



Machine invention systems: a (r)evolution of the invention process?

Dragos-Cristian Vasilescu¹ · Michael Filzmoser¹

Received: 23 September 2020 / Accepted: 2 October 2020 / Published online: 23 November 2020
© The Author(s) 2020

Abstract

Current developments in fields such as quantum physics, fine arts, robotics, cognitive sciences or defense and security indicate the emergence of creative systems capable of producing new and innovative solutions through combinations of machine learning algorithms. These systems, called machine invention systems, challenge the established invention paradigm in promising the automation of – at least parts of – the innovation process. This paper’s main contribution is twofold. Based on the identified state-of-the-art examples in the above mentioned fields, key components for machine invention systems and their relations are identified, creating a conceptual model as well as proposing a working definition for machine invention systems. The differences and delimitations to other concepts in the field of machine learning and artificial intelligence, such as machine discovery systems are discussed as well. Furthermore, the paper briefly addresses the social and societal implications and limitations that come with the adoption of the technology. Because of their revolutionizing potential, there are widespread implications to consider from ethical and moral implications to policymaking and societal changes, like changes in the job structure. The discussion part approaches some of these implications, as well as solutions to some of the proposed challenges. The paper concludes by discussing some of the systemic benefits that can be accessed through machine invention.

Keywords Machine invention · Invention process · Automated invention · Creative systems · Machine learning

1 Introduction

Artificial intelligence and machine learning are evolving technologies with the potential of significant influence on economy and society. Owing to technology cycles significant shortening (Weaver et al. 2017) and faster adoption rates, the society needs to pay increased attention to the impact of these technologies, as well as their implications (Linstone 2011). Computerization and digital transformation are expected to replace manual and unpleasant dull, dirty and dangerous (Lin et al. 2012) work – as robotics did in the past in the industrial production – but also to assume more complex mental activities. For example jobs in service and administrative areas—the type of work that has the potential of intrinsic motivation and self-actualization (Frey and Osborne 2017).

The recent developments in gaming, such as OpenAI’s Dota 2 (OpenAI 2017), DeepMind’s AlphaGo (Silver et al.

2017) or AlphaGo Zero (Silver et al. 2017) suggest that one activity that could be automated in the future is the creation of new information, models and camera-ready content – i.e., invention. In the current understanding, humanity’s inventions and innovations stem from human reasoning and logic (Galanakis 2006). Therefore, we are our own judges of what is good or bad in the first instance. With the advent of machine invention systems, we are on the brink of changing this fundamental paradigm. This technology opens, besides its challenges—one of which is the loss of more creative jobs – a plethora of opportunities and new research avenues, expansion of current research fields and improving and thereby advancing our knowledge. As machine invention systems take advantage of all their gained knowledge and capitalize on their ability to handle large amounts of complex data, they can provide a unique out-of-the-box perspective to all fields of study. The technology promises to provide a second point of view – for the first time in history – on all human-related activities.

This paper raises awareness on the emergence of machine invention systems, that have the potential to revolutionize the way inventions are generated. Emerging fields that have wide ranging implications to society, like this one, require careful

✉ Dragos-Cristian Vasilescu
dragos.vasilescu@tuwien.ac.at

¹ Institute of Management Science, Vienna University of Technology, Vienna, Austria

thought and analysis. It is hoped and desired that articles such as this one will help developers create more robust, ethical and sustainable systems, that not only perform their function, but also protect or at least do not infringe on the well-being of our society. To achieve this goal and influence the development of this field, this paper conceptualizes and identifies the components and operations of such systems, their limitations from a societal perspective, and provides a basic understanding of their potential.

For this purpose, a model of the components of machine invention systems and their relations is derived from state-of-the-art application cases in different domains. Based on this model, the impacts and implications of this novel technology for the work environment are discussed, as well as challenges with respect to issues as ethics or intellectual property rights for society in general.

The remainder of this paper is structured as follows: in the subsequent section we apply the case study method to describe current examples of developments of machine invention systems in various areas and for various purposes. Building on this data we derive a concept of machine invention systems in the correspondent section, generalizing the features of the examples discussed previously. Thereafter we analyze the opportunities, as well as the challenges machine invention systems will pose in different areas such as integration in, and coordination with, human work processes, as well as property rights and ethical issues. The final section gives an outlook on necessary future research on machine invention systems.

2 State-of-the-art applications

Comparative case study analysis (Yin 1981, 2013; Eisenhardt 1989) is applied to generate insights on machine invention systems from current observable application cases. This exploratory qualitative method enables the identification of commonalities and patterns of a phenomenon by comparing several cases. The case study method enables theory building – which is especially important at this early stage in the field of machine invention systems. The novelty of the field also precludes representative large-sample quantitative studies and analyses (Eisenhardt 1989). The research process follows predefined steps: (1) case search in literature and practice, (2) selection of relevant and interesting cases, (3) instrument determination for data collection, (4) data collection from multiple sources, (5) case analysis, (6) cross-case comparison to validate insights from individual cases, (7) model and theory building based on the insights by cross-case comparison, and finally (8) incorporation of the results into literature and the state-of-the-art.

Given the novelty of the field and lack of a standardized terminology – which is established in the subsequent section – the search for cases is challenging. Applications are not systematically presented in scientific publications. Therefore, the search for cases was conducted using multiple sources, such as newspaper articles, webpages, interviews with experts and scientific publications, where available. Once a specific case was identified, additional materials were searched to insure a complete and accurate description for each individual case as a basis for the analysis.

The case study analysis is then applied to filter results and determine selection criteria for the inclusion of identified systems in the study. Owing to the extensive nature of the research and the limited space available, the cases reported here are presented in a simplified manner, only addressing the necessary aspects that are relevant towards obtaining a working definition. Other aspects, such as the interface between academia and industry, institutional arrangements, openness of the teams or the history of the emergence of the projects, among others, were filtered out from this paper.

The selection criteria for the analysis are *system output*, *system input*, the *type of the system* and the distinction between optimization or innovation as *system purpose*. Concerning the *system output*, we evaluate if it is a model or content not requiring further processing or interpretation, a pattern-free solution, a novel data recombination or another result not directly attributable to the input data set – under these conditions the case was classified as a potential machine invention system. In addition, *system inputs* were analyzed and all systems that feature discrete search spaces were eliminated, as this indicates an optimizing approach rather than an invention process. Cases as Deep Blue (Campbell et al. 2002) or novel wire-antenna designs (Altshuler and Linden 1997) were excluded due to their limited system inputs. Based on the *system type*, all systems that only provide support information for human users to review and decide – e.g., decision support systems or expert systems – were eliminated from the analysis. Some of the cases removed by this criteria were intelligent clinical training systems (Haddawy and Suebnukarn 2010), intelligent support for conflict resolution (Sycara 1993) and the Integrated Communications Officer (INCO) expert system project (Rasmussen et al. 1990). Finally, the selection separates cases into optimization and innovation. Optimization is based on the changes in the parametric design until no further performance improvements are possible, whereas innovation requires a novel approach carried out through qualitative changes and not just the variation of parametric values (Leon et al. 2007). Based on the premise of searching for invention systems, cases of pure optimization were removed from the analyses. The remaining examples are categorized based on their field of application.

We then excluded from the study all cases that did not match our selection criteria. The final case selection, i.e., the applications presented in this section (see Table 1 for a summary of cases grouped by similarities), are a non-exhaustive list of examples that, at least partly, include machine invention systems.

The number of cases of machine invention systems, starting with two examples found at the beginning of this research in 2014, is rapidly growing in the recent years. Although both physical and virtual applications can be found in these examples, the latter is much better represented. Cyber–physical systems, cyber security, creativity applications, cryptography and knowledge generation systems seem to be the main fields of application in the current state of development of machine invention systems.

Table 1 shows that robotics is one of the major application domains of machine invention system. Through machine invention systems virtual or physical robots learn to move. In terms of physical robots, one representative example is Darwin, a robot developed by Pieter Abbeel at UC Berkeley (Knight 2015). Using several simulated neuronal networks, also called deep-learning networks, Darwin tries to “imagine” how actions are supposed to be performed to accomplish a given task – like standing up or keeping its balance. This analysis provides a baseline for the robots’ actions and coordinates a second neuronal network that is in charge of moving the robots’ joints under the influence of sensor responses and range of possible motions.

The DeepMind project, on the other hand, developed a simulation software on robot locomotion learning (Heess et al. 2017). Given a series of constraints and a constantly changing environment, the virtual robot distributed proximal policy optimization algorithm creates, simulates, optimizes and adapts the movements of the robot with no given rules on how to perform such tasks. In the simulation, the virtual robots develop non-trivial locomotion skills difficult to program, such as jumping, crouching or turning in an unpredictable non-controlled environment.

In the domain of defense and security, a genetics-based machine learning system discovers complex new combat maneuvers for combat fighters (Smith et al. 2004). The

system uses learning classifier systems to identify high expected payoff rules and based on reinforcement learning creates new and high-performance maneuvers that have not been considered or used before.

In the field of cyber-security, unsupervised learning algorithms can check continuous streams of data by self-adjusting loops and adapt the search patterns based on a continuous learning process (Das Gupta 2017). The learning classifier systems are used in this case to identify the nature of the threat and search for models relevant to those threats. Besides reducing the time needed to identify an attack, the learning algorithm can improve the models based on the identified threats and thereby continuously improves the capabilities of the system.

In a number of applications machine systems artistically create unique forms of art. In this category, Kulitta (Quick 2014, 2015) is developing new music scores based on a structural abstract generation, musical interpretation and learning algorithms. In visual arts, convolutional neural networks can process images by the same principles of abstraction and object recognition. Using content representation, these neural networks create a new image with the same recognized objects, combined with a variety of styles learned by abstracting different picture techniques (Gatys et al. 2015).

Also, cognitive sciences can take advantage of machine invention systems. Especially in the field of quantum physics, where at first glance phenomena often seem counterintuitive for researchers, a computer algorithm called Melvin (Krenn et al. 2016) finds arrangements of quantum building blocks that produce viable solutions. The algorithm is learning by identifying useful groups of elements that ultimately lead to suggested new experiments to help researchers understand and expand their knowledge of quantum effects in fringe areas like quantum cryptography or quantum entanglement.

Based on these identified cases, we can derive the characteristics of machine invention systems and develop a definition of these systems. Because not all the output of the current examples is in a ready-to-be-used format by an end customer, but by scientists or experts – the common denominator between all of the above examples lies in the invention aspect of the output, rather than coining all output as innovations. Therefore, the term machine invention system describes more accurately the purpose and use of these systems than the term machine innovation would.

3 Machine invention systems

In the applications described above, a commonality is the learning process that enables the system to create new inventions. Therefore, a clear delimitation of where the

Table 1 Overview of considered application cases

Domain	Field	Application
Robotics	Physical Locomotion	Darwin
	Virtual Locomotion	DeepMind Locomotion
Defense and security	Aircraft maneuvers	LCS fighter
	Cybersecurity	LCS cybersecurity
Fine arts	Musical composition	Kulitta
	Image creation	CNN imaging
Cognitive science	Quantum physics	Melvin

learning process stops and the invention process starts is essential for the development of a definition for machine invention systems. The three main categories of machine learning are supervised learning, unsupervised learning and reinforcement learning (Jordan and Mitchell 2015), together with other categories like semi-supervised learning, transduction and learning to learn (Ayodele 2010), among others. The learning algorithm most frequently found in the above examples is the unsupervised learning process (among others Darwin, DeepMind Locomotion, CNN Imaging and Melvin). Unsupervised learning analyzes unlabeled data to find patterns (Jordan and Mitchell 2015) or high-level linkages (Le 2013) by using techniques like clustering or dimension reduction (Hofmann 2001). In recent years, an increasing trend of combining unsupervised learning with deep learning can be observed (Le 2013; Stollenga et al. 2014; LeCun et al. 2015). Deep learning is attempting to extract the deeper meaning of identified patterns, for example, the identification of complex contours in face-recognition instead of just simple shapes (LeCun et al. 2015; Gatys et al. 2015). Even though the algorithms used by the various applications presented above are different in nature, they do not require assisted learning, and are able to infer linkages from unstructured data.

Although machine learning systems include similar learning processes, like unsupervised learning with deep learning, the main difference to mere machine learning lies in the output. Unsupervised learning yields patterns, clusters, probabilities of events or linkages between different causes and their effects – machine invention systems go a step further, using the identified patterns and concepts to develop new scenarios, theories or experiments.

Figure 1 shows the general structure, components and operations of machine invention systems. In the first part, the learning module learns and interprets the input data or, in some applications, generates new simulated data from which it can learn. In order to achieve this result – depending on the application – the system will use a series of machine learning algorithms simultaneously. Besides unsupervised and deep learning, we can find examples of reinforcement learning, transduction and learning to learn processes in the above applications as well. Finding patterns or hidden linkages within the data, as described earlier in this section, is the task of the various machine

learning algorithms that are part of the first module. For example, Darwin simulates in its learning module how a certain movement could be made, based on the given mechanical limitations of the system, and uses the found solutions to train a high-level deep learning network on how to perform the task.

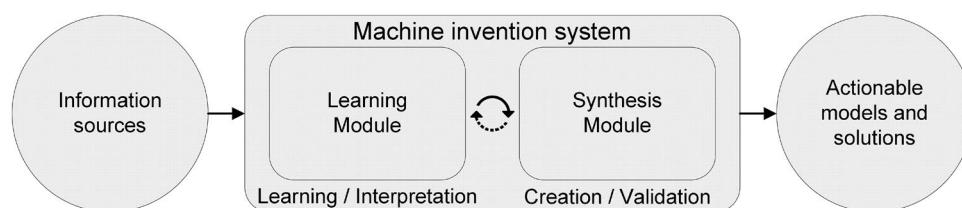
In the synthesis module, the emerging patterns are conceptualized, sets of rules are derived, models are created and validated and the potential results are predicted. Using Darwin again to exemplify: in the synthesis module, the information acquired from the learning module is used and combined with sensory inputs giving information about the real-world environment to produce viable movement solutions.

While all machine learning algorithms have feed forward capabilities, not all allow for feedback loops, like e.g. the convolutional neural networks used in creating new imagery (Gatys et al. 2015). There are, however, special cases of convolutional neural networks that allow for feedback loops (Stollenga et al. 2014). As a rule, the presence of feedback loops between the synthesis and the machine learning module increases the quality and the refinement of the delivered output, like in the case of neural networks (Stollenga et al. 2014). The delimitation between the two modules of the machine invention systems does not have to be apparent in practice, it is merely introduced for the purpose of better explaining the model.

Based on the above model and cases, we can define machine invention systems as follows: “*machine invention systems are cyber-physical or virtual systems that can create new actionable models and innovative pattern-free solutions by processing and extracting higher-level concepts and models from unorganized information sources.*”

This definition integrates a series of characteristics that machine invention systems must possess. Based on their type they are (1) cyber-physical or virtual systems, (2) that can extract information from the environment, search by themselves or generate their own data through simulations. Furthermore, they need (3) data processing abilities in the form of algorithms that can analyze the information, (4) with the ability of such systems to extrapolate the information into concepts, models or “expected-payoff” substructures. And finally, use the extrapolated information to (5) identify prescriptive models, discover novel data recombination

Fig. 1 Main components of a machine invention system



possibilities together with their expected results and generate pattern-free solutions.

It should be noted, that in our definition, we refer to cyber-physical systems just in the same way as (Lee 2008) defined them. The purpose here was to differentiate between the fully virtual algorithms, like Kulitta or DeepMind Locomotion, versus applications that use live sensorial inputs, like Darwin.

Although all different parts of the system can be attributed to different machine learning paradigms – such as clustering and dimension reduction (Hofmann 2001) for the learning module and connectivism, analogy and discovery (Domingos 2015) for the synthesis module—the particular order and combination of these paradigms generates the specific type of output of machine invention systems. This particular design allows the target function, i.e. output to be viewed as an invention. The machine invention systems are a subdomain of machine learning, as all individual components of these systems can be attributed to this larger category. Consequently, all machine invention systems are part of the domain of artificial intelligence (Michalski et al. 1983).

Machine invention systems are comparable with the much older and more established subdomain of artificial intelligence called machine discovery systems (Klosge and Żytkow 1994). Although historically the machine discovery systems started to emerge in the mid 1970's and through 1980's, most were tasked with finding low-level linkages in databases, chemistry and physics (Żytkow 1993). A few examples of such systems are STAHL and DALTON (1987), REVOLVER (1989), GELL-MANN (1990) and MECHEM (1992) (Valdes-Perez et al. 1993). The machine discovery systems were mostly used to interpret scientific reasoning, rediscover known facts or formulate general concepts of scientific activities (Valdes-Perez et al. 1993). There are, however, significant differences between the two. Żytkow explains the difference between “discovery” and “invention” in a clear fashion. “Discovery” pertains to discovering existing natural laws and constants, “invention” means the creation of a complex model or system that does not exist in nature (Żytkow 1993). Given the classification of machine invention systems in this section, both systems belong to the same family of artificial intelligence systems, but feature significant differences: they use fundamentally different types of data – labeled in case of machine discovery, unlabeled in case of machine invention. Furthermore, they apply distinct algorithms—although this difference could also be attributed to the significant advances in computational power.

From a machine learning perspective, none of the individual components of the above model is new, however, the particular structure and design of machine invention systems generates significant opportunities and challenges for society as illustrated in the subsequent section.

4 Opportunities and challenges

The development of machine invention systems could logically be seen as incremental in nature, considering its machine learning components already existed for some time. However, it challenges the fundamental way we view the invention process itself. Based on the work of Norman and Verganti (2014), this development represents a radical change for our society, as machine invention systems meet all the requirements for radical innovations (Dahlin and Behrens 2005): They are (1) a novel and unique approach and (2) have a large impact on the development of future technology.

Not only future technology is affected by this development, but the very core of what we consider human-centric activities. Innovation was a task always attributed to animals (Reader et al. 2016), but with the advent of machine invention systems, this basic paradigm is challenged. An intersection of machine systems in this sector was expected however, based on their extreme technological evolution in the past 50 years (Barnet 2004).

The necessary interaction of humans and machine invention systems, will vary based on the cultural and historical differences between and within organizations. Concerning the implementation approach, both technology and human aspects require careful consideration. An instrumental approach that considers technology to be exogenous, homogeneous, stable and predictable in its effects on human work, seems inappropriate for integrating a machine invention system in an organization's idiosyncrasies and routines. Innovation and product development might evolve to a sociomaterial practice in organizations, which needs to be addressed by appropriate studies in organization, science and technology research (Orlikowski 2007). The human–machine interaction in this domain should use both human ingenuity and the computational capabilities of machine invention systems.

Another important aspect refers to the initial design and creation of such systems. When it comes to complex systems in novel fields, the limited perspective of the technician or developer is not enough to consider the large variety of usage situations – intended or not – and consequences of the system being created. In order to mitigate at least some of the potential issues that might arise later, and to consider as many requirements as possible, appropriate technological designs must depart from the classical unilateral design through developers. A solution here is the participatory design (B.-N.Sanders 2002; Muller and Kuhn 1993) in which multiple stakeholders are engaged in the design process from the early stages on. Optimally, these stakeholders will continue to play a role through all of the lifecycle phases of the system – from design, prototyping, revisions, usage and end-of-life. This design approach is crucial also due to the responsibility the developers carry in most cases

(Thekkilakattil and Dodig-Crnkovic 2015) – especially when the systems have failed or are malfunctioning.

As with many autonomous systems, the placement of the human–machine interface is of great importance when it comes to designing the system for the best balance between efficiency and control. Here, an interface at the beginning of the process and/or at the very end, serving as a human audit of the results, seems to achieve this type of balance (Vasilescu and Filzmoser 2020).

One lurking issue here seems to be the loss of jobs to new automation methods that reduce the areas of human labor (Frey and Osborne 2017). Currently, the jobs being slowly replaced are typically repetitive, poorly cognitive and sometimes dangerous. Phasing out such jobs is not a negative development in general, although it is posing tremendous threats to those employees that currently occupy these positions. However, thinking ahead it means a shift toward a more skilled workforce in the future. There is also evidence suggesting that even though routine jobs are already declining, low-level service-oriented jobs are increasing (Cully 1999). Reich (1992) predicts that both “symbolic-analysts” and service-oriented jobs will gain importance in the twenty-first century workforce due to the inherent value placed on human interaction as opposed to technology-only interactions (Ganguli and Roy 2011).

From a different perspective, the deskilling of the workforce due to reduced interactions is another important issue (Cooley 1987). Although machine invention does increase efficiency and offers new avenues for development, the social isolation surrounding working with machines can lead not only to psychological effects (Vega and Brennan 2000), but to deskilling of the staff. One of the most basic forms of learning, “learning by doing”, (Arrow 1971) can no longer be used to learn the intricacies of a new skill when it comes to machine invention systems, as the people working with these systems become merely observers, maintainers and sometimes decision factors on the outputs.

As a society, it is also unclear what the acceptance rates and limitations of the machine-generated solutions will be. Just because a new technology appears, it does not mean it will be accepted and implemented. Acceptance is closely related to social interactions and other factors such as political and economic issues (Grunwald 2000). Proper planning and policymaking can steer the development of these systems and avoid the pitfalls (social, environmental, financial, among others) of letting the technology follow its own independent technological evolution based on the interests of a few. Even if machine invention systems can generate new content, models and applications, it is currently unclear to which extent the industries applying them or the policy makers will decide if the generated knowledge will be automatically implemented, or to which degree there will be a human decision-maker involved in the implementation

process. Some limitations should help alleviate part of the concerns. Given the tremendous potential of such systems it is clear that capabilities alone do not necessarily mean they should be acted upon. Some guiding (ethical) principles should be embedded into such autonomous knowledge-generating systems. The question of acceptance can be viewed from a different perspective as well, as the technology itself changes how humans perceive reality (Ciborra 2006, 2007). This effect is enhanced by the humans’ acceptance of new technologies appearing over time, as they use and become more comfortable with them (Klamer and Allouch 2010; De Graaf and Allouch 2013). In the end, machine agency cannot be separated from human agency – they both have to be analyzed together.

A significant challenge faced by the machine invention systems is the fact that the expected output is – in the best case – a novel model, application or actionable experiment and therefore should be protected by patenting, copyright or other intellectual property rights. Yet, the current policymaking is still working on the premise of human-only inventions. Changes in global policymaking are needed, as autonomous technologies become more accepted by society. Such a difficulty has become apparent in the case of accidents involving driverless cars (Ganesh 2017), where the accountability cannot be clearly assigned as in the traditional driver-based system.

Another, perhaps more important discussion regarding the generation of novel solutions to problems through such systems lies in the arising ethical issues. For example, a robot equipped with a machine invention system could generate solutions for the most efficient way of taking life in specific scenarios; these solutions could completely defy human logic and instincts, and therefore cannot be counteracted by humans. Such ethical concerns should be thoroughly discussed and systemically applied to all autonomous-prone technologies. There are already models to assess the ethical implications of emerging technologies, like the ethical technology assessment, the techno-ethical scenario or the Ethical Issues of Emerging ICT Applications (ETICA) approach (Busby et al. 2008; Brey 2012), but there is no globally-accepted axiom or set of principles regarding the ethics of autonomous technologies yet. Recently, an increasing trend toward ethical considerations can be found in literature (Lin et al. 2012; Nørskov 2017; Lanfranchi 2017) and in practice (DeepMind 2018).

The human invention process is chaotic and often fruitless. Nevertheless, the human invention process brought to life all the technological advancements of our society until recently. A machine attempting to innovate on the other hand has a more methodological approach, by virtue of its nature. In contrast, the machine invention process can generate large numbers of results, but due to the lack of basic understanding and discernibility, these results need to be analyzed and

curated. This further analysis should be conducted in first instance by the systems themselves for quality purposes, and in a second instance by human decision makers, due to ethical considerations. Although this means that only a fraction of the computed solutions are useable, the large processing capabilities of modern systems mitigate, in part, this shortcoming.

Based on these considerations, advancements in the direction of machine invention would have the potential to create significantly different inventions compared to human inventions. With the advancement of machine learning and the innate capability of assessing and connecting large amounts of data, their potential to realize previously unrelated connections is far beyond human capability. From creating new connections in data, there are only a few more steps to derive new inventions. The potential for different inventions, together with a larger number of inventions altogether, can be a tremendous asset for all fields, industries, economy and society in general.

5 Conclusion and outlook

This study identifies cases of machine invention systems in use in different domains and based on a comparative case study analysis derives a model and definition of machine invention systems. Several findings can be derived from the analyses of the cases described in this paper. Based on the cases, we propose that cyber-physical or virtual systems can generate inventions. Such inventions can be different in nature, from prescriptive models to pattern-free solutions. Furthermore, machine invention systems use machine learning algorithms and known machine learning paradigms to generate their output. However, these findings also generate questions that need to be addressed in future research: do the applied techniques differ for various domains and applications, and why? Are machine invention systems really superior to human-only innovation processes, or is a combination preferable? If so, how should the optimal interaction between human and machine in innovation tasks look like? And how can we overcome the resistance towards acceptance of machine learning systems themselves as well as towards the results they produce?

Considering the early stage of the emerging machine invention systems, there is only a restricted set of cases this study can build upon and therefore a number of limitations exist. Currently, the limitations of such systems are unclear, as are the fields, aside the mentioned ones, in which machine invention systems can be applied. Although the selected cases presented in this paper are very different from each other, the addition of new cases under the umbrella of machine invention systems can add relevant details or bring

changes to the concept of machine invention as it is currently understood.

The challenges and opportunities presented in this paper show the importance of further developing the theoretical framework encompassing machine invention systems, in order to better understand the boundaries, capabilities and limitations of the current state-of-the-art. Future research in this field has to focus on an in-depth analysis of the algorithms and principles that machine invention systems use. Based on these analyses, the range of possible inventions, use cases, and application fields can be derived, and additional application possibilities in previously unexplored areas uncovered.

Moreover, additional research is required concerning the optimal conditions about the acceptance of machine invention systems in today's various work environments to complement the existing systems. This raises important research questions: (i) To which extent can and should invention be automated and (ii) How can and should human employees and machine invention systems interact? The goal is to provide insights on reducing the deployment risks of this technology and allow for a smooth diffusion in the industrial applications. Such complementary innovation processes have to be developed to implement and support the collaboration of two very different approaches to invention, human and machine alike.

It is important to mention that technology itself is not neutral (Balabanian 2006) and given the intrinsic connection between human and technology, all stakeholders should play their roles in the development and beneficial use of machine invention systems. Some nefarious uses are easier to prevent than others – i.e. through policymaking – nevertheless, due to the potential capabilities of such technologies, wide-ranging implications should be considered on a case-by-case basis. It is the hope of the authors that this paper represents an important step for further investigations and toward the ethical development of this field. This technology presents a significant addition and change to our current innovation processes, and through careful development and implementation such systems can have deep and positive consequences for our society.

Funding Open access funding provided by TU Wien (TUW).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Altshuler EE, Linden DS (1997) Wire-antenna designs using genetic algorithms. *IEEE Antennas Propag Mag* 39:33–43. <https://doi.org/10.1109/74.584498>
- Arrow KJ (1971) The economic implications of learning by doing. *Readings in the Theory of Growth*. Springer, London, pp 131–149
- Ayodele TO (2010) Types of machine learning algorithms. In: *New advances in machine learning*. InTech, pp 19–48
- B.-N. Sanders E, (2002) From user-centered to participatory design approaches. In: Frascara J (ed) *Design and the Social Sciences*. CRC Press, Florida, pp 1–8
- Balabanian N (2006) On the presumed neutrality of technology. *IEEE Technol Soc Mag* 25:15–25
- Barnet B (2004) Technical machines and evolution. *CTheory* 1:3–16
- Brey PAE (2012) Anticipatory ethics for emerging technologies. *Nano-Ethics* 6:1–13. <https://doi.org/10.1007/s11569-012-0141-7>
- Busby H, Hervey T, Mohr A (2008) Ethical EU law? The influence of the European group on ethics in science and new technologies. *Eur Law Rev* 33:803–842
- Campbell M, Hoane AJ, Hsu F (2002) Deep Blue. *Artif Intell* 134:57–83. [https://doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/10.1016/S0004-3702(01)00129-1)
- Ciborra C (2006) *Imbrication of representations: risk and digital technologies*. Palgrave Macmillan, London
- Ciborra C (2007) Digital technologies and risk: A critical review. *Risk Complex ICT Cheltenham, UK Edw Elgar*, pp 23–45
- Cooley M (1987) *Architect or Bee? The human price of technology*. The Hogarth Press, London
- Cully M (1999) A more or less skilled workforce? Changes in the occupational composition of employment, 1993–1999. *Aust Bull Labour* Adelaide 25:98–104
- Dahlin KB, Behrens DM (2005) When is an invention really radical? *Res Policy* 34:717–737. <https://doi.org/10.1016/j.respol.2005.03.009>
- Das Gupta S (2017) Detecting threats with analytics and machine learning (ML). *J Econ Perspect* 31(2):87–106
- De Graaf MM, Allouch SB (2013) Exploring influencing variables for the acceptance of social robots. *Robot Auton Syst* 61:1476–1486
- DeepMind (2018) *DeepMind Ethics & Society*. In: *DeepMind*. <https://deepmind.com/applied/deepmind-ethics-society/>. (Accessed 15 Feb 2018)
- Domingos P (2015) *The master algorithm: how the quest for the ultimate learning machine will remake our world*. Basic Books, New York
- Eisenhardt KM (1989) Building theories from case study research. *Acad Manag Rev* Briarcliff Manor 14:532–550
- Frey CB, Osborne MA (2017) The future of employment: How susceptible are jobs to computerisation? *Technol Forecast Soc Change* 114:254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Galanakis K (2006) Innovation process. Make sense using systems thinking. *Technovation* 26:1222–1232. <https://doi.org/10.1016/j.technovation.2005.07.002>
- Ganesh MI (2017) Entanglement: machine learning and human ethics in driver-less car crashes. *Mach Res* 6:76–87
- Ganguli S, Roy SK (2011) Generic technology-based service quality dimensions in banking: Impact on customer satisfaction and loyalty. *Int J Bank Mark* 29:168–189. <https://doi.org/10.1108/02652321111107648>
- Gatys LA, Ecker AS, Bethge M (2015) A Neural Algorithm of Artistic Style. *Arxiv Preprint* 4:1508.06576. <https://doi.org/10.1167/16.12.326>
- Grunwald A (2000) *Technology policy between long-term planning requirements and short-ranged acceptance problems* New challenges for technology assessment. *Vision assessment: Shaping technology in 21st century society*. Springer, London, pp 99–147
- Haddawy P, Suebnukarn S (2010) Special Topic – Editorial intelligent training intelligent clinical training systems. *Methods Inform Med* 49(04):388–389
- Heess N, Tb D, Sriram S et al (2017) Emergence of locomotion behaviours in rich environments. *Arxiv* 1:70702286
- Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. *Mach Learn* 42:177–196
- Jordan MI, Mitchell TM (2015) *Machine learning: Trends, perspectives, and prospects*. *Science* 349:255–260. <https://doi.org/10.1126/science.aaa8415>
- Klamer T, Allouch SB (2010) Acceptance and use of a social robot by elderly users in a domestic environment. 2010 4th International Conference on Pervasive Computing Technologies for Healthcare. IEEE, New Jersey, pp 1–8
- Klosge W, Zytow JM (1994) *Machine Discovery Terminology*. KDD Workshop 463
- Knight W (2015) A Robotic Toddler Uses Deep Learning to Walk. In: *MIT Technol. Rev.* <https://www.technologyreview.com/s/542921/robot-toddler-learns-to-stand-by-imagining-how-to-do-it/>. (Accessed 13 Sep 2017)
- Krenn M, Malik M, Fickler R et al (2016) Automated search for new quantum experiments. *Phys Rev Lett* 116:090405. <https://doi.org/10.1103/PhysRevLett.116.090405>
- Lanfranchi V (2017) *Machine Learning and Social Media in Crisis Management: Agility vs. Ethics*. IMT Mines Albi-Carmaux, École Mines-Télécom
- Le QV (2013) Building high-level features using large scale unsupervised learning. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. pp 8595–8598
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Lee EA (2008) *Cyber physical systems: design challenges*. IEEE, New Jersey, pp 363–369
- Leon N, Cueva J, Villarreal C et al (2007) Automatic shape variations for optimization and innovation. In: León-Rovira N (ed) *Trends in Computer Aided Innovation*. Springer, US, Boston, MA, pp 179–188
- Lin P, Abney K, Bekey GA (eds) (2012) *Robot ethics: the ethical and social implications of robotics*. MIT Press, Cambridge, Mass
- Linstone HA (2011) Three eras of technology foresight. *Technovation* 31:69–76. <https://doi.org/10.1016/j.technovation.2010.10.001>
- Michalski RS, Carbonell JG, Mitchell TM (1983) *Machine learning: an artificial intelligence approach*. Springer, Berlin
- Muller MJ, Kuhn S (1993) Participatory design. *Commun ACM* 36:24–28. <https://doi.org/10.1145/153571.255960>
- Norman DA, Verganti R (2014) Incremental and radical innovation: design research vs Technology and Meaning Change. *Des Issues* 30:78–96. https://doi.org/10.1162/DESI_a_00250
- Nørskov M (2017) *Social robots: boundaries, potential, challenges*. Taylor & Francis
- OpenAI (2017) *OpenAI Dota 2 1v1 bot*. In: *OpenAI*. <https://openai.com/the-international/>. (Accessed 7 Dec 2018)
- Orlikowski WJ (2007) Sociomaterial practices: exploring technology at work. *Organ Stud* 28:1435–1448. <https://doi.org/10.1177/0170840607081138>
- Quick D (2014) *Kulitta: A Framework for Automated Music Composition*. Yale University, Yale
- Quick D (2015) *Composing with Kulitta*. *Int Comput Music Conf Proc* 1:306–309
- Rasmussen AN, Muratore JF, Heindel TA (1990) The INCO expert system project: CLIPS in shuttle mission control. *First CLIPS Conf* 1:305

