MONITORING AND CONTROL OF AI ARTIFACTS: A RESEARCH AGENDA

Hiroshi Koga, Sachiko Yanagihara

Kansai University (Japan), University of Toyama (Japan)

koga@res.kutc.kansai-u.ac.jp; sachiko@eco.u-toyama.ac.jp

ABSTRACT

The purpose of this paper is to find a future research agenda through examination of the concept of AI artifacts. To that end, this paper is organized as follows: First, First, what AI artifacts are discussed. Next, the characteristics of AI artifacts are clarified, that is, the following two points. (1) AI artifacts contain organizational context and human agency, (2) AI artifacts fuse boundaries with natural objects. Finally, the impact is examined and future research agendas are proposed.

KEYWORDS: IT artifacts, AI artifacts, responsibility, privacy, sociomateliarity.

1. INTRODUCTION

The purpose of this paper is to focus on the idea of "AI artifacts", which has recently attracted attention in information systems research, to examine its significance, and to present a research agenda for AI artifacts.

In recent years, the expression "IT artifacts" (*signifiant*: symbolic expression) is often used instead of "information systems" in research on information systems (especially management information systems).

One of the reasons for this is the need to expand the traditional concept of information systems, as devices such as smartphone and software (app) such as SNS have penetrated daily organizational life. In other words, behind the expression "IT artifacts," we can find an attitude of daring to focus on the technical aspect under the situation where the conventional organizational behavior and various information systems are being fused.

Let's take a look back at the conventional discussion of information system development. Once upon a time, the "introduction" or "implementation" of information systems was discussed. And what happens after the introduction and implementation has been perceived as an organizational problem. However, as mentioned above, information devices are now deeply involved in all organizational behaviors. Therefore, "transformation of the organization and information systems" through the use of information systems has come to be recognized as an important management issue. This should be easy to understand, for example, by referring to the discussion of DX (digital transformation) and IoT (internet to things).

In this paper, therefore, we would like to revisit and consider the idea of AI artifacts, and then present a future research agenda, including its relevance to information ethics and business ethics.

2. PRIOR RESEARCH ON IT/AI ARTIFACTS

Herbert A. Simon was probably the first person to use the term "artifacts" in information systems research. The winner of the Alfred Nobel Memorial Prize for Economics at the National Bank of Sweden, he was not only a pioneer in modern management theory, but also an erudite and versatile person who made outstanding achievements in artificial intelligence theory and information systems research.

2.1. Revisiting the science of artifacts by H. A. Simon

Simon's work on artifacts is the well-known "The Science of Artifacts". In the same book, the differences between the "natural sciences" and the "science of artifacts" are repeatedly emphasized. According to him, natural science is a "nomothetic science," a discipline in search of universal truths. There, natural laws are assumed to exist "objectively". Moreover, natural law can be thought of in terms of purpose, i.e., value in isolation (there is, of course, no denying that science has developed in relation to the problem of God's existence).

The science of artifacts, on the other hand, aims to "construct useful things". Therefore, the evaluation of the constructed artifacts is more important than the elucidation of objective laws or the validation of the theory. For this reason, evaluation criteria (objectives and values) are important.

Simon distinguishes between these two very different views of science, and argues that science of artifacts should aim at the "science of design".

Here, we would like to expand on Simon's argument and consider it. The natural sciences are heavily influenced by "Christianity". The object of the natural sciences is the "world of God's making". He probably thought that by analyzing the natural world, he could objectively prove God's existence if he could reveal the "laws of nature," which are the blueprints of the world-building that God used. Newton, the last alchemist, is said to have discovered the "Law of Universal Attraction" as proof of the existence of God.

On the other hand, the science of artifacts covers artifacts (things made by people); if we take artifacts broadly, we can say that humanity, such as poetry and literature, and social science, such as laws and institutions, are "artifact sciences. Here, "satisfaction" is important rather than optimization, and "description (idiographic)" and "evaluation of the artifact" are more important than "law-establishment".

Of course, as Schön (1983) critiques, it must be said that Simon's concept of design is only an effective strategy for "structured problems" and falls within the bounds of "technical rationality". Schön argued that attention should be paid to trial and error and reflection in the execution phase. In the field of information systems research, from the late 1980s onwards, a research trend emerged to focus on the execution process as pointed out by Schön.

2.2. IT artifacts as s a representation of an identity for information systems research

One of the reasons why the information systems researchers refer to "IT artifacts" is that they are aware of the difference between them and computing or computer science. If we only focus on the technical aspects, we can call it a "computer system".

