Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 180 (2021) 180–189

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing

Towards digital cognitive clones for the decision-makers: adversarial training experiments

Mariia Golovianko [a], Svitlana Gryshko [b], Vagan Terziyan [c], Tuure Tuunanen [d]

[a]*Department of Artificial Intelligence, Kharkiv National University of Radioelectronics, Kharkiv, 61166, Ukraine*
[b]*Department of Economic Cybernetics, Kharkiv National University of Radioelectronics, Kharkiv, 61166, Ukraine*
[c,d]*Faculty of Information Technology, University of Jyväskylä, Jyväskylä, 40100, Finland*

Abstract

There can be many reasons for anyone to make a digital copy (clone) of own decision-making behavior. This enables virtual presence of a professional decision-maker simultaneously in many places and processes of Industry 4.0. Such clone can be used as one's responsible representative when the human is not available. Pi-Mind ("Patented Intelligence") is a technology, which enables "cloning" cognitive skills of humans using adversarial machine learning. In this paper, we present a cyber-physical environment as an adversarial learning ecosystem for cloning image classification skills. The physical component of the environment is provided by the logistic laboratory with camera-surveillance over the conveyors. The digital component of the environment contains special modifications of Generative Adversarial Networks, which include a human-operator as a trainer, an autonomous Pi-Mind clone as a trainee (a discriminator) and a smart digital adversary as a challenger (generator of sophisticated decision situations, emergencies and attacks, which supposedly catalyzes the cloning process).

Keywords: digital cloning, machine learning, Generative Adversarial Networks, cybersecurity, logistics

1. Introduction

Digital Twin is usually understood as a digital replica of a living or a nonliving entity that enables seamless data circulation between physical and virtual worlds [1]. This kind of interaction between worlds is traditionally controlled remotely from some application or a digital space. The idea of making digital twins autonomous, agent-based, self-interested and self-managed entities emerges as a Smart Resource concept [2]. Such autonomous digital replicas of various objects may communicate, negotiate and collaborate with each other enabling digitalization of complex self-managed processes supported by the specific middleware [3].

10.1016/j.procs.2021.01.155

An appropriate, the so-called Pi-Mind ("Pi" stays for "Patented Intelligence") technology is suggested to digitalize ("clone") human decision capability for Industry 4.0 processes by means of autonomic intelligence, semantic modeling and deep learning [4]. Two different cloning approaches are used: explicit transfer of knowledge, i.e., preferences and values used for personal decision-making, from human to an autonomous agent (top-down AI), and twinning by machine learning (bottom-up AI).

Experiments on top-down explicit cloning have been performed at the TRUST-Portal (http://portal.dovira.eu), an academic digital ecosystem for humans and their digital clones [5]. Later the concept evolved towards the idea of the Collective Intelligence as the combination of both human and machine decision-making using the principles of individual and group cloning [6]. Personal clones are supposed to co-evolve synchronously with the correspondent humans and continuously learn to compromise between individual and collective choices.

However, the straightforward top-down approach fails when the decision space (a set of possible decisions) is unknown in advance or when the human cannot explicitly formalize the way to make a certain choice. In this case clones are to be pre-trained to make decisions in previously unknown or critical situations and respond proactively to new threats.

Our research questions are the following: what kind of machine learning (ML) techniques and what kind of ML training environments would enable capturing personal decision skills (i.e., formalized preferences, values, biases and intuition) in order to apply these skills for autonomous decision-making in complex or adversarial decision situations.

In this paper, we test a cyber-physical adversarial learning environment for cognitive cloning of critical decision-making for Industry 4.0. The environment is based on the physical infrastructure of an industrial logistics laboratory and neural network architectures from the family of Generative Adversarial Networks (GANs) representing the adversarial ML paradigm [7]. The modified GAN architecture is applied to generate continuously evolving situations aiming to confuse the clone that learns to find decisions correctly with respect to the response from the respective human. In order to catalyze the training performance various adversarial contexts are simulated in the real laboratory environment [8].

