse health outcomes.

If an AI-generated result or recommendation leads to inaccurate conclusions, misdiagnosis, or false positives, this could amount to negligence. Even if AI is only used as an aid for referrals, if these referrals prompt investigations or procedures that are unnecessary, invasive, or lead to poorer patient outcomes, then liability may arise.

Moreover, the blurring of responsibilities between the medical device and health professionals is a concern. As fast technological progress is outpacing the ability of quality assurance teams to understand how best to provide oversight, it further increases the liability risk.

A respective use case illustration is presented on the following page.

USE CASE ILLUSTRATION: Medical devices

Examples of Risk Management Service	Examples of Tools & Techniques
Maintain an accurate representation and documentation of the AI-based medical device and its development process, including extensive risk assessment and hazard analysis, which must be continuously updated as new risks are discovered. Proactively maintain post-market surveillance for any issues that may arise concerning safety of a medical device.	Use algorithmic design history file (ADHF) to document the design input, output, review, verification, validation, transfer, and changes.
Publish details of health care algorithm development and disclose key information about how AI system is trained, for example the size, diversity and provenance of the training datasets.	To enhance confidence in AI-assisted medical diagnoses, healthcare companies can develop solutions that assign each diagnosis a confidence score and explain the probability and contribution of each patient symptom (e.g., vital signs, signals from medical reports, lifestyle traits) to that diagnosis. Clinical professionals can then understand why the conclusion was made and make a different one if required.
Provide technical support for high-quality data-gathering efforts.	Adoption of data quality standards, such as the UK NHS Digital Data Quality Maturity Index[124].
Provide medical education to prepare health-care providers to evaluate and interpret the AI systems available in the evolving health-care environment.	Examples include: a) MIT Sloan Management Executive Education – Artificial Intelligence in Healthcare b) Harvard School Public of Health – Applied Artificial Intelligence for Health Care c) University of Manchester, NHS Health Education England – AI for Healthcare: Equipping the Workforce for Digital Transformation
Adopt healthcare AI products approved by national competent authorities.	Use of certified AI/ML-based Software as a Medical Device (SaMD) thar delivers safe and effective software functionality to improve the quality of care that patients receive[125]. It should adhere to good ML practices, patient-centered approach including transparency to users, regulated scientific methods related to algorithm bias, and proven real-world performance.

Outlook

Increased cross-industry collaboration can help raise public confidence in risk mitigation methods and inspire trust in a digital society

Liability is an important tool for regulators to mitigate potential harm and already applies to the use of AI systems in well-recognized cases such as those related to medical malpractice. AI developers and users can expect new liability laws that aim to mitigate risks specific to the use of AI. In 2020, the European Parliament adopted a non-binding resolution on a civil liability regime for AI. In 2021, the EU Commission published a proposal for the first binding legal framework for AI, which is expected to prompt future regulatory initiatives in jurisdictions outside of the European Union[126].

Beyond regulatory interest, broader recognition of algorithmic risk among developers, insurers, and the public has led to increased availability of tools that can help developers build responsibly and reduce risk. Standards bodies such as the International Organization for Standardization and International Electrotechnical Commission are working on formalizing AI risk management guidance[127]. Increased awareness of risk mitigation resources such as these, developed with technical and procedural rigor, will serve to reduce algorithmic risk, increase AI adoption, and help insurers evaluate and reinforce a strong risk posture.

Technology companies and insurers can pair their strengths to mitigate algorithmic risk across industries and motivate broader adoption of risk management best practices. Technology companies have firsthand insight into cutting-edge methods to reduce risk from development to deployment, while insurance companies have a long history of incentivizing the use of best practices for enhanced safety and security in industries from transportation to healthcare and many others. Increased cross-industry collaboration can help raise public confidence in risk mitigation methods and inspire trust in a digital society.

As Microsoft's Chief Responsible AI Officer puts it, "We all need to think deeply about and account for sociotechnical impacts (of AI systems) … As the adoption of AI technologies accelerates, new and complex ethical challenges will arise[128]". Algorithmic risk management is neither solely a technical problem nor solely an operational one. Effectively mitigating this risk requires a life cycle approach to development and a holistic evaluation of where and how AI will be used within organizations and by customers.

