Institute of Risk Management

IRM Digital Ethics Guidelines

irm

Developing risk professionals

Contents

Foreword

The last few years have seen a rapid expansion of digital technology into all of our professional and personal lives. This expansion was further hastened in 2020 by the health crisis, and it will continue to have a growing impact on the work of risk managers and society at large.

Failing to consider the ethical implications of new technology has created challenges historically: the first Industrial Revolution brought violent demonstrations and the rise of the Luddite movement intent on destroying the new spinning machines, which they thought would push a skilled workforce into unemployment.

What is often called the fourth Industrial Revolution will bring similar challenges to many people. These modern technologies sit in our pockets and on our desks. The decisions they facilitate every minute can have profound consequences for the jobs and daily lives of billions of individuals around the world.

As risk professionals, we have a duty to ensure that new technologies are developed and used in ways that will do the minimum amount of harm while striking a balance with the opportunities for our organisations and communities. This balance can lead to an ethical dilemma.

Consideration of the digital world's ethics requires us to ask pertinent questions of the people who are deploying the technology. It does not require a deep understanding of the technology itself. This digital ethics guideline has been created for the risk professional audience without expecting an in-depth knowledge of modern technology.

These guidelines, created by the Professional Standards Committee, comprise three principles. Each is clarified and illustrated with examples in basic language. It has been a welcome and timely piece of work which I hope you will find useful as we move further into the digitally-enabled world.

Iain Wright CFIRM

Chair, IRM

March 2021

About The Author

This publication is authorised by the Professional Standards Committee on behalf of the Institute of Risk Management. It has been developed by Mark Turner BSc (Technology) CFIRM from diverse external sources and the wider risk management community. Mark is the founder and managing director of Emsity Ltd, a risk management software consultancy.

Application of the Guideline

The Fourth Industrial Revolution (4IR) is characterised by the melding of many advanced technologies, resulting in blurring boundaries between physical, digital, and biological systems in highly complex networks.

Additionally, Artificial Intelligence technologies are advancing at breakneck speed, introducing decision-making processes independent of human involvement. Decisions made using such technologies, either directly or through human agents, can and will have wide-ranging consequences for individuals, societies, and the natural world for potentially hundreds of years.

The IRM recognises that risk professionals are instrumental in ensuring that technologies are challenged to generate 'good' results for people, organisations, society, and biological ecosystems. To help bring clarity to a technically challenging and highly complex subject, the Professional Standards Committee of the IRM has undertaken this research to support its members in being better prepared when faced with potentially ethically challenging technological scenarios.

Technology that is appropriately and ethically applied can help organisations to protect their profitability, efficiency, resilience, reputation, and attractiveness as a place to work. The guidelines presented in this document have been created to help the risk professional ask the types of questions necessary to balance the value that digital technologies undoubtedly provide against potential harm. Such a balancing act requires the consideration of the ethics of doing business in a digital environment.

When considering ethics, it is necessary to think beyond mere legalities and to consider whether a decision or action is fundamentally right or wrong. This changes the decision-making process's dimension from compliance to one of principle; the risk professional should have a strong moral compass to know right from wrong and the fortitude to challenge unethical behaviour.

Ethics must be considered in the context of the culture within which the ethical decision is being made. If an organisation's ethical culture is poor, then challenging the digital ethics in isolation will at best fall on deaf ears, or worse, result in the curtailing of your career. As such, it is recommended that these guidelines be applied with full appreciation and understanding of the current cultural climate in which you find yourself.

However, remember that if you are a member of the IRM in whatever capacity, then you are expected to abide by the IRM's Code of Conduct. If the ethical culture of an organisation is deemed to require change, the risk professional can support the development of attitudes and behaviours to improve the culture.

It should also be remembered that what is right and legal today may become wrong or even illegal in the future. The application of these guidelines should be regularly reviewed to ensure that previous decisions still pass the test of right or wrong.