The rest of the paper is organized as follows: in Section 2 we discuss on how adversarial AI could be useful for cognitive cloning; in Section 3 we describe the used research methodology; in Section 4 we report on related work; in Section 5, we suggest an innovative GAN architecture for cloning; Section 6 describes our solution design and our pilot studies within adversarial learning lab and we conclude in Section 7.

2. Why adversarial learning?

Our working and living environments are becoming more and more sophisticated and so are decisions made both by humans and artificial systems. Due to fast learning capabilities and high performance, artificial systems become reliable partners and/or human representatives in digitally enabled decision processes. Except when successfully attacked by an adversary or finding themselves in new, earlier unobservable situations. From this perspective, in order to stay reliable a system should be trained not only to perform its basically designed functions, but also to protect itself from previously unknown threats, as well as to act like the correspondent human in unknown situations.

Our approach is based on artificial generation of new challenging conditions (i.e., training content) to train the system. Training data prepared beforehand by human is obviously not enough for predictions of how the decision situation would evolve, that`s why we rely on artificial generators, particularly GANs. The idea behind GANs is that an intelligent system learns from confrontation with a strong, constantly evolving, hardly predictable artificial adversary. This approach is similar to those used to stimulate the acquired immunity in the human body.

3. Research methodology

Our study applies design science research methodology (DSRM) [9] to develop a specific adversarial digital training environment as a design artifact and a supporting information technology (IT) based on recent AI advances for training autonomous agents ("cognitive clones") that twin human's business-as-usual and critical decision-making skills. Clones are supposed to act as responsible digital substitutes of their (temporarily unavailable, disabled or inactive) human donors in critical and emergent situations.

Human-clone knowledge transfer described in this paper is based on the continuous adversarial learning (according to the before mentioned bottom-up approach) tested in a real cyber-physical laboratory environment within the NATO SPS project "Cyber-Defence for Intelligent Systems" (http://recode.bg/natog5511). The requirement of large well-formed validated datasets for comprehensive trainings and hardly formalized decision processes forced us to use the capacities of real logistic systems, producing and processing big data on industrial business-as-usual and critical decision-making. In this project we aim at proving the concept that Pi-Mind clones can be trained to enhance civil and military security infrastructures and take over the control on certain operations acting on behalf of security officers. A set of experiments are launched in which the artificial workers "observe" the interroll cassette conveyer, an analog of those used in airports for luggage distribution and inspection. Their task is to prevent any potential danger caused by the loads of the cassettes on the conveyer applying personal judgement and expertise of the "cloned" security experts. Two types of decision processes based on the image recognition are launched: in the business-as-usual environment and in the conditions of adversarial attacks.

Besides specific deep learning related metrics used to evaluate the performance of each generative model during training, such as the Fréchet Inception Distance (FID) [10], we use other metrics to evaluate the artificial workers` quality after they are already trained to make decisions. First the prediction accuracy of the classification models is calculated as the percentage of the number of correct predictions (decisions) out of the total number of predictions. This evaluation represents how good a model is, but it is not enough to decide on the appropriateness of an artificial worker as a replacement for the human. With this aim we also find important to measure (i) the correlation between an artificial worker and a human expert indicating the percentage of the matching decisions, (ii) the actual correctness of the decisions by human experts calculated as the classification accuracy and (iii) the correctness of the human experts after he/she knows the decision made by an artificial worker.

4. Related work

Various proactive security approaches use modeling and simulation environments to address previously unknown potential emergencies or adversarial attacks on critical infrastructure, which include also human modelling frameworks. For example, human behavior modeling and simulation of complex scenarios can be done for improvement of commander decision-making processes [11]. Agent driven simulations enable building autonomous models of people and groups based on multiple parameters: ethnical, religion, social, educational, political, age, gender, health, etc.