However, this is because we want to emphasize the aspect of an inherent social construct that is colored by background factors such as organizational culture in the place where it is used. In other words, we understand that the feature of an artifact is that it does not function as we expect it to.

In addition, information systems have "interpretive flexibility". It depends on the context of the organization whether e-mail is seen as a tool for freely expressing opinions or as a surveillance tool that censors the content of speech to identify problematic people. In other words, "the meaning of an artifact can be interpreted very differently depending on the situation in which it is placed.

From the above, IT artifacts are characterized by the fact that their uses and effects are not known until they are used. There is a reason why objective laws are so hard to figure out.

Such an understanding is close to the position of Schön's critique of Simon. Nevertheless, the result follows Simon's position of "making something useful" and "idiographic rather than nomothetic science". That's why Simon is being re-evaluated. And so information systems research, which focuses on "practice," emerged as the science of artifacts that went beyond Simon.

Wanda J. Orlikowski a leading commentator on the flexibility of interpretation, defines IT artifacts in a paper co-authored with Iacono as follows It is, they say, "those bundles of material and cultural properties packaged in some socially recognizable form such as hardware and /or software" (Orlikowski & Iacono, 2001, p.121).

Furthermore, Orlikowski & Iacono (2000) offer the following five premises of IT artifacts; That is, (1) IT artifacts, by definition, are not natural, neutral, universal, or given. (2) IT artifacts are always embedded in some time, place, discourse, and community. (3) IT artifacts are usually made up of a multiplicity of often fragile and fragmentary components, whose interconnections are often partial and provisional and which require bridging, integration, and articulation in order for them to work together. (4) IT artifacts are neither fixed nor independent, but they emerge from ongoing social and economic practices. (5) IT artifacts are not static or unchanging, but dynamic.

Thus, IT artifacts tend to emphasize organizational aspects rather than technical characteristics. Therefore, it should be called “social artifact” or "socio-materiality". From such a perspective, Lee et al. (2015) referred to the subject in the field of information system research as "IS artifacts", which are subclasses: (1) information artifacts, (2) technology artifacts, and (3) social artifacts. It points out the need to focus on the interaction between them (Lee, Thomas & Baskerville, 2015).

As described above, a characteristic of the research approach that focuses on IT artifacts is that the significance of IT artifacts is considered to be constructed in organizational practice, based on the social constructive perspective of information systems. Therefore, the main focus of such a research approach is to focus on organizational information practices and to describe the actions of IT artifacts in the process.

Nowadays, information systems and organizations are inseparably related to each other in the actual organizational information behavior such as DX. Based on this premise, IT artifacts can be said to focus on the IT elements in an organization. This has led to the emergence of a position that emphasizes two distinct characteristics of IT artifacts: organization and technology (materiality). It is the so-called socio-material school.

For example, Orlikowski (2008) used the term "entanglement" to refer to the inseparability of organizational and technical elements, which is a term from quantum mechanics. Similarly, Leonardi (2012) used the word "entanglement" to describe such a relationship, referring to the way the tiles overlap (or the way the stones in the riverbed are rounded and aligned with the flow of the river).

In Japanese Buddhist terminology, the aspect in which two things are united and inseparable is called *Ni-Ni-Fu-NI*(而二不二). Because there is light, there is shade. We can't eliminate the light and take out only the shadows. The mind and body cannot be separated (*Shiki-Shin-Fu-Ni*:色身不二). There are ideas such as the inability to separate the environment from the organism (*E-Sho-Fu-NI*:依正不二). In this respect, the idea of social materiality can be said to be extremely oriental.

2.3. The concept of AI artifacts: embedding the Organizational Context

On the other hand, it is for AI artifacts that technical aspects are often emphasized. For example, “such as artificial neural networks, specifically focusing on deep neural networks” by Tuncali *et.al*. (2018, p.1) or “data and AI models being used in the process of AI system development” by Maksimov *et. al*. (2018, p.2).

However, Rankin's argument can also be understood as "incorporating the social aspect as a function involved in intellectual judgment. If this is the case, AI artifacts can be seen to contain the social contextual or practical aspects of IT artifacts as their own functions.

Behind such a strong technical orientation, it is thought that the organizational context is embedded in machine learning. In other words, embedding organizational context means that the intelligence activities that have been entrusted to human beings have been entrusted to artifacts.

Traditionally, without fear of misunderstanding, IT artifacts have been responsible for "mechanical processing" in the field of organizational practice related to information processing. In the case of AI artifacts, however, the execution of value judgments will be automated through machine learning. In other words, the aim of artifacts has shifted from "mechanical processing of formal information" to "mechanical generation of semantic information". For example, it has become possible to predict the number of months of pregnancy based on purchase histories that are little more than formal information, or to sell the behavioral patterns of people who have declined job offers, in order to expand business opportunities through the generation of "personal information, that is, semantic information".

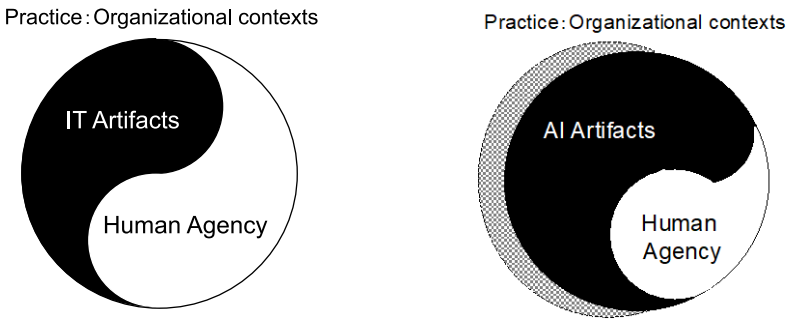
As a result, the ethics of the operation of AI artifacts have come to be questioned. At this time, AI ethics is indeed an organizational practice and is deeply related to organizational culture and values. So far, it can be said that the key to informatization practice using AI artifacts lies in organizational practice.

The left figure of Figure 1 illustrates the relationship between IT artifacts and organizational factors in organizational practice. This is in support of the Yin-Yang diagram. The whole (circle) indicates the organizational practice. A schematic representation of the complex intertwining of IT artifacts and tissue elements, with yin and yang occurring in the circle.

For organizational members, there is no awareness of yin and yang in the process of practice. No, you can't practice it consciously. When riding a bicycle, being aware of the movement of the

pedals and chain can lead to accidents. Even in organizational practice, you are probably doing it without being aware of IT artifacts and organizational practices. Hence, The left figure of Figure 1 is shown as a complex shape of a yin-yang diagram rather than a semicircle to represent the nature of the IT artifacts and organizational factors interfering with each other rather than sharply distinguishing them.

Figure 1. Differences between AI artifacts and IT artifacts.



Source: Drawing by authors

By the way, the Yin-Yang diagram seems to clearly distinguish between IT artifacts and organizational agency. Indeed, from a microscopic point of view, we can distinguish between devices such as smartphones and software. From a macroscopic perspective, however, they work together to form a hybrid.

Then, we want you to imagine a meal scene, for example. Like the attitude of picking up a camera thinking, "Photogenic! (or, to use a more recent expression, *instagrammable*)" while holding the smartphone (that is, camera) in his hands. There, the smart phone and the human being have become one, and the act of eating itself is being transformed. In other words, like the "pointillism" of an impressionist painting, the image of a unified whole is the basis of the idea of "IT artifacts", in the same way that the dots of various paintbrushes may appear up close, but the landscape appears from a distance.

Next, the right figure in Figure 1 is a simplified diagram of the relationship between AI artifacts and organizational factors in organizational practice. AI artifacts can be used to transcend specific contexts (as in the old Empty Systems in Expert Systems), and AI algorithms sometimes face ethical issues, as in the case of the U.S. Target's Pregnancy Index, in order to surrender the values of a particular organizational context. We consider these features to be the transcendence of the organizational context of AI artifacts. In the right figure of Figure 1, these characteristics are shown in the "shaded area" as "the expansion or transcendence of the context of organizational practice".

In addition, as pointed out in the previous section, AI artifacts may encompass organizational factors. Hence, AI artifacts in organizational contexts are not only intricately intertwined with organizational factors, but also serve as substitutes for some organizational factors. Therefore, the balance of yin and yang is different in comparison to the left diagram in Figure 1.

The implication shown schematically is that when considering AI artifacts, the scope of information ethics will be expanded because parts traditionally positioned in business and organizational ethics will be embedded in AI artifacts.

3. RETHINKING THE CONCEPT OF AI ARTIFACTS

In the previous section, we discussed the differences between IT and AI artifacts, focusing on previous research. The perspective of IT artifacts can be understood as a concept proposed for organizational practice as a hybrid of IT and organization, in order to focus on the micro-IT and organizational relationships therein.

