Safety in the digital age. Old and new problems.

By Jean-Christophe Le Coze

International Institute of Leadership & Safety Culture

IILSC Insights

Message from Dr. Andrew Sharman, IILSC Chief Executive Officer

At the International Institute of Leadership & Safety Culture (IILSC), we are committed to empowering leaders to create sustainable cultures of care that drive exceptional organisational performance. Our IILSC Insights whitepapers, crafted in collaboration with world-class experts in leadership, culture and workplace safety, offer valuable insights and actionable strategies for leaders and organisations striving to thrive in a rapidly changing world. Through these in-depth resources, we aim to foster collaboration, spark innovation, and equip leaders at all levels with the knowledge and tools to cultivate resilient and high-performing teams. Dive into these thought-provoking pieces to discover how you can influence change, shape safer work environments and build a lasting culture of care within your organisation.

Join us on this journey toward safer, more inclusive and future-ready workplaces.

Author of this IILSC Insights white paper

JEAN-CHRISTOPHE LE COZE

Jean-Christophe Le Coze is research director at INERIS.
He has 25 years of experience in safety, promoting a sociological, historical and epistemological approach to the field. He is the author of several books and articles on this subject, specialising in safety-critical systems, and has collaborated with a wide range of public and private organisations over the years across all sectors. He is an Associate Editor of the journal Safety Science and is a member of several Think Tanks.

Executive Summary

This IILSC Insight explores the implications of new technologies, such as Artificial Intelligence, automation and cyber security, on safety. The author joins the dots between word processing software and aviation safety for a unique illustration of the risks we face when introducing new technology into a work environment.

The author argues that technological advancements affect our relationship with the world as we know it, reshaping the environment around us, adding to the socio-technical fabric of our lives.

The digital world shapes a new context for our social interactions, our identity, our cognition and our imagination, and this has implications for our safety at work.

Small encounters, big questions

As I write this white paper, the new version of the Microsoft software program, Word, anticipates the words that I am about to type. It is based, I reckon, on one of these algorithms that we talk about a lot these days. For those who haven't yet experienced this new feature, let me explain. When I intend to write and start writing a word, like the word "algorithm" for instance, the software detects that there is a high probability after I enter the first three letters of the word, that I want to write "algorithm". Sometimes it only takes the first two letters of that word for the algorithm to make a proposition, most likely (as I can infer from by knowledge of how algorithms work) if I have already used several times the word before in the text. This increases the likelihood that I will be about to use it again.

Therefore, following the first three or two letters that I am typing, namely "alg" or "al", the software suggests the word "algorithm". I do not have to write the rest of that word; I can validate this proposition, and move on to the next word, perhaps this time again helped by the algorithm. I need to press each time a key which validates the proposition of word, which makes the cursor move immediately to the end of my word, so I can write the next one. If the proposed word does not correspond to what I had in mind, I continue to write my word as planned, and the suggestion automatically disappears.

So, I need to adjust to this added functionality, slightly changing my habits. I have not been trained for it, it is rather intuitive, like the predictive text on my smart phone when I send short messages if I select this option (which is not always convenient when switching languages for instance). This affects the way I use the keyboard, the different keys I press. I must accept or reject the suggestion, which requires me to think first then to press a key that I have never used so many times before, each time when I accept a suggestion. But it also somehow affects my relationship with the developing text.

I do not fully understand how the algorithm predicts what I am about to write although I expect it to be based on computer power, statistics and big data as explained by their designers, and promoters. Sometimes the software works correctly (meaning it predicts the word I want to write), sometimes it doesn't and sometimes it does not suggest a word at all when I expected it to. I don't know why.

This simple example is one of many encounters with what is known as the new, unfolding digital age. Algorithms, machine learning, big data and artificial intelligence (AI) are indeed the key words of these current transformations. Following a first wave of internet development coupled with the spread of personal computers in the 1990s, the 2010s brought a second level of connectedness through smart phones and tablets, generating a massive amount of data from private and public activities.

It is this new environment, built over 30 years, made of big data produced by the daily activities of people working, traveling, reading, buying and communicating, which provides an opportunity for the proliferation of algorithms, machine learning and a new generation of artificial intelligence AI (NSCT, 2016a, 2016b), beyond the first generation of GOFAI (good, old-fashioned artificial intelligence).

It affects our relationships with the world as we know it, reshaping our environment as many waves of technological changes have done in the past, from fire, to printing, to radio waves or to flying, cumulating in a mix of old and new (Edgerton, 2011). Adding to the sociotechnical fabric of our everyday lives, the digital world shapes a new context for our social interactions, our identity, our cognition, our imagination (Couldry, Hepp, 2017). The best example of the past two decades, which is a case everybody can relate to, is the smart phone.

My simple example of text assistant might not seem to change fundamentally what it is that I

do. It is one concrete example affecting work, but which remains, at this stage, uncertain in its scope, and slightly confusing. Does it fundamentally transform my approach to writing? Not only from an embodied perspective, namely in the way that I use the keyboard with my fingers and coordinate my movements to do so (i.e., automatising new patterns, developing new heuristics) which is an obvious one; but also from a less obvious point of view, regarding this time deeper thinking processes that we know are connected to our bodies, our movements but also to our material, our technological environment?

While it might not seem important, what happens when I slowly increase my dependence on the contribution of other agents, such as an algorithm that helps me write? One answer is that I slowly get used to being supported by non-human entities, machine learning entities that modify my habits, my thought processes (e.g., attention, memory, decision-making). Should I be pleased that this algorithm might, in time, get used to my style of writing? It could become a new, tailor-made assistant that I find so useful that its loss would result in some level of dissatisfaction, discontent or inconvenience. I would indeed need to return to older ways of writing, which were more cognitively demanding without this algorithmic help.

Should I be concerned beyond this cognitive dimension of delegating what I used to do in the past by myself? I do not know. Some believe this incremental process constitutes a slippery slope over which humans slowly lose control (Frischmann, Sellinger, 2018). But these algorithms have been designed by humans, by engineers. These engineers are employed by companies. Who are they? Who employs them? How do they do it? Why do they do it? How do they program it?

Because they are designed by humans, because they use data produced by humans, these programs are not value neutral. They embed choices, trade-offs. They also depend on informational or data infrastructures which must be made accessible to feed the algorithms or be created instead to ensure the development of these new software capabilities. By delving deeper into this small-scale encounter of a new software's functionality, we realise quickly of course that it is a giant corporation, Microsoft, which is behind all this.

Established as a multinational on the market for several decades, it is at the forefront of the digitalisation of our daily lives, personal and professional. A company with great power, whose managers strategically decide to invest in research and development to maintain their competitive edge over other giant corporations (i.e., GAFAM), caught in a race for a share of the immense business that this promised digital eldorado represents.
There are long chains of mediations between my writing experience and the strategy of Microsoft, which includes the negotiation of the organisation that I work for with the salespeople of this multinational to upgrade the software that I am using to write this white paper. I have no idea what is included in these contractual negotiations in terms of the data that I generate while using the software and how the algorithm may fine-tune its ability to predict the next word that I am about to write, tailoring its assistance to my usual selection of vocabulary for instance.

Should I be concerned with this facet of my writing experience? Are there also privacy issues associated with this? Considering that I also use Outlook from the same corporation, which is also a central software of my daily professional activities as I depend on my emails. These questions, when added, might not be in the end trivial questions. Are the two applications, Word and Outlook, communicating? Does this question matter? Should the cybersecurity events that we regularly hear of and the lack of updates about Microsoft's latest versions of its operating systems by its users be a concern too, for my data, for my work and for my company?
Over the past decade, such questions have been abundantly and hotly debated across

the world: data privacy protection, cybersecurity, algorithmic biases or AI (or algorithm) transparency... and a very visible thesis is that of the "surveillance capitalism" one (Zubboff, 2019). These debates are still ongoing, and more and more so. In Europe, directives on AI, data protection and cybersecurity have been released in an attempt to regulate this new and expanding landscape. This, some would say, taking a deep view of the past, is the story of humanity. To return to the cases of technological inventions and innovations mentioned above – fire, printing, radio waves or flying – entirely reconfigured our experience of ourselves, of societies, of the world.

They, too, for the most recent ones (radio waves, flying), were developed by engineers in corporations of growing size, power and influence during the 20th century, steered by managers and regulated by states, in highly complex and contentious interactions. Acquainted with this history of co-evolution between humanity and technology, between corporations, states and civil society, the digital age might seem just like another phase in the ever-changing course of societies.

So, what's new, what's old? Aren't these not perennial, recurring problems? Automation for instance, and the interactions between humans and machines and particularly computers, have already been extensively researched when they developed in the 1970s. Let's illustrate this from an explicit safety angle.

Boeing 737 Max

If I fail to make a good use of the new software described in my experience above, if I don't understand how it works and I am not trained, the resulting complications which come with it remain fairly minor problems. If it annoys me, I can probably find a way, somehow, to turn it off. Yet very similar circumstances and complications can be fatal in different contexts, in safety-critical systems for instance. A well-known example is the 737 Max of Boeing.

It illustrates how some of the questions triggered above by my small-scale encounter with a new functionality can become of the utmost importance in different contexts. It involves similar cognitive processes, of people grappling with new algorithms that require adapted responses, but the results in this second example are of two crashes involving two different Airlines, of identical, brand-new aircrafts recently purchased. What is quite problematic from a safety point of view is that these two events showed a deliberate design choice made discreetly (namely without notifying the authorities, FAA – Federal Aviation Administration) by engineers and managers of Boeing. This was a choice made discreetly for commercial reasons. Boeing was a company known for its highest safety standards, in an aviation industry equally renowned for its safe performance, including the human factor dimension.

The loss of control of the aircraft was the result of the lack of knowledge by the pilots of the presence of a software (MCAS - Manoeuvering Characteristics Augmentation System) set up to correct automatically the aircraft angle, to avoid the risk of stalling. This software was designed to be activated when the angle was reaching a certain value, a risk created by the power of the new engines fitted on the aircraft's fuselage. This angle was calculated by a probe on the side of the aircraft. In the two cases of crashes, the angle was not to blame. The failure was in the probe triggering the software's automatic corrective angle procedure in a default mode (there was no redundancy), which activated the two-horizontal stabilizers (or rear wings) of the aircraft tail. This caused the aircraft to dive in a manner completely unexpected for the pilots.

They tried to restore the aircraft's balance faced with a totally unexpected behaviour that they couldn't understand without adequate training, namely not knowing that it was the software that was activating this corrective functionality. This was knowledge that would have helped them understand the situation. The procedural fix issued between the first (in 2018) and second crash (in 2019) by Boeing, without changing the software design, to better inform pilots about the principles of its automatic corrective angle procedure, proved insufficient to protect the second crew, although they tried to apply it.

As explained, because of the size and power of the new engines, this hidden functionality was meant to correct an angle that would lead to a stall. But it was concealed by engineers and managers because it helped speed up the certification process of the aircraft and avoid a costly training program for pilots when sold with the aircrafts, cutting down costs for buyers. This was a competitive advantage in the market, dominated by Airbus. This competitive advantage would help to increase shareholder value.

The questions derived from my small encounter of the functionalities of Microsoft's predictive text software now strongly resonates in the context of a safety-critical system, aviation. Engineers, choices, software, algorithms, design, humans, multinationals, strategies and regulations shaped a specific outcome: the story of a program hidden from the view of its most affected primary users, the pilots, affecting in turn the passengers in the most dramatic of ways (Robison, 2021). This is one of the types of challenge for safety that the expansion of the digital age brings.

Safety in the digital age, challenges

Transparency of AI tools, an understanding for users of what software, machine learning processes or algorithms do when they interact with them, is essential. It affects users' ability to manage and handle situations. The safety implications are obviously enormous in all kinds of sectors, such as hospitals, railways or the nuclear industry, to take just these three examples. It concerns safety risks, such as surgery errors, explosions or derailment, but also occupational safety risks. Misunderstanding from communication issues when using digital devices augmented by algorithms, can potentially lead to injuries, or even fatalities, when work involves coordinating many people exposed to hazardous situations.

To understand how these affect the everyday practices of people at the sharp end requires anticipation, foresight. To achieve this, sufficient attention must be paid by people at the blunt end, at the highest-level (i.e., engineers, managers), during the introduction of new digital tools, generating new working conditions and patterns of interactions among people. This vocabulary of sharp-blunt end, from the 1980s and borrowed to one the most influential authors of the field of safety, James Reason, also reminds us of what is retrospectively simplified as "human error" (Reason, 1990). Errors are products of systems in which people often compensate for those systems' imperfections, including their design, as happened to the pilots of the Boeing 737 Max with dire consequences.

Along with many other psychologists and sociologists who have been interested in these issues at several levels of sociotechnical systems for decades (Turner, 1978, Perrow, 1984, Vaughan, 1996), Reason pointed at the importance of thinking about the complexity

of organisations, systems, networks, their hierarchies, their differentiation across expertise and the need for coordination and cooperation among a wide range of people, horizontally, vertically.

The digital age adds other topics to what could be described as a new version of the automation problem. These topics, introduced earlier with my questions about cybersecurity or data privacy, potentially further complexify organisations. Again, it may be argued that security is not a discovery, that the protection of personal data isn't either, and this is not entirely wrong, but new practices are slowly being established and required for securing and protecting data and performing safely (which are now increasingly regulated, as indicated).

Cybersecurity for instance, requires a range of practices by employees which were not previously expected of them, when they use their computers, when they work from home, when they travel. The ideas, tools and approaches developed in the safety field and inspired by 50 years of research in psychology, ergonomics and sociology since the pioneering work already mentioned in the two previous paragraphs, will be precious to practitioners, including safety professionals, when dealing with these problems. They emphasise the complexities of practices at the sharp and blunt end, such as in human factors, high-reliability organisations and system effects studies, all of them contributing to a human-centred understanding and approach (Hollnagel, 2020, Vaughan, 2021, Shneiderman, 2022).

The phrase "safety in the digital age"(Le Coze, Antonsen, 2023) captures these new challenges, which come with the massive changes of the operating landscape discussed in this white paper. These changes affect the conditions of safe performance in multidimensional ways, implying the need for multidisciplinary treatment of the problem (Le Coze, 2019). And this trend of digitalisation, which brings issues of AI transparency, cybersecurity and data privacy, is not a phenomenon to be understood in isolation. Other trends affecting safety are financialisation, globalisation and global warming, to only mention some of the most talked about and obvious ones (Le Coze, 2020, 2023).

Conclusion

As in previous eras, digital societies present great potential for improvements in supporting humans in their tasks while posing new challenges for sustained safe performance of public and private organisations, particularly so in safety-critical contexts. These challenges range from AI transparency, cybersecurity and data privacy protection, renewing the context in which safety must be addressed by workers, engineers, managers and regulators in their interactions.

REFERENCES

Couldry, n., Hepp, A. 2017. The Mediated Construction of Reality. Cambridge, UK: Polity Press

Edgerton, D. 2011. The shock of the old. Technology and global history since 1900. Oxford, UK: University Press.

Frischmann, B., Sellinger, E. 2018. Re-Engineering Humanity. Cambridge, Cambridge University Press.

Hollnagel, E. 2020. Synesis. The Unification of Productivity, Quality, Safety and Reliability. Abingdon, Oxon, UK: Routledge.

Le Coze, JC. 2023. Coupling and Complexity at the global scale. flows, networks, interconnectedness and synchronicity (e.g. Covid-19). Safety Science. 106193.

Le Coze, JC., Antonsen, S. (eds) 2023. Safety in the digital age. Sociotechnical Perspectives on Algorithms and Machine Learning. SpringerBriefs in Applied Sciences and Technology. Springer, Cham

Le Coze, JC. 2020. Post Normal Accident. Revisiting Perrow's classic. Boca Raton, FL. CRC Press. Taylor & Francis Group.

Le Coze, JC (ed). 2019. Safety Science Research. Evolution, challenges and new directions. Boca Raton, FL: CRC Press, Taylor&FrancisGroup.

NSTC. 2016a. Big data: a report on algorithmic systems, opportunity, and civil rights. Executive Office of the president. Retrieved in October 2019 at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

NSTC. 2016b. Preparing for the future of AI. Executive Office of the president. Retrieved in October 2019 at https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Perrow, C. (1984). Normal Accidents, Living with High-Risk Technologies. first ed. Princeton University Press, Princeton.

Robison, P. 2021. Flying blind. The 737 Max tragedy and the fall of Boeing. Doubleday.

Shneiderman, B. 2022.Human-centered AI.Oxford: Oxford University Press.

Turner, B, A. 1978. Man-made disasters. The failure of foresight. London: Wykeham Publications.

Vaughan, D. 1996. The Challenger launch decision: risky technology, culture and deviance at NASA,
University of Chicago Press, Chicago.

Vaughan, D. 2021. Dead Reckoning. Air traffic control, systems and risks. Chicago: Chicago University Press.

Zuboff, S. 2019. The age of surveillance capitalism. The Fight for a Human Future at the New Frontier of Power.London: Profile books.

INTERESTED IN MORE IILSC INSIGHTS?

This white paper is part of the IILSC Insights series, aimed at informing and inspiring high impact leadership in safety.
To receive more IILSC Insights, sign up for our IILSC newsletter at www.iilsc.com/join-us

ABOUT THE INTERNATIONAL INSTITUTE OF LEADERSHIP & SAFETY CULTURE (IILSC)

The International Institute of Leadership and Safety Culture (IILSC) has a clear mission: **To empower leaders to embrace an ever-changing world, create a sustainable culture of care and drive organisational performance.**

Owned by the executive education club CEDEP, the Institute is a global hub for leaders to meet, talk, learn, and create safety excellence. Through executive education, consulting, prestigious events, and digital learning, IILSC is creating a worldwide network of leaders from the C-suite, the OSH profession, and beyond that will turbo-charge advances in safety and health at work.

For more information, visit www.iilsc.com or scan the QR code

Would you like to propose a topic and write an IILSC Insights white paper? If so, please contact us at contact@iilsc.com

ABOUT CEDEP

CEDEP is an independent not-for-profit executive education club providing a unique and safe space for global leaders to reflect, explore, collaborate, peer-learn, grow, and succeed. It is co-run by its international members from diverse and non-competing industries who understand the value of building long-term relationships and tackling real-life business challenges within a collaborative learning community.

CEDEP empowers leaders to shape organisations for a more sustainable and positive future with transformational leadership development programs and learning experiences, co-designed with its academic team, members, clients, and non-resident faculty from the world's top business schools.

For more information, visit www.cedep.fr