- Reader SM, Morand-Ferron J, Flynn E (2016) Animal and human innovation: novel problems and novel solutions. *Philos Trans R Soc B Biol Sci* 371:20150182. <https://doi.org/10.1098/rstb.2015.0182>
- Reich RB (1992) *The work of nations: preparing ourselves for 21st century capitalism*. 1st Vintage, Books. Vintage Books, New York
- Silver D, Schrittwieser J, Simonyan K et al (2017) Mastering the game of Go without human knowledge. *Nature* 550:354–359. <https://doi.org/10.1038/nature24270>
- Smith RE, El-Fallah A, Ravichandran B et al (2004) *The Fighter Aircraft LCS: A Real-World, Machine Innovation Application. Applications of Learning Classifier Systems*. Springer, Berlin, Heidelberg, pp 113–142
- Stollenga MF, Masci J, Gomez F, Schmidhuber J (2014) Deep networks with internal selective attention through feedback connections. *Adv Neural Inform Proces Syst* 4:3545–3553
- Sycara KP (1993) Machine learning for intelligent support of conflict resolution. *Decis Support Syst* 10:121–136. [https://doi.org/10.1016/0167-9236\(93\)90034-Z](https://doi.org/10.1016/0167-9236(93)90034-Z)
- Thekkilakattil A, Dodig-Crnkovic G (2015) Ethics Aspects of Embedded and Cyber-Physical Systems. In: 2015 IEEE 39th Annual Computer Software and Applications Conference. pp 39–44
- Valdes-Perez RE, Simon HA, Zytkow JM (1993) Scientific model-building as search in matrix spaces. *AAAI* 7:472–478
- Vasilescu DC, Filzmoser M (2020) Organizational and societal implications of the adoption of machine invention systems. In: Göcke B, Rosenthal-von Pütten A (eds) *Artificial intelligence: reflections in philosophy theology and the social sciences*. Mentis, Leiden, Netherlands, pp 377–392
- Vega G, Brennan L (2000) Isolation and technology: the human disconnect. *J Organ Change Manag* 13:468–481
- Yin RK (1981) The case study crisis: some answers. *Adm Sci Q* 1:58–65
- Yin RK (2013) Validity and generalization in future case study evaluations. *Evaluation* 19:321–332. <https://doi.org/10.1177/1356389013497081>
- Żytkow JM (1993) Introduction: cognitive autonomy in machine discovery. *Mach Learn* 12:7–16. <https://doi.org/10.1007/BF00993058>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.