Zurich and Microsoft are committed to advancing the responsible use of AI and data, and in finding novel ways to tackle algorithmic risk to improve trust in technology, thereby helping to address "some of the world's most pressing challenges" as referenced in the fourth opening sentence of this paper. We trust that this paper will encourage more organizations to do the same.

> *We all need to think deeply about and account for **sociotechnical impacts of AI systems**.*
>
> *Zurich and Microsoft are committed to **advancing the responsible use of AI and data**, and in **finding novel ways to tackle algorithmic risk to improve trust** in technology, thereby helping to address the world's most pressing challenges.*

References and comments

Chapter 1: Executive summary

[1] For the purpose of this white paper, AI is broadly understood as "intelligence exhibited by machines whereby machines mimic cognitive functions associated with human minds; cognitive functions include all aspects of learning, perceiving, problem solving and reasoning" (*Rockwell Anyoha*, History of Artificial Intelligence, Science in the News, 2017, https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/).
[2] *Gray, R.* "The A-Z of how artificial intelligence is changing the world", BBC, November 19, 2018 https://www.bbc.com/future/gallery/20181115-a-guide-to-how-artificial-intelligence-is-changing-the-world
[3] *McKinsey Global Institute*, "Notes from the AI frontier: Modeling the impact of AI on the world economy", September 4, 2018, https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy.
[4] *Tomašev, N., Cornebise, J., Hutter, F. et al.* AI for social good: unlocking the opportunity for positive impact. Nat Commun 11, 2468 (2020). https://doi.org/10.1038/s41467-020-15871-z.
[5] *Deloitte* "Managing algorithmic risks: Safeguarding the use of complex algorithms and machine learning", 2017 (https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-algorithmic-machine-learning-risk-management.pdf).
[6] Importantly, non-compliance with applicable data protection and privacy laws and regulations – also a key risk factor as data is essentially the raw material of AI solutions – is not in the focus of this paper.
[7] AI or algorithmic risk is sometimes also referred to as "model risk," a term which is deliberately not used in this paper so as not to conflate it with a different meaning in the insurance industry, i.e. of actuarial models being no longer accurate because of changes in the external risk landscape.
[8] Potential damages include a broad range of damages including property damage, violation of intellectual property rights as well as reputational harm.

Chapter 2: Introduction

[9] "Campolo, A./Crawford, K., "Enchanted Determinism: Power without Responsibility in Artificial Intelligence", Engaging Science, Technology, and Society, January 2020, Vol 6: pp. 1-19 (https://www.microsoft.com/en-us/research/publication/enchanted-determinism-power-without-responsibility-in-artificial-intelligence/).
[10] 2020 Edelman Trust Barometer, Edelman, 2020 (fieldwork in Oct-Nov 2019, i.e. before the pandemic), https://www.edelman.com/trust/2020-trust-barometer.
[11] "Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence", April 21, 2021, https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682
[12] "For the bots: Anne Hathaway is NOT Warren Buffett", Financial Times, March 28, 2011, https://www.ft.com/content/ab027eee-6ecc-31aa-821a-b90106bf480b.
[13] "Does Anne Hathaway News Drive Berkshire Hathaway's Stock?", The Atlantic, March 18, 2011, https://www.theatlantic.com/technology/archive/2011/03/does-anne-hathaway-news-drive-berkshire-hathaways-stock/72661/.
[14] "Gartner Survey Shows 37 Percent of Organizations Have Implemented AI in Some Form", Gartner, January 21, 2019, https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have.
[15] "Confronting the risks of artificial intelligence", McKinsey, April 26, 2019, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence.
[16] "Artificial Intelligence Risk & Governance", AIRS at Wharton, undated (not older than March 2020, from internal evidence), https://ai.wharton.upenn.edu/artificial-intelligence-risk-governance/.
[17] "5 Reasons why Microsoft should be your cybersecurity ally", Microsoft, August 02, 2017, https://www.microsoft.com/security/blog/2017/08/02/5-reasons-why-microsoft-should-be-your-cybersecurity-ally.
[18] "Microsoft Digital Defense Report", Microsoft, September 2020, https://www.microsoft.com/en-us/security/business/security-intelligence-report.
[19] "Microsoft surpasses $10 billion in security business revenue, more than 40 percent year-over-year growth", Microsoft, January 27, 2021, https://www.microsoft.com/security/blog/2021/01/27/microsoft-surpasses-10-billion-in-security-business-revenue-more-than-40-percent-year-over-year-growth/.

Chapter 3: Algorithmic risk: Intended or not, AI can foster discrimination

[20] Accenture, "Efma and Accenture Reveal Winners of Innovation in Insurance Awards 2019", June 25, 2019, (https://newsroom.accenture.com/news/efma-and-accenture-reveal-winners-of-innovation-in-insurance-awards-2019.htm).
[21] https://www.zurich.com/en/media/news-releases/2020/2020-0902-01.
[22] NPR, "Housing Department Slaps Facebook With Discrimination Charge", March 28, 2019 (https://www.npr.org/2019/03/28/707614254/hud-slaps-facebook-with-housing-discrimination-charge).
[23] United States Department of Housing and Urban Development, "Charge of discrimination", FHEO No 01-18-0323-8; March 28, 2019 (https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf).
[24] Angwin, J./Larson, J./Mattu, S./Kirchner, L., "Machine Bias", ProPublica, May 23, 2016, (https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing).
[25] Manyika, J./Silberg, J./Presten, B., "What Do We Do About the Biases in AI?", Harvard Business Review, October 25, 2019 (https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai).

References and comments

Chapter 3: Algorithmic risk: Intended or not, AI can foster discrimination

[26] Rudin, C./Wang, C./Coker, B., "The Age of Secrecy and Unfairness in Recidivism Prediction", Harvard Data Science Review, March 31, 2020 (https://hdsr.mitpress.mit.edu/pub/7z10o269/release/4).
[27] Obermeyer, Z./Powers, B./Vogeli, C./Mullainathan, S., "Dissecting racial bias in an algorithm used to manage the health of populations", Science October 25, 2019: Vol. 366, Issue 6464, pp. 447-453, (https://science.sciencemag.org/content/366/6464/447).
[28] Reuters, "Amazon scraps secret AI recruiting tool that showed bias against women", October 11, 2018 (https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G).

Chapter 4: Data and design flaws as key triggers of algorithmic liability

[29] See also McKinsey, Derisking AI, 2020, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development#.
[30] For a valuable toolkit, see UK Information Commissioner's Office: AI & data protection risk mitigation toolkit, 2021, https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ai-and-data-protection-risk-mitigation-and-management-toolkit; KPMG, AI Risk and Control Matrix, 2018 https://assets.kpmg/content/dam/kpmg/uk/pdf/2018/09/ai-risk-and-controls-matrix.pdf.
[31] See also PricewaterhouseCoopers, Model risk management of AI machine learning systems, 2020, https://www.pwc.co.uk/data-analytics/documents/model-risk-management-of-ai-machine-learning-systems.pdf.

Chapter 5: How to determine algorithmic liability?

[32] The purpose of this Chapter is to provide a foundational understanding of legal considerations and challenges around algorithmic liability and does, by no means, provide any legal advice. To answer liability questions in any specific case will require a detailed analysis of various factors, including the ones outlined in this paper, and professional legal advice for the affected jurisdictions.
[33] CMS Cameron McKenna Nabarro Olswang, "Artificial Intelligence - Who is liable when AI fails to perform?" (https://cms.law/en/gbr/publication/artificial-intelligence-who-is-liable-when-ai-fails-to-perform).
[34] For an overview of the state of AI regulation across the globe, KPMG, The Shape of AI Governance To Come, 2021 (https://assets.kpmg/content/dam/kpmg/xx/pdf/2021/01/the-shape-of-ai-governance-to-come.pdf).
[35] However, if the algorithm might be qualified as a service, products liability law might not be applicable. See Chopra, S./White, L. F., A Legal Theory For Autonomous Artificial Agents, 2011, pp. 29-69.
[36] Importantly, on April 21, 2021, the EU Commission has published progressive draft legislation on the regulation of AI systems that is envisaged to be issued as binding legislation towards the end of 2022 (https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence).
[37] Infantino, M./Wang, W., "Algorithmic Torts; A Prospective Comparative Overview", Transnational Law & Contemporary Problems, Vol. 28, pp. 309 (325 et seq.) with further references (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3225576).
[38] For the U.S. perspective, see Dempsey, J. X., "Artificial Intelligence: An Introduction to the Legal, Policy and Ethical Issues", August 10, 2020, pp. 9-34, providing an overview and selected case studies. In China, algorithmic torts have been focusing on intellectual property and unfair competition issues, such as the infringement through misuse of algorithms upon intellectual property rights and rights of dissemination on information networks. See Infantino, M./Wang, W., supra, pp. 309 (334 et seq.) with further references.
[39] For an analysis of relevant U.S. AI litigation cases in area of product liability law: Niehaus, S./Nguyen, H.: "Artificial Intelligence and Tort Liability: The Evolving Landscape", Practice Note 2020.
[40] Infantino, M./Wang, W., supra, pp. 309 (326 et seq.).
[41] Villasenor, J., Products liability as a way to address AI harms, 2019 (https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/).
[42] Villasenor, J., supra.
[43] Giuffrida, I., Liability for AI Decision-Making: Some Legal and Ethical Considerations, Fordham Law Review Vol. 88, Issue 2, 2019, p. 439 (443) (https://fordhamlawreview.org/wp-content/uploads/2019/11/Giuffrida_November_S_3.pdf).
[44] Giuffrida, I., ibid.
[45] Villasenor, J., supra.
[46] For highlights of the global policy discussion, see Chapter 6.
[47] For example, EIOPA´s GDE (Consultative Expert Group on Digital Ethics in Insurance) Report: Artificial Intelligence governance principles: towards ethical and trustworthy Artificial Intelligence in the European insurance sector, June 2021 (https://www.eiopa.europa.eu/content/eiopa-publishes-report-artificial-intelligence-governance-principles_en).
[48] For an overview of regulatory action across industries, see Greene, N./Higbee, D./Schlossberg, B., "AI, Machine Learning & Big Data", 2020, Global Legal Insights | United States, https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/usa. Also, Hanna, M., "We Don't Need More Guidelines or Frameworks on Ethical AI Use. It's Time for Regulatory Action", BRINK News (July 25, 2019) (https://www.brinknews.com/we-dont-need-more-guidelines-or-frameworks-on-ethical-ai-use-its-time-for-regulatory-action); World Economic Forum, "AI Governance—A Holistic Approach to Implement Ethics into AI", 2019 (https://www.weforum.org/whitepapers/ai-governance-a-holistic-approach-to-implement-ethics-into-ai).
[49] Chae, Y, U.S. "AI Regulation Guide: Legislative Overview and Practical Considerations", Journal of Robotics, Artificial Intelligence and Law Volume 3, No. 1, 2020. See also, Martin, J., "Regulation of Artificial Intelligence in Selected Jurisdictions | United States", Law Library of Congress, 2020, pp. 27 et seq.

References and comments

Chapter 5: How to determine algorithmic liability?

[50] Cal. Civ. Code § 1798.185(a)(16). See also Weaver, J. F., AI Under the California Privacy Rights Act, Journal of Robotics, Artificial Intelligence and Law Volume 4, No. 2, 2021 (https://www.cov.com/-/media/files/corporate/publications/2021/03/uspto-releases-report-on-artificial-intelligence-and-intellectual-property-policy.pdf).

[51] Va. Code 59.1-573(A)(5).

[52] Federal Trade Commission (FTC), Using Artificial Intelligence and Algorithms, with respect to the Fair Credit Reporting Act (FCRA) and Equal Credit Opportunity Act (ECOA) (https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms).

[53] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL) (https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html).

[54] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)) (https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html).

[55] European Commission’s proposal of an Artificial Intelligence Act, supra (fn. 1).