So far, it can be said that the concept of IT artifacts is closely related to "*sociomateriality*", which is a recent keyword in the field of information systems research. Now, AI artifacts are an extension of IT artifacts. Therefore, I would like to assume that the focus is on everyday practice.

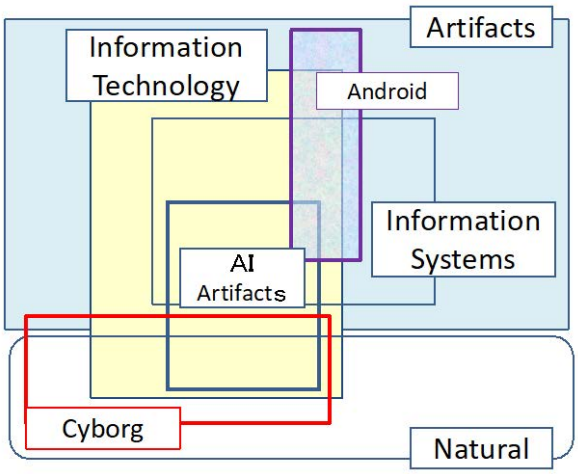
AI artifacts are more closely intertwined with organizational agencies than IT artifacts. As mentioned above, some of the organizational agency is directly incorporated into the AI artifacts, since decisions, etc., traditionally taken by humans are delegated to the system. In short, AI artifacts are encroaching on the realm of traditional organizational agency.

At this time, in the realm of computing, there is an idea of "objects" that focuses on organizational routine acts. By everyday practice, however, we mean "here and now" practices that are situated in personal, organizational, and social contexts, unlike objects as abstracted procedures. In order to do so, we would like to confirm that the assumptions are different from those of object-oriented approach.

Of course, the core algorithmic part of an AI artifact can be context-free, like the abstracted procedures of an object-oriented approach. Therefore, in the right figure of Figure 1, the shaded area shows the possibility of AI artifacts to go out of their original place of practice.

In this section, we consider the comprehensive relationship between AI and IT artifacts and natural objects and humans. Then, figure 2 was drawn by the authors to show the comprehensive relationship between AI and IT artifacts.

Figure 2. Differences between AI artifacts and IT artifacts.



Source: Drawing by authors

First, IT artifacts are concepts that encompass information systems (IS). Of course, it is also true that there are some IS that do not use IT. It should be noted that there is a non-IT based IS for providing and sharing information, for example, hallway bulletins and various documents.

Among the ISs that use IT, AI artifacts are the ones that use AI. However, some IT artifacts, including AI artifacts, may "cross the border" into natural objects. This is the major difference between IT artifacts and AI artifacts: humans with IT artifacts (e.g., artificial organs) are referred to as "cyborgs" in this paper. In particular, when AI is functioning as a substitute for the brain, the brain, which is in charge of decision-making and action, is an artifact even though it is natural in appearance, and it is considered to be a part of the fusion of the natural and the artifact, which should be the original meaning.

Conventional research on brain function has suggested that the brain is in charge of human behavior and that the human body is driven by the brain's control. In fact, it is common for brain diseases to cause a loss of functions that would have been controlled by the site.

On the other hand, the brain may be able to substitute functions in other parts of the body. In addition, it has recently become known that the brain does not control the brain top-down, but rather that the brain gives instructions for the first time when transmitters are transmitted to the brain from each organ (Valadi et.al., 2007).

Furthermore, recently, "bio-3D printers" that use cells to artificially create tissues have entered the stage of practical application, with some companies expecting to release them by the spring of 2020 (Ohshita, 2019).

However, this is the first time that live cells containing the genetic information of a patient can be grown and used like ink to create the organs themselves. Considering the fact that we are entering the clinical trial phase of artificial organs (even if they are only for experiments, there is the problem of destroying living cells), it can be said that we have completely surpassed the conventional concept of artificial organs and have entered into the field where IT can create natural organs. However, if it becomes possible to create neurons in the same way, the day will come when IT-created brain neurons will be implanted, in other words, brains as AI artifacts will be implanted in humans.

In organizational practice, IT artifacts and organizational agency formed a hybrid. However, AI artifacts may be integrated with human organs such as the brain to form hybrids (cyborgs). This in itself is likely to be an important issue for cyborg ethics. This point will be discussed in more detail in the next section.