Human reliability is a crucial element in ensuring infrastructure performance in emergencies. Petrillo et al. study various aspects of human behavior that can influence operator reliability by simulating human errors in emergency condition and argue that the human factor is often a reason for various accidents [12]. Their multi-criteria technique for analyzing complex decisions is based on an integration of fuzzy cognitive maps and the analytic hierarchy process. Such approach allows considering synchronously multiple complementary, contradictory and competitive factors during simulations and identifying and assessing the most frequently occurring and the most critical human factors in emergency management, which is essential for the timely provision of the appropriate safety and policy measures.

Longo [13] proposes an advanced architecture to discover security gaps and to enhance protection capability against threats in marine ports (specifically container terminal). An appropriate simulation environment is used for investigation of how container inspection operations can be profitably integrated with other terminal processes. The aim is to reduce delays and resource consumption. A special risk management platform generates virtual risks and threats scenarios (e.g., inclusion of radiological materials, weapons, etc. within containers) that can affect container terminal operations. These are associated with the security procedures trying to detect and counter risks and threats within the container terminal. The aim is to reengineer security policies, procedures and infrastructures. Padovano et al. [14] enhance simulation environments up to smart factories using digital twins as autonomous service providers. Such service-oriented digital twins are deployed within a real factory-based test environment and demonstrate essential improvements on the quality and productivity of corresponding processes.

In this paper, we use cognitive clones as models of human security operators within our experimental adversarial simulations related to the control of the logistics conveyor (container content inspection) and reported in Section 6. Our study applies adversarial learning as a catalyzer of the cognitive cloning process. Traditionally adversarial

learning is used as a tool for proactive protection of the AI systems against various vulnerabilities and sophisticated attacks. The vulnerability of machine learning algorithms to adversarial samples is widely recognized. Such samples can lead to the wrong behavior of the learned models while the threat is well hidden from humans. Ren et al. [15] make a good summary of the theoretical foundations, algorithms, and applications of adversarial attack techniques and their potential impact in physical-world scenarios. They also admit the absence of a defense mechanism that achieves both efficiency and effectiveness against adversarial samples. Various solutions and architectures related to adversarial learning are not only focusing on a particular domain; there are also domain-transferable representations [16] aiming to reduce domain shift.

A case of adversarial attacks, which attracts huge attention and creates big concerns nowadays, is known as deepfakes. Powered by the latest AI technologies, deepfake generators create such content (pictures, audio, and video) that is almost impossible for human observers to classify correctly. The destructive impact from the deepfake is huge. A good overview and taxonomy of different deepfake types and associated risks is presented in [17].

For safety- or security-critical domains (self-driving cars or malware detection), the quality and predictability of the automated decision-making is extremely important, especially for the so-called corner case inputs. Manually labeled data fails to provide needed content for training. The DeepXplore framework [18] has been elaborated for systematically testing real-world systems based on deep learning. This framework uses the maximization of differential behaviors and neuron coverage and efficiently finds thousands of incorrect corner case behaviors, which can be used to retrain the corresponding decision model to improve its accuracy.

Zhang et al. [16] apply GANs to deliver driving scene-based test generation with various weather conditions. They sample images from public datasets and YouTube videos (snowy or rainy scenes) and use them for training together with transformed images as generated tests. Terziyan and Nikulin [19] suggest generating adversarial examples, which are located deeply within the largest voids (ignorance or confusion zones of the decision space), to facilitate learning. More cases and approaches on the adversarial examples' generation techniques can be found in the review [20].

However, adversarial attacks and deepfakes may play also a positive role for the objectives of cloning decision-making skills. When clones are being trained to copy a human decision behavior, it is difficult to predict all potential and sophisticated decision situations. Specially trained adversarial generators (producing decision situations as adversarial samples or deepfakes) can be useful to facilitate training by helping the clone to capture the corner cases and the boundaries for the decision space of a particular decision-maker.

5. GAN architecture for cognitive cloning

The traditional GAN [7] and its major modifications [21, 22, 23] is a machine-learning infrastructure for a kind of a game between Discriminator, which is an AI entity driven by a neural network (a potential decision maker), and Generator, which is a neural network driven artificial adversary (a challenger for the Discriminator). Both Discriminