POST-TRUTH SOCIETY:
THE AI-DRIVEN SOCIETY WHERE NO ONE IS RESPONSIBLE

Tatsuya Yamazaki, Kiyoshi Murata, Yohko Orito, Kazuyuki Shimizu

Meiji University (Japan), Meiji University (Japan), Ehime University (Japan),
Meiji University (Japan)

tyamazaki@meiji.ac.jp; kmurata@meiji.ac.jp; orito.yohko.mm@ehime-u.ac.jp;
shimizuk@meiji.ac.jp

ABSTRACT

This study deals with a post-truth society, which would advent due to the widespread use of artificial intelligence (AI)-based information systems using machine learning methods such as deep learning. In that society, the truths about individuals, groups, organisations, communities, society, nations and the world would become meaningless or worthless, and the situation surrounding the four factors that erode accountability in computing – many hands, bugs, blaming the computer or the computer as a scapegoat and ownership without liability (Nissenbaum, 1996) – would become worse due to the unpredictability and uncontrollability of the behaviour of AI-based systems, leading to the lack of responsibility and accountability in AI computing. To prevent the emergence of the post-truth society and regain responsibility and accountability in computing, everyone – not only ICT engineers but also end-users – has to acquire the sufficient knowledge and skill for good computing practices, in particular the ability to consider socially and ethically, through undergoing well-organised ICT educational programmes.

KEYWORDS: post-truth society, AI-based systems, unpredictability, uncontrollability, responsibility, accountability.

1. INTRODUCTION

This study deals with a post-truth society, which would advent due to the widespread use of artificial intelligence (AI)-based systems using machine learning methods such as deep learning. In that society, people would be encased in filter bubbles (Pariser, 2011) in various aspects of their everyday and social lives where what they know is unconsciously controlled by machine learning algorithms, and thus it would become very difficult for them to discover the real truth about the world. It's well known that personalised political advertisements delivered by Cambridge Analytica at the US presidential election and in the UK national referendum on membership of the EU in 2016 have allegedly contributed to the advent of post-truth politics and the resultant social fragmentation, although many have cast doubt on the effectiveness of the ads used to control voting behaviour. However, the wave of 'post-truth' ripples across society and individual lives, as well as politics.

In fact, information people can acquire in their everyday lives tends to be controlled by AI-based information systems which analyse large-scale personal databases to provide individual users with pseudo-personalised data services. Search results, postings and ads individuals view online have already been pseudo-personalised. Such data services are intended to steer individual behaviour in a way that is convenient for organisations which operate those systems. As people become increasingly dependent on the systems in terms of their information acquisition and decision making, people's thought, speech and behaviour would strongly be affected by algorithms and data used in the AI-based systems, and ultimately the systems would determine what people can know and create people's own pseudo-personalised truth.

Additionally, it has become hard for an individual to successfully control his/her identity, because information on him/her created by AI-based systems, which might actually contain stigmatic one, remains accessible online and/or in organisational databases for long periods of time, and many of others who access it can easily believe in the contents of it as the reality of him/her regardless of whether they are true or not. The truths about individuals, groups, organisations, nations and so on would become meaningless or worthless for society resulting in the emergence of the post-truth society, and people would be forced to live their post-truth lives in despair. An actual example of an AI application which functions as a threat to personal identity is one to create a deepfake, a doctored video in which a person can be made to appear as if they are doing and saying anything (Cook, 2019a). Many people including politicians and famous figures have become victims of the AI applications to masterfully edit deepfakes, being distorted their digital identities. The utilisation of this sort of AI software which can be used to conceal the truth and replace it by fakes could threaten democracy and suppress individual freedom. When it comes to deepfake porn videos, the AI applications could lead to curtailing freedom of expression and violating human dignity – especially of women – although some take a negative attitude towards regulating such contents, ironically on the ground of the protection for freedom of expression. Eventually, deepfake AI applications have not been effectively regulated so far, whereas technological efforts to fight against deepfake videos are continued (Kemeny, 2018). Here, a serious problem is that it is very difficult to find people responsible for the victims' damage created by deepfakes (Cook, 2019b).

The difficulty in clarifying the locus of responsibility is quite characteristic of the post-truth society. In this society, AI-based systems tend to function as black-boxes because their autonomous behaviour based on machine learning is not only unintelligible but also unpredictable and uncontrollable even for engineers who engage in the development and operation of the systems. When the systems are networked and work with other AI-based systems, the unpredictability and uncontrollability can be exacerbated. In addition, free/libre and open-source software (FLOSS) is often incorporated in the systems. Consequently, it is not unusual that it's very difficult to decide who is responsible, accountable and/or liable for harm caused by operations of AI-based systems. However, we cannot overlook such a technology-driven vacuum of responsibility/accountability in society.

2. A VACUUM OF RESPONSIBILITY/ACCOUNTABILITY IN AI COMPUTING

2.1. Nissenbaum's four barriers to accountability in computing

The autonomous functioning of AI-based systems using machine learning techniques leads to the unpredictability and uncontrollability of the behaviour of the systems, and provides parties relevant to the development and use of the systems, such as software engineers and system

developers, with a good excuse to evade their responsibility and/or accountability for harm the systems can bring. In fact, for example, it is not easy to decide who has to take a responsibility for a fatal traffic accident caused by an autonomous car. No one would be willing or able to be responsible for anything happen owing to the systems in the post-truth society.

More than twenty years ago when a computerised society centred on the Internet was emerging, Nissenbaum (1996) pointed out that there are four factors which erode and obscure accountability or answerability for failures, risks and harm computing brings about. This means that those who are involved in information system development and deployment work in an environment where it's hard to clarify the locus of accountability or answerability in computing, and thus it is difficult for them to appropriately take accountability for negative outcomes related to their work, no matter how conscientious they are. Consequently, developing and maintaining a professional attitude in the field of computing are extremely difficult.

According to her, the four barriers to accountability or answerability in computing are as follows:

(a) Many hands: Information systems are developed not by single programmers working in isolation but by groups or organisations. Such groups or organisations are composed of various people with a diverse range of skills and expertise such as designers, engineers, programmers, managers and salespeople. Consequently, when a system gives rise to harm, it's hard to identify who is accountable.

(b) Bugs: It is commonly recognised that bugs – a variety kinds of software errors including modelling, design and coding errors – are inevitably exist in a computer system, especially as it grows larger in scale. Therefore, harm and inconveniences caused by bugs can't be helped, and it is unreasonable to hold programmers, system engineers and designers to account for imperfections in their systems.

(c) The computer as scapegoat: When some kind of error or damage occurs, the computer systems, not human agents, associated with it are blamed. This would result in underestimating human agents' roles in and responsibility for it, and end up in the situation where none is accountable for an error or a damage.

(d) Ownership without liability: The software industry tends to demand maximal protection of the property rights to their products while denying accountability, as well as liability, for any harm their software would bring to the extent possible.

We are now in the early days of an AI-driven computerised society. However, the situation surrounding the four factors has become worse rather than better with the development and spread of information and communication technology (ICT) centred on AI.

2.2. Possible controversial scenarios

Let us consider the following possible scenarios, which would be or have been realised by introducing AI-based systems:

(a) An autonomous car killed a pedestrian. However, the growing use of driverless cars has dramatically cut down on traffic fatalities.

(b) A robot security guard accidentally killed a burglar. However, the introduction of security guard robots has highly enhanced security at office buildings and alleviated a chronic shortage of nightwatches.

(c) A robot soldier killed an opposing human soldier. Since the setting up of robot troops, the number of war dead has sharply decreased.

(d) An AI-based advertisement delivery system sent an irrelevant ad to a person based on customer profiling.

Much controversy seems to exist over Scenarios (a) – (c). Each homicide committed by the AI-controlled autonomous robot conflicts with Isaac Asimov's first/zeroth law of robotics that states a robot may not injure a human being/humanity or, through inaction, allow a human being/humanity to come to harm, whereas the homicide may be justified from a utilitarian perspective. On the other hand, Scenario (d) may seem less controversial. Receiving an irrelevant and unsolicited ad is annoying for anyone, but one just has to ignore it. However, the irrelevancy of the ad may mean incorrect personal profiling of the person was conducted, and this can lead to the distortion of his/her digital identity which would cast a negative influence over his/her life for years.

The feature common to the four scenarios is that it's very difficult to clarify who is responsible to what extent. Behind this is the unpredictability and uncontrollability of the behaviour of AI-based systems and the worsened situation surrounding Nissenbaum's four factors in the age of AI.

2.3. Obscured accountability due to the usage of FLOSS as a programme module

The recent circumstances surrounding responsible development and use of ICT have been complicated. One of the causes which have brought about the complication is the unpredictability and uncontrollability of the behaviour of AI-based systems mentioned above. Another cause is the widespread use of free/libre and open source software (FLOSS). In general, the quality of FLOSS is believed high on the ground of Linus's Law, which asserts 'given enough eyeballs, all bugs are shallow' (Raymond, 1999). However, this belief was questioned when serious bugs, the Heartbleed and Freak bugs, were discovered in the OpenSSL cryptographic software library in the mid-2010s. The discovery of those bugs revealed the fact that it was hard to ensure enough eyeballs in the processes of developing and revising this widely-used open source software (Yadron, 2014). Nevertheless, FLOSS is widely used in AI-based systems as a core programme module. Actually, for example, Hadoop and Spark have been incorporated into many of those systems for big data processing and machine learning. Additionally, it is not unusual that the source codes of programmes for AI-related data processing developed by for-profit ICT companies are disclosed so that the further development of the programmes can be conducted as a FLOSS project.

Modular design of computer programmes, which has been adopted in software development for a long time as a standard software design concept, has also contributed to the complication. This design concept assumes that a computer programme is a set of modules which are functionally independent with each other. The adoption of the concept is expected to make a software development faster, less expensive and more secure, owing to the reusability of

software modules whose quality and safety have been demonstrated, despite the unfortunate accidents caused by bugs hidden in the reused software module of Therac-25 (Leveson & Turner, 1993). Therefore, using FLOSS as a software module is considered as a good way to ensure the high quality and low-cost development of AI-based systems. This means that many people who are not necessarily personally identified can contribute to the development of AI-based systems, and FLOSS in which bugs are hidden can be incorporated into those systems. In addition, FLOSS providers usually disclaim responsibility for any damage or harm brought by the use of the software. Consequently, many hands and bugs still remain as barriers to accountability in AI computing in a more serious fashion, and it is extremely difficult to fill the vacuum of accountability in AI computing.

2.4. Scapegoated end-users

Those who engage in the development of ICT-based products and services including FLOSS, AI-related technologies and social media seem to be compelled to shift the responsibility regarding the quality of them to end-users, because the responsibility is too heavy to bear. In fact, online service users are required to agree to a detailed terms and conditions imposed by service providers prior to using the services. Such an informed consent scheme enables providers of online services to bear no responsibility for any trouble their users would face while using the services and to avoid the associated litigation risks. In addition, they can attribute responsibility for negative impacts or harm caused, for example, by personal data leaks, flaming and the spread of disinformation to end-users. Computing professionals working for such service providers can free themselves from accountability in computing. However, those who can fill this vacuum of accountability are only those computing professionals.

2.5. Autonomous systems as scapegoat

Pasquale (2015) pointed out that society has been becoming a black box due to the use of cutting-edge ICT such as Internet of Things (IoT), big data, AI and robots. The confusion of responsibility and accountability in computing seems to already be intractable. In particular, the operation of autonomous systems into which AI technology is incorporated would lead responsible people to claim that ‘it’s the system’s fault’ to evade their accountability when it causes harm. The unpredictability and uncontrollability of the behaviour of AI-based systems would create an opportunity to justify this claim and promote relevant people’s attitude of dodging responsibility by shifting the blame to those systems. These tendencies would become stronger, when an AI-based system is networked and works with other AI-based systems.

Needless to say, AI-based systems cannot become responsible or accountable agents even if they behave completely autonomously. It seems to be reasonable that the owners of such systems take responsibility in system behaviour. However, it’s quite usual that users of a system are forced to accept absolving its owner from his/her responsibility and using the system on their own responsibility in advance of using the system. This means that the barrier of ownership without liability exists in AI computing.

Autonomous robots controlled by AI-based systems will increasingly be used in various places such as production sites, offices, hospitals, nursing homes and schools, working symbiotically with people. However, if no one can take any responsibility in malfunction of those robots and resultant property destruction and physical or mental harm as well as in unexpected harm, our

future society in which the truth of the malfunction or harm has no meaning and no value would entail serious risks.

2.6. Changes in the meaning of a bug

Bugs hidden in AI-based systems may exert significant negative impacts on people's everyday and/or social lives, given the increasingly pervasive use of such systems. The trouble is that it's unclear who is responsible to explain the circumstances surrounding such negative impacts and who is liable to compensate for the resultant losses. It seems to be necessary to reconsider the meaning or definition of a bug in AI computing.

That is, even if there is neither logical nor coding error in the programmes of an AI-based system, we need to consider that there exist bugs when the system behaves in a manner that the developers of it have not intended and expected and the behaviour harms people, society and/or the environment. This type of bug may remain hidden, or the elimination of it may be prohibitively costly. If this is the case, the only possible way of debugging is to stop operating the system. This seems to undermine the value of the systems. However, continuous operation of such an AI-based system ignoring harm it brings would far more force down the value of it, and lead to losing the public's interest in AI. Responsible operation of AI-based systems is the only way of preserving the social value of them.

3. REGAINING RESPONSIBILITY/ACCOUNTABILITY IN COMPUTING

The risks entailed in the emergence of the post-truth society, where the truth has become less meaningful and worthy and no one is willing or able to accept his/her responsibility, demonstrate the social significance of the accountable management of AI artefacts through properly monitoring and controlling them, though this is really a tough challenge. However, if such management is failed, we would face the disruptions of social lives of individuals, the erosion in local communities, social fragmentation and the ruin of democracy, because AI-based systems are increasingly exerting significant influences over what we can know about our friends, acquaintances, communities, society and the world.

It is unrealistic and impractical to provide AI artefacts with legal personality and to question their responsibility. Instead, of course, organisations and/or individuals which engage in the development and use of AI systems should take their responsibility and accountability for the technological and social quality of them. Nowadays, a large majority of ICT-based system developments and operations are conducted by business organisations. The speed of ICT developments is very fast often being referred to as dog year or mouse year, and cutting-edge ICT is rapidly deployed by business organisations without disclosing sufficient information about the deployment because it is conducted in a competitive environment. Therefore, unless business organisations develop and use ICT based on the idea of 'ethics by design' taking their responsibility and accountability to the current and future generations, responses to the harm brought about by novel ICT can be made only afterwards. Only those working for organisations which engage in the development and operation of ICT-based systems can proactively address the risk of harm the operation of the systems would bring. People outside the organisations can just respond to ethical – not to speak of legal and technological – issues related to the development and use of ICT.

However, major players in the ICT industry who lead the development and use of cutting-edge ICT including AI technology seem not to willingly take their responsibility commensurate with the tremendous impact of their business activities on society. Actually, many ICT companies have maintained an attitude of 'innovative first, consider consequences afterwards'. But, if they fail to behave as professionals, their development and use of cutting-edge ICT may bring about serious social harm.

It is alleged that, in the current computerised society, there is a chronic shortage of qualified – well-trained and high-skilled – ICT engineers. As AI-based systems penetrate into society and economy, such engineers are expected to play a pivotal role in proactively addressing ethical and social issues which can be caused by AI-based systems. However, it is not easy for them to fill such a role, because of the troublesome features of AI-based systems – the unpredictability and uncontrollability of their behaviour – and the resultant obscured locus of responsibility and/or accountability in AI computing. In addition, the majority of them work for for-profit organisations, whose working environment often make it hard for them to develop their professional outlook (Murata, 2013). These suggest that not only software engineers, who have been required to build up an attitude of professionalism, but a wider range of people who are involved in computing, including end-users, need to accept their professional responsibility depending on where they stand in the AI-driven information society. In order to prevent the advent of the post-truth society, the attitude we need to develop is 'everyone has to take his/her respective responsibility in computing'.

For a wide range of people to cultivate such an attitude, appropriate ICT educational programmes must be developed. The contents and methods of the education have to be carefully examined and regularly revised, given that ICT engineers, let alone end-users, have not necessarily studied computer science and engineering at their schools and that their skill and knowledge have to be continuously renewed in accordance with the rapid advancement of ICT. The ability to consider socially and ethically has to be acquired by everyone through undergoing the educational programmes. These cannot be effective only for particular people, groups, organisations, communities and countries. In this respect, the educational programmes should be developed and revised by a non-profit body independent from any for-profit organisation and government agency, and setting up a system to issue various levels of ICT professional licences authorised by the body may be effective to encourage people to develop their professional outlook in computing suitable to their positions.