Testing The Principles

When applying the principles presented in the guidelines, it is worth considering four basic tests:

1. Accountability: Are appropriate and effective governance and oversight mechanisms in place to safeguard all potential stakeholders in the technology?
2. Fairness: Could the technology discriminate, either intentionally or inadvertently, against individuals or social groups?
3. Transparency: Would the technology's actions and processes be straightforward for a knowledgeable human to understand?
4. Sustainability: Are the real-world consequences of the technology on individuals and society understood?

Each principle has a few examples of potential tests, but it is down to the risk professional to ensure that appropriate tests are considered and implemented.

Principle One

If something should not be done in the physical world, it should not be done in the digital world.

Being ethical goes beyond legal compliance. It is about doing what is right for the organisation and society at large. While it may be easy for many people to determine what is right and wrong in the real world, it may be less straightforward when considering the digital world.

The IRM code of ethics (https://www.theirm.org/media/5044/codeofconduct2015.pdf) can be used as a checklist and reference on what constitutes ethical behaviour in the physical or digital domains.

Digital technologies quickly transcend national and international borders. Laws and moral codes vary around the world.

What may be legal and morally acceptable in one country at one point in time may be considered illegal, immoral, or unethical in another country or at a later date. Risk professionals need to use their judgment and their understanding of different cultures to be sure that their application of digital technology has the widest ethical acceptability.

Where in doubt, the risk professional has an ethical responsibility to research and confirm the ethical applicability. Given that ethical principles can change over time, all digital technologies impacting humans should be reviewed throughout the life cycle of the technology - during its conception, throughout its use, and even during its archiving or disposal if appropriate.

Example questions for testing the principle:

- Is it clear who would be held accountable in a court of law (civil or criminal) if the principle is not upheld?
- Can the technology be weaponised?
- Might the technology discriminate against individuals or social groups in a way that is deemed unfair?
- Does the designed concept of fairness carry across borders?
- How transparent and legal are the gathering, processing, and retention of data and the technology's resulting actions?

Where sources are not transparent, can the complete legality of the data be proven?

- Can the technology put individual lives, societies, or biological ecosystems in danger, even unintentionally? How can you be sure?
- Can the technology be used to subvert democratic or economic principles to gain an unfair advantage to individuals or social groups?

Example: You would not want to use stolen or counterfeit goods to manufacture your physical goods. Therefore, why would you use unlicenced or cracked (essentially stolen) or unverified (potentially counterfeit) software within your organisation or your digital product?

The worldwide trade in cracked and unlicensed software in 2018 was close to $50 billion, with $20 billion accounted for in North America and Western Europe alone. Approximately 37% of software in use around the world is stolen, and malware from unlicensed software costs companies worldwide nearly $359 billion a year[1].

Potential Actions: Audits of software licences should be routinely conducted across an organisation's estate to ensure that all software products are correctly licenced and issued and are being used by the licencing agreements. Also, the audit should ascertain that the software is authentic and procured from reputable suppliers.

Principle Two

Data should be secured to ensure that it is complete, correct, consistent, and current as far as reasonably practicable.

Data represents all digital activities' foundations, and digital activities are found in every complex human system. These include transportation systems, energy and water systems, food production systems, defence and security systems, and many more. Where data security is compromised, the potential for ethical breaches increases significantly, with resulting impacts on individuals or groups, e.g., customers, employees, associated organisations, the wider population, etc.

Ensuring that data, particularly personal and commercially sensitive data, is protected against internal and external threats is an ethical imperative, both digitally and physically. Just as a building or office would be locked and access controlled, data and digital technology should be secured. Staff should be trained to defend against phishing and other social engineering attacks, and protocols should be implemented and tested to ensure that information and communications technology (ICT) security is robust.

Building organisational resilience against cyber attacks is not only good for business but aids in building confidence and trust around data security for ethical decision-making. There may also be a financial benefit in the form of reducing insurance costs[2]

Incorrect or obsolete data is potentially dangerous. Ensuring that data is not only complete and current but also untainted is critical. Intentionally false or misleading data and information have been proven to create economic and physical damage[3]. Data integrity should be rigorously assessed, mainly where there is a risk to individual lives, social groups, or the wider environment.

The computer-based analysis results should be consistent and transparent such that recipients and decision-makers can trust the data they are using. Obfuscating messages by presenting data in an inconsistent or deliberately confusing way can easily be seen as unethical, particularly if the resulting decisions favour an individual or social group.

While these guidelines cannot specify the distinction between basic marketing, propaganda, and outright unethical behaviour, risk professionals should use their risk professionals should use their judgement to ensure that ownership and accountability for potentially unethical behaviour are correctly apportioned and that unethical behaviour is appropriately called out when suspected.

Example questions for testing the principle:

- Does the data have clear ownership and accountability?
- Who is responsible for checking data integrity?
- Are outputs presented in a fair way that is transparent and clear?
- Has the data been robustly secured, and can its provenance be verified?
- Do decisions made from the data unfairly affect individuals or social groups?
- Can errors in the data cause harm to individuals, social groups, or biological ecosystems either now or in the future?
- Can errors in the data, either deliberate or accidental, subvert the democratic or economic principles to advantage an individual or social group?

Example: Data security failures cost companies millions of pounds each year through lost data, regulatory fines, and share price impacts, not to mention the loss of trust from customers and brand damage.

Potential Actions: Risk professionals from all departments should be conscious of data and digital system security risks.

Operating in functional silos can lead to potential threats being missed. Organisations should emphasise to their employees that network and data security is a risk affecting everyone and that everyone has a part to play in mitigating it.

If appropriate for your organisation, risk teams could organise a cross-functional risk workshop on data security to break down silos and seek fresh perspectives.

The agenda could include, for example, decision-making algorithms in recruitment or other HR activities; how the security of sensitive data is maintained in finance, mergers & acquisitions (M&A), research & development. Authorisation for communications releases over social media and response plans to negative social media comments; physical systems at risk from data corruption or malicious manipulation.

Principle Three

Data sources and resulting outputs should be confirmed as free from bias as far as reasonably practicable.

All humans have natural prejudices[4] and cognitive bias. Indeed science has identified at least 185 biases and heuristics[5]. When people create data, it is easy for these human biases to be incorporated into it. When the data is subsequently used by technology to perform calculations and make decisions, the human bias in the data can manifest itself in the machines' output.

Decisions made or suggestions proposed by machines are only as good as the data inputs ('garbage in, garbage out,' in computer science jargon). Where machine learning is used, care should be taken in selecting data used to train the models (which is the way artificial intelligence 'learns' how to classify data).

Indeed, the source data not only contain bias, but the decision on which data to use can also have a human bias.

For example, while it may be possible to train a recruitment algorithm to review job applications without disclosing sensitive variables such as gender, race, or sexual orientation, the machine could still favour a particular phrase such as "executed" or "captured." This would appear more frequently in applications from men.

This can occur if the machine training has been based on data that has not been appropriately sampled to give equal weight to both genders or other potentially discriminatory characteristics.[6]

The need for accurate data interpretation and transparency when relying on digital technology is crucial wherever decisions are being made about people, be they individual or social groups. Being able to explain the source of the data used to train machine learning algorithms in non-technical terms and articulate the decision-making algorithm to a third party helps remove bias.

Also, training human decision-makers who may rely on digital suggestions can introduce a more human element into the decision-making process. However, even this can be called into question if decision-makers hold a similar bias to the algorithm.

Example questions to test the principle:

- Has the training data been adequately sampled to ensure fair representation of all key characteristics?
- Has data used to train machine learning algorithms been sampled from a diverse enough population to ensure fairness is inherent in the process?
- Who has responsibility for tools used within your organisation used for decision-making? Have they been trained to look for bias both in the algorithms and in themselves?
- For tools used by customers outside of your organisation, are safeguards in place to detect unfair decision-making?
- Are results being audited and reviewed to identify potential bias?
- How much tolerance does the process have for bias?

For example, making suggestions for holiday destinations may have more tolerance for bias than making a job offer or arresting potential criminals.

- If bias is discovered in a system, what plans are in place to deal with the fallout?

Example: A job suggestion algorithm used on a social media platform presented higher paid job roles more often to men than women. The algorithm had learned this behaviour as men tended to search for and click on higher-paying jobs than women[7].

Potential Actions: During the technology development, the development team should be challenged on the data sources on which they are building their system, and also the conclusions from the data and how these were derived.

It would be appropriate for 'data bias' to be an entry in the project risk register and the necessary action allocated to a person or team with proper knowledge and training in identifying bias. The results of the risk actions would be made available to the management team.

After the technology has been published or released, the project team should ensure a process to frequently test the results to look for bias, particularly if the algorithm continues to 'learn' based on its use.

Conclusion

Digital ethics is a complex subject that requires a cross-functional approach to ensure all aspects have been considered. No single risk professional will ever be able to perceive all of the ethical risks within an organisation, so building relationships with colleagues in all business areas is essential in safeguarding the organisation from digital ethical vulnerabilities.

Applying the principles and example questions suggested in this guideline can help identify the digital ethical risks within an organisation. However, this should not be considered an exhaustive list, as digital technologies are progressing and developing at an extraordinary pace.

Risk professionals should remain abreast of risk trends and adapt their ways of working to accommodate technical developments.

Risk professionals also need to remain vigilant to changes within their organisation, particularly where new digital technologies have been adopted. The three principles outlined in this paper can be used to challenge decision-makers and ensure that the right people are asking the right questions at the right time.

An understanding of techno-jargon is not needed. However, the application of digital ethical thinking should be demonstrated clearly and reassuringly. Risk professionals should be empowered and encouraged to look for such demonstrations, ask the appropriate questions, and flag all senior management concerns.

Sources

[1] BSA Global Survey 2018 https://gss.bsa.org/

[2] Better Security Measures May Reduce Cyber Insurance Premiums | Mimecast Blog

[3] How fake data could lead to failed crops and other woes - BBC News

[4] Humans are wired for prejudice, but that doesn't have to be the end of the story (theconversation.com)

[5] List of cognitive biases - Wikipedia

[6] Amazon scraps secret AI recruiting tool that showed bias against women | Reuters

[7] Women less likely to be shown ads for high-paid jobs on Google, study shows | Google | The Guardian

The IRM's Digital Risk Management Certificate

Distance learning

International recognition

Relevant for all sectors

Qualification overview

The IRM Digital Risk Management Certificate is the essential qualification for tomorrow's risk practitioner.

The Digital Risk Management Certificate is a stand-alone qualification and can be completed in as little as six months. This qualification introduces the concepts of digital risk management and will equip risk practitioners with the ability to apply their skills in an increasingly fast-paced and changing digital world. It explains how new technologies and digitalisation are disrupting businesses, changing the risk environment for organisations of all types and posing new ethical challenges.

It looks at how appropriate risk management tools and techniques can be applied, adapted and developed in this digital context. The qualification also provides a detailed introduction to cyber security principles and practices.

What our students say

Robert Luu
Director of Customer Success
Galvanize, Singapore

"Whether you're directly in risk management practice or not, it is a great program to immerse yourself in to grasp the foundational knowledge that touches on a variety of topics of today, and the technological advancement of the future."

Emma Duggan
Risk Manager
Experian, United Kingdom

"The IRM's Digital Risk Management Certificate is extremely relevant to my role and I would urge risk professionals to consider it. It is very relevant for anyone working in technological development, as risk is everyone's responsibility."

For more information, please visit:

www.theirm.org/digi

Institute of Risk Management
2nd Floor, Sackville House
143–149 Fenchurch Street
London
EC3M 6BN

www.theirm.org

irm

Developing risk professionals