[56] The proposed Artificial Intelligence Act (supra) lays down a risk methodology to define “high-risk” AI systems that pose significant risks to the health and safety or fundamental rights of persons. A detailed compilation of high-risk AI systems in contained in its Annex III.

[57] See Title II of proposed Artificial Intelligence Act.

[58] Strict liability means that a party can be held liable despite the absence of (proven) fault. See, for example, non-binding EP Resolution on Civil Liability Regime for AI, supra (fn. 54).

[59] If there are several operators, all operators should be jointly and severally liable while having the right to recourse proportionately against each other; is of the opinion that the proportions of liability should be determined by the respective degrees of control the operators had over the risk connected with the operation and functioning of the AI-system. See EP Resolution on Civil Liability Regime for AI, supra (fn. 54).

[60] Council Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, 2016 O.J. (L119) 1, Rec. 71, Art. 22(1) (“GDPR”). There are several exceptions to this right, including if the automated decision-making or processing is necessary for entering into, or performance of, a contract between the data subject and a data controller. Id., at Art. 22(2).

[61] Supra.

[62] See Roberts, H./Cowls, J./Morley, J./Taddeo, M./Wang, V./Floridi, L., “The Chinese Approach to artificial intelligence: an analysis of policy, ethics and regulation”, AI & Society (2021) 36, pp. 59 (60 et seq.) (https://www.researchgate.net/publication/342246048_The_Chinese_approach_to_artificial_intelligence_an_analysis_of_policy_ethics_and_regulation).

[63] Effective as of Jan. 1, 2021, tort law is part of the PRC Civil Code.

[64] See Xuanfeng Ning, S./Wu, H., “AI, Machine Learning & Big Data | China”, Global Legal Insights, 2020 (https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/china).

[65] See Art. 1223 PCR Civil Code as well as the Interpretations of the Supreme People's Court on Several Issues concerning the Application of Law in the Hearing of Cases Involving Disputes over the Liability for Medical Damage (effective on Jan 1, 2021).

[66] The Road Traffic Safety Law (Revised Draft) issued by the Ministry of Public Security of China on March 24, 2021, marks a significant milestone as the first proposal of Chinese legislation on autonomous vehicles at the basic legal level. See Mark Schaub/Atticus Zhao, China’s Legislation on Autonomous Cars Rolls Out, April 2021 (https://www.chinalawinsight.com/2021/04/articles/corporate-ma/chinas-legislation-on-autonomous-cars-rolls-out/).

[67] The State Administration for Market Regulation (SAMR) and Standardization Administration of China (SAC) jointly published the Information Security Technology – Personal Information Security Specification (GB/T 35273-2020) (PI Specification) proposed by the National Information Security Standardization Technical Committee (TC260) on March 6, 2020.

[68] See Hong, Y., “Responses and explanations to the five major concerns of the Personal Information Security Code”, 2018 (https://mp.weixin.qq.com/s/rSW-Ayu6zNXw87itYHc).

[69] Sacks, S., “New China Data Privacy Standard Looks More Far-Reaching than GDPR. Center for Strategic and International Studies”, 2018 (https://www.csis.org/analysis/new-china-data-privacystandard-looks-more-far-reaching-gdpr).

[70] Chinese E-Commerce Law, effective on January 1st, 2019 (promulgated by the 5th session of the Standing Committee of the Thirteenth National People’s Congress, Stand. Comm. Nat’l People’s Cong., August 31, 2018.

[71] Giuffrida, I., supra (fn. 43), p. 439 (444 et seq.).

[72] To date, the European Commission does not consider it necessary to provide devices or autonomous systems a legal personality, as the harm these may cause can and should be attributable to existing persons or bodies.

[73] Supra (fn. 1). The European Commission’s proposal envisages a conformity assessment by a third party or the provider itself and, for stand-alone AI systems, registration in a central database set up and maintained by the Commission. For the conformity assessment, different procedures apply depending on the type of system and whether the system is already covered by existing product safety legislation listed in Annex II of the proposal.

[74] See Infantino, M./Weiwei Wang, W., supra, pp. 309 (330).