Even if engineers who engage in the development and operation of ICT-based information systems have sufficient technological knowledge and skill as ICT professionals, their lack of the knowledge and skill, as well as work habits, to ethically and socially consider would lead to serious social harm caused by their well-meaning development and operation of information systems. As a practical matter, however, it's not necessarily so easy for engineers to develop and maintain their professional outlook and appropriately address ethical and social issues. Even those ICT engineers who are well-trained and full-fledged to behave as responsible professionals would encounter difficulties in accurately predicting and properly dealing with the long-term social consequences, as well as even the immediate social impacts, of their development and operation of information systems. ICT engineers are required to humbly face up to their ineludible cognitive and intellectual limitations.

4. CONCLUSIONS

This study has examined ethical and social issues we would need to address in a post-truth society, which would advent due to the widespread use of AI-based systems using machine learning methods such as deep learning. In that society, the truths about individuals, groups, organisations, communities, society, nations and the world would become meaningless or worthless, and the situation surrounding the four factors that erode accountability in computing would become worse due to the unpredictability and uncontrollability of the behaviour of AI-based systems, leading to the lack of responsibility and accountability in AI computing. To prevent the emergence of the post-truth society and regain responsibility and accountability in computing, everyone – not only ICT engineers but also end-users – has to acquire the sufficient knowledge and skill for good computing practices, in particular the ability to consider socially and ethically, through undergoing well-organised ICT educational programmes.

We are now experiencing a hard time due to the coronavirus (COVID-19) disease pandemic. One of the most serious problems with the pandemic is a lack of accurate information as to the characteristics of the new virus. Some people who occupy professional and responsible positions in healthcare or infectious disease prophylaxis have provided inaccurate and/or wrong information about the disease. However, no one seem to have taken accountability for their misinformation delivery. Medical policies to prevent the spread of the disease differ from community to community as well as from country to country, causing general confusion as to how people can prevent themselves from being infectious. Many lay people freewheelingly deliver their irresponsible criticism to those policies online, and spread questionable or false information on the disease using social media.

The current messy situation surrounding the coronavirus is similar to the post-truth society depicted in this paper in terms of the meaninglessness and worthlessness of truths and the absence of responsible and accountable people. The pandemic will end when an effective therapy is established or a specific medicine is developed. However, the widespread use of AI-based systems will continue to be expanded due to the irresistible convenience those systems provide to the general public, although we cannot expect to have a specific cure for the social pathology which comes into existence in the post-truth society.

ACKNOWLEDGEMENTS

This study was supported by the Japan Society for Management and Information Grant-in-Aid for SIG ‘Monitoring and Control of AI Artefacts’, the Seikei University Grant-in-Aid 2020 for the research on ‘Monitoring and Control of AI Artefacts: Consideration from Economic, Social and Legal Perspectives’, and JSPS Grants-in-Aid for Scientific Research (C) 20K01920, 19K12528 and 17K03879.

REFERENCES

Cook, J. (2019a, June 12). Deepfake videos and the threat of not knowing what’s real. *Huffpost*. Retrieved from https://www.huffpost.com/entry/deepfake-videos-and-the-threat-of-not-knowing-whats-real_n_5cf97068e4b0b08cf7eb2278.

Cook, J. (2019b, June 23). Here’s what it’s like to see yourself in a deepfake porn video: there’s almost nothing you can do to get a fake sex tape of yourself taken offline. *Huffpost*.

Retrieved from https://www.huffpost.com/entry/deepfake-porn-heres-what-its-like-to-see-yourself_n_5d0d0faee4b0a3941861fced.

Kemeny, R. (2018, July 10). AIs created our fake video dystopia but now they could help fix it: new software developed by artificial intelligence researchers could help in the fight against so-called deepfake videos. *Wired*. Retrieved from https://www.wired.co.uk/article/deepfake-fake-videos-artificial-intelligence.

Leveson, N. G. & Turner, C. S. (1993). An Investigation of the Therac-25 Accidents. *IEEE Computer*, 26 (7), 18-41.

Murata, K. (2013). Construction of an Appropriately Professional Working Environment for IT Professionals: A Key Element of Quality IT-Enabled Services. In Uesugi, S. (ed.), *IT Enabled Services* (pp. 61-75). Wien: Springer.

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, 2(1), 25-42.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Raymond, E. S. (1999). *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopol, CA: O'Reilly Media.

Yadron, D. (2014, April 11). Heartbleed Bug's 'Voluntary' Origins: Internet Security Relies on a Small Team of Coders, Most of Them Volunteers; Flaw Was a Fluke. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/programmer-says-flub-not-ill-intent-behind-heartbleed-bug-1397225513.