4. THE RAINBOW CHALLENGE FACING AI ARTIFACTS: THE RESEARCH AGENDA

In this section, we present the research agenda for the management of AI artifacts. As we have repeatedly emphasized, AI artifacts have the following characteristics.

1. Two functions that organizational agencies have assumed: cultural constraint functions and intelligence functions such as judgments, which are now being delegated to AI artifacts
2. The phenomenon of hybridization (fusion) with human beings, that is, a boundary fusion with the human world, is emerging.

From these points of view, the management of AI artifacts will face the following seven challenges. Since there are seven challenges, we will refer to them as rainbow challenges in this paper. This is because, in Japan, the rainbow is considered to be composed of seven colors, and these issues are interrelated, just as the boundaries between the colors of the rainbow are ambiguous. Hence, I dared to title it Rainbow.

(A) Issues Related to System Delegation

(A-1) Privacy violations related to data use: Snooping for confidential information through lifelog analysis

(A-2) Responsibility for the accident: AI developers vs. users

(A-3) Employment Issues through System Delegation: Unemployment vs. Job Creation

(A-4) The loss of personal development opportunities: the Bernard's organizational theory perspective

(B) Issues related to human-machine hybrids

(B-1) Discussion of the view of artificial life

(B-2) The New Disparity Problem

(B-3) International Comparison of Attitudes toward Hybridization of Humans and Artifacts

The following is a brief description of each issue.

4.1. Privacy violations related to data use

First, there is a risk that data analysis generates "sensitive information". For example, the invasion of privacy will be seen as a problem, such as an US company (e.g. Target Corporation) predicting the number of gestation weeks of customers. Needless to say, in the field of business administration, customers are also considered to be organizational members (contributors) (Barnard, 1938). This is because it is difficult to continue organizational activities if customer contributions are supported.

4.2. Responsibility for the accident

As is often pointed out, there is an "*aporia*" in which the responsibility for traffic accidents in the case of automated driving lies with the system developer or user. On Japanese legal issues, for example, there are papers like the following (cf. Kawasi, 2020; Ninomiya, 2020). While there are a variety of moral, legal, and technical issues, we will only point out this point.

4.3. Employment Issues through System Delegation

The challenge of whether AI will take away jobs is also important, as has been widely debated in the past (cf. Frey & Osborne, 2017). In Japan, Arai (2017) argue that the more important issue

is the decline in human basic academic skills, rather than the replacement of many jobs with AI. Others argue that data scientists will be replaced by AI in the near future. In any case, the key issue is whether technology induces deskilling (cf. Braverman, 1974).

4.4. The loss of personal development opportunities

Delegation of intelligence /judgment functions creates the following problems. For example, where is the responsibility for accidents in autonomous driving? In the academic field of business administration, "responsibility" has been considered a key factor for the simultaneous development of individuals and organizations (Barnard, 1938). Therefore, if the responsibility is ambiguous, the organization may collapse.

4.5. Discussion of the view of artificial life

The question is whether artificial objects should be treated as life forms. This is a conundrum. Here, we would like to present a Buddhist episode that is suggestive in considering whether or not cyborgs can be considered human.

A traveler was resting in a cave. Demon/鬼 A comes carrying a corpse on his back. Demon A was about to eat with the corpse. Then another demon, B, came along. He said, "It's mine," and tried to take the body from the first demon A. The first demon A said, "Well then, let's ask the traveler there who this corpse belongs to. The traveler honestly replied that this was what the first demon A had brought. After hearing this, Demon B got angry and tore off the traveler's right arm and ate it. Then the first demon A, took pity on him and tore off the corpse's right hand and put it on the traveler. However, demon B, who came later, tore off a traveler's leg and ate it. Then the first demon, A, tore the leg off the corpse again and attached it to me. This happened repeatedly, and the demons left the cave.

The traveler thought that he had been saved, but on reflection, he couldn't tell if the corpse had really become him or not.

This may be an extreme example. However, when cyborgs become more advanced, it will be a challenge to consider whether we are a living organism or a robot.

4.6. The New Disparity Problem

In the world of sports today, the development of tools has also contributed greatly to record breaking. In the old days, an innovation in the material of the bar vaulting pole was a huge boost to the record. In recent years, marathon shoes have encouraged good records.

At the Paralympics, it is difficult to compete without a dedicated prosthetic leg or wheelchair. If these are regarded as cyborgs in the broadest sense, there is a danger of creating a disparity in the possibility of introducing AI artifacts (and by extension, economic power) as a requirement for *cyborgization*.