References and comments

Chapter 6: Principles and tools to manage algorithmic liability risk

[75] Microsoft, “Leaders look to embrace AI, and high-growth companies are seeing the benefits”, March 5, 2019 (https://news.microsoft.com/europe/features/leaders-look-to-embrace-ai-and-high-growth-companies-are-seeing-the-benefits/).

[76] Microsoft, "Microsoft launches business school focused on AI strategy, culture and responsibility", March 11, 2019 (https://blogs.microsoft.com/ai/ai-business-school/).

[77] Microsoft, "Put AI into action with AI business school" (https://www.microsoft.com/en-us/ai/ai-business-school?rtc=1). 78 Lewins, N./Morrisey, D., “Responsible AI in financial services: governance and risk management” (https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE49Atf).

[79] Microsoft, “Machine learning operations (MLOps)” (https://azure.microsoft.com/en-us/services/machine-learning/mlops/).

[80] Microsoft, “Creating AI glass boxes – Open sourcing a library to enable intelligibility in machine learning”, May 10, 2019 (https://www.microsoft.com/en-us/research/blog/creating-ai-glass-boxes-open-sourcing-a-library-to-enable-intelligibility-in-machine-learning/).

[81] Microsoft, "Model interpretability in Azure Machine Learning (preview)", February 25, 2021 (https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability).

[82] Gebru, T., et al., “Datasheets for Datasets”, Cornell University, March 19, 2020 (https://arxiv.org/abs/1803.09010).

[83] Ribeiro, M. T./Singh, S./Guestrin, C., “Why Should I Trust You? Explaining the Predictions of Any Classifier”, August 9, 2016, University of Washington (https://arxiv.org/pdf/1602.04938.pdf).

[84] Agarwal, A./Beygelzimer, A./Dudik, M./Langford, J./Wallach, H., “A Reductions Approach to Fair Classification”, Association for Computing Machinery, March 2018 (https://www.microsoft.com/en-us/research/publication/a-reductions-approach-to-fair-classification/).

[85] Fairlearn: A toolkit for assessing and improving fairness in AI, various authors at Microsoft and Allovus Design, September 22, 2020 (https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/).

[86] Marshall, A., “Securing the future of AI and machine learning at Microsoft”, February 7, 2019, https://www.microsoft.com/security/blog/2019/02/07/securing-the-future-of-ai-and-machine-learning-at-microsoft/.

[87] Microsoft SEAL (https://www.microsoft.com/en-us/research/project/microsoft-seal/).

[88] Secure Multi-Party Computation, Microsoft, June 10, 2011 (https://www.microsoft.com/en-us/research/project/multi-party-computation/).

[89] Dwork, C., “Differential Privacy”, July 2006 (https://www.microsoft.com/en-us/research/publication/differential-privacy/).

[90] SmartNoise, A differential privacy toolkit for analytics and machine learning, https://smartnoise.org/.

[91] Artificial Intelligence and the GDPR Challenge, Microsoft (https://digitaltransformation.instantmagazine.com/pub/ai-and-gdpr-2/cover/#!/cover-copy/).

[92] Schneider, J./Abraham, R./Meske, C., “AI governance for businesses” (https://arxiv.org/ftp/arxiv/papers/2011/2011.10672.pdf).

[93] ISO/IEC CD 23894.2, Information Technology — Artificial Intelligence — Risk Management, https://www.iso.org/standard/77304.html.

[94] Microsoft AI principles, Microsoft, https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6.

[95] European Commission, Ethics Guidelines for Trustworthy AI, April 8, 2019 (https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai).

[96] “Zurich announces industry-leading data commitment”, September 3, 2019, Zurich, https://www.zurich.com/en/media/news-releases/2019/2019-0903-01. Going beyond legal requirements, Zurich’s Data Commitment includes a promise never to sell customers’ personal data nor to share personal data without being fully transparent, meaning customers will always be notified if their personal data is shared, and with whom. Further, any third party with whom Zurich does share personal data is bound by an enforceable contract, which sets out how that personal data can be used. Zurich pledges to use data in the best interest of its customers. For example, data-driven insights will enable Zurich to provide innovative services that help to prevent incidents, expanding the traditional protection offered by insurance. These include smart services for home protection, and to improve health and well-being, as well as travel insurance that keeps customers out of harm’s way.

[97] Neil, L./Mittal, S./Joshi, A., “Mining Threat Intelligence about Open-Source Projects and Libraries from Code Repository Issues and Bug Reports”, (https://ebiquity.umbc.edu/_file_directory_/papers/897.pdf).

[98] CodeQL for research (https://securitylab.github.com/tools/codeql/).

Chapter 7: Managing algorithmic liability risk: Insurance for emerging exposure

[99] “Taking control: Artificial intelligence and insurance”, Lloyds, Emerging Risk Report 2019 Technology (https://assets.lloyds.com/assets/pdf-taking-control-aireport-2019-final/1/pdf-taking-control-aireport-2019-final.PDF).

[100] Balakrishnan, T./Chui, M./Hall, B./Henke, N., “McKinsey Survey on the State of AI”, November 17, 2020 (https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020).

[101] Kumar, R. S. S./Nagle, F., “The Case for AI Insurance”, Harvard Business Review, April 29, 2020 (https://hbr.org/2020/04/the-case-for-ai-insurance).

[102] AIG, Robotics Shield, “End-to-End Risk Management for the Booming Robotics Industry”, 2020, (https://www.lexingtoninsurance.com/content/dam/lexington-insurance/america-canada/us/documents/aig-robotics-shield-hs.pdf).

[103] Munich Re, “InsureAI – Guarantee the performance of your Artificial Intelligence systems” (https://www.munichre.com/en/solutions/for-industry-clients/insure-ai.html.)

References and comments

Chapter 7: Managing algorithmic liability risk: Insurance for emerging exposure

[104] Buehler, K./Dooley, R./Grennan, L./Singla, A., "Getting to know—and manage—your biggest AI risks", McKinsey article, May 3, 2021 (https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/getting-to-know-and-manage-your-biggest-ai-risks).

[105] Google and Partnership AI, "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing", 3 January 2020 (https://arxiv.org/pdf/2001.00973.pdf).

[106] "Responsible Use of Technology: The Microsoft Case Study" White Paper, WEF in collaboration with the Markkula Center for Applied Ethics at Santa Clara University, USA, February 2021 (https://www.weforum.org/whitepapers/responsible-use-of-technology-the-microsoft-case-study).

[107] Koshiyama, A. et al., "Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms", 15 February 2021 (https://cdei.blog.gov.uk/2021/04/15/the-need-for-effective-ai-assurance/).

[108] Villasenor, J., supra.

[109] SwissRe - Hitachi partnership on IIoT (Industrial Internet of Things (https://corporatesolutions.swissre.com/insights/knowledge/risk-)managers-mitigate-risks-posed-by-AI.html).

[110] Xiang, W., et al., "Verification for Machine Learning, Autonomy, and Neural Networks Survey", October 5, 2018 (https://arxiv.org/pdf/1810.01989.pdf).

[111] WEF and BSR White Paper, "Responsible Use of Technology", April 2019 (http://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology.pdf).

[112] IEEE P7000™ Projects, OCEANIS (Open Community for Ethics in Autonomous and Intelligent Systems) (https://ethicsstandards.org/p7000/).

[113] SAFAD principles, 2019 (https://www.connectedautomateddriving.eu/wp-content/uploads/2019/09/Safety_First_for_Automated_Driving.pdf).

[114] https://www.iec.ch/safety.

[115] https://www.sae.org/standards/content/arp4761/.

[116] Babic, B./Cohen, I. G./Evgeniou, T./Gerke, S., "When machine learning goes off the rails", Harvard Business Review, January-February 2021 (https://hbr.org/2021/01/when-machine-learning-goes-off-the-rails).

[117] Accenture, "Know your threat – AI is the new attack surface", 2019 (https://www.accenture.com/_acnmedia/Accenture/Redesign-Assets/DotCom/Documents/Global/1/Accenture-Trustworthy-AI-POV-Updated.pdf).

[118] Microsoft SEAL provides a set of encryption libraries that allow computations to be performed directly on encrypted data (https://www.microsoft.com/en-us/research/project/microsoft-seal/).

[119] Developing Open Source Tools for Differential Privacy (https://opendp.org/).

[120] https://www.privacy-regulation.eu/en/article-35-data-protection-impact-assessment-GDPR.htm.

[121] "Chatbots RESET - A Framework for Governing Responsible Use of Conversational AI in Healthcare", World Economic Forum in collaboration with Mitsubishi Chemical Holdings Corporation, December 2020 (https://www.weforum.org/reports/chatbots-reset-a-framework-for-governing-responsible-use-of-conversational-ai-in-healthcare).

[122] ITU/WHO Focus Group on Artificial Intelligence for Health (https://www.itu.int/en/ITUT/focusgroups/ai4h/Pages/default.aspx#:~:text=The%20ITU/WHO%20Focus%20Group,diagnosis,%20triage%20or%20treatment%20decisions (https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FG-AI4H_Whitepaper.pdf).

[123] UK Center for Data Ethics and Innovation, "CDEI AI Barometer report", June 2020 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894170/CDEI_AI_Barometer.pdf.

[124] Data Quality Maturity Index, NHS Digital, February 2021 (https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/data-quality#current-data-quality-maturity-index-dqmi-).

[125] U.S. Food and Drug Administration (FDA), Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan, January 2021; https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

Chapter 8: Outlook

[126] Proposed Artificial Intelligence Act by the European Commission, supra (fn. 1).

[127] ISO/IEC CD 23894.2 – Information Technology — Artificial Intelligence — Risk Management, International Organization for Standardization (https://www.iso.org/standard/77304.html?browse=tc).

[128] "The building blocks of Microsoft's responsible AI program", Natasha Crampton, Microsoft, January 19, 2021, https://blogs.microsoft.com/on-the-issues/2021/01/19/microsoft-responsible-ai-program/

Authors

Rui Manuel Melo Da Silva Ferreira, Ph.D., MBA
Chief Data Governance Officer
Zurich Insurance Company Ltd

Elisabeth Bechtold, Ph.D., LL.M.
Global Lead Data Governance & Oversight
Zurich Insurance Company Ltd

Franziska-Juliette Klebôn, lic.oec. MBA
Data & AI Lead
Microsoft Switzerland

Srikanth Chander Madani, MBA
Industry Advocate, Worldwide Financial Services
Microsoft Corp.

Christian Bucher, MC AI Engineer, BC IFA
Global Data & AI Principal
Microsoft Switzerland

Rachel Azafrani, MSc.
Security Strategist
Microsoft Corp.

This document is provided "as-is." Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

This document does not provide you with any legal rights to any intellectual property in any Microsoft or Zurich product. You may copy and use this document for your internal, reference purposes. Neither Microsoft Corporation, nor Zurich Insurance Group Ltd nor any of its subsidiaries accept liability for any loss arising from the use or distribution of this publication.

This publication has been jointly prepared by Zurich Insurance Group Ltd and Microsoft Corporation, and the opinions expressed therein are those of Zurich Insurance Group Ltd and Microsoft Corporation as of the date of writing and are subject to change without notice.

This publication has been produced solely for informational purposes. All information contained in this publication have been compiled and obtained from sources believed to be reliable and credible but no representation or warranty, express or implied, is made by Zurich Insurance Group Ltd or any of its subsidiaries (the ‘Group’) as to their accuracy or completeness. The Group hereby disclaims any duty to update any information in this publication.

This publication is not intended to be legal, underwriting, financial, investment or any other type of professional advice. The Group disclaims any and all liability whatsoever resulting from the use of or reliance upon this publication. Certain statements in this publication are forward-looking statements, including, but not limited to, statements that are predictions of or indicate future events, trends, plans, developments or objectives. Undue reliance should not be placed on such statements because, by their nature, they are subject to known and unknown risks and uncertainties and can be affected by numerous unforeseeable factors.

The subject matter of this publication is also not tied to any specific insurance product nor will it ensure coverage under any insurance policy. This publication does not constitute an offer or an invitation for the sale or purchase of securities in any jurisdiction.