4.7. International Comparison of Attitudes toward Hybridization of Humans and Artifacts

The hybrid of natural and artificial objects can be rephrased as a hybrid of real and virtual. Incidentally, as with cricket, baseball and "*Ya-Kyu* (Japanese/野球) ", there are regional differences in attitudes towards hybrids. In baseball, the United States, it is no exaggeration to think that privacy can be infringed if it can provide an excellent customer experience. In cricket, that is, in Europe, it is important not to infringe on individual rights such as the right to be forgotten. In any case, in these areas, it seems to understand that AI artifacts should be under human control. On the other hand, in *Ya-Kyu*, that is, in Japan, the attitude toward hybrids is affinity.

For example, in Japan, industrial robots are given names (for example, names of female idols such as Momoe/百恵, Junko/淳子 and so on). This is because, like humans, robots are considered “comrades”. In Buddhism, it is considered "all things have the Buddha nature/一切衆生悉有仏性". Therefore, AI artifacts are also considered to be equal to humans and have little resistance to accepting AI artifacts as friends.

5. CONCLUSION

AI artifacts differ in nature from traditional IT artifacts. AI artifacts (1) merge with natural objects and the real world, and (2) come to include organizational context. Therefore, the danger of producing unintended results cannot be denied. This is a reason why AI artifacts need to be monitored and controlled. The following agendas can be pointed out as specific monitoring and control issues.

As it has been described above, we have definitely confirmed that more research is urgently needed to explore a variety of new phenomena of AI artifact monitoring and control.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP 17K03909 and JP20K01899, the Japan Society for Management and Information Grant-in Aid for SIG “Monitoring and Control of AI Artefacts”, and the Seikei University Grant-in-Aid 2020 for the research on “Monitoring and Control of AI Artefacts: Consideration from Economic, Social and Legal Perspectives”.

REFERENCES

Arai, N. (2017) Could an AI pass the entrance exam for the University of Tokyo? (Ted presentation) https://www.ted.com/talks/noriko_arai_can_a_robot_pass_a_university_entrance_exam?language=ja (2020.3.31. access)

Barnard, C.I. (1968). *The Functions of the Executive*. Harvard university press.

Braverman, Harry (1974) *Labor and monopoly capita*l. New York: Monthly Review.

Frey, C.B., & Osborne, M.A. (2017) The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254-280.

Kawai, Y.(2020). Attribution of Causality and Responsibility to a Robot. *Journal of the Robotics Society of Japan*. 38(1), 32-36 (in Japanese).

Leonardi, P.M. (2012). *Car crashes without cars: Lessons about simulation technology and organizational change from automotive design*. MIT Press.

Maksimov, Y.V., Fricker, S.A. and Tutschku, K. (2018). Artifact Compatibility for Enabling Collaboration in the Artificial Intelligence Ecosystem. *International Conference of Software Business*, pp.56-71.

Ninomiya, Y.(2020). A Direction of Automated Vehicles for Social Implementation. *Journal of the Robotics Society of Japan*. 38(1), 47-51(in Japanese).

Ohshita, J. (2019). 3D Printer, Human Tissue Using Cells, Evaluating Ricoh and JSR's New Drugs, including Blood Vessels and Dura, Nihon Keizai Shimbun Morning Edition, August 26, 2019. (in Japanese).

Orlikowski, W.J. & Iacono, C.S. (2000). The Truth Is Not Out There: An enacted view of the digital economy. Brynjolfsson, E. and Kahin, B. eds. *Understanding the Digital Economy: Data, Tools, and Research*. MIT Press, 352–380.

Orlikowski, W.J. & Iacono, C.S. (2001). Research commentary: Desperately seeking the "IT" in IT research: A call to theorizing the IT artifact. *Information Systems Research*, 12(2), 121-134.

Rankin, T. (1987). The Turing paradigm: A critical assessment. *Dialogue*, 29(2-3), 50-55.

Schön, D. A. (1983) *The Reflective Practitioner: How Professionals Think in Action*, Basic Book

Simon, H.A. (1969) *The Sciences of the Artificial*. MIT Press, Cambridge, Mass

Tuncali, C.E., Ito, H., Kapinski, J. and Deshmukh, J.V. (2018). Reasoning about safety of learning-enabled components in autonomous cyber-physical systems. *2018 55th ACM/ESDA/IEEE Design Automation Conference* (DAC), pp. 1-6.

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., & Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature cell biology*, 9(6), 654. [https://www.nature.com/articles/ncb1596 (2020.3.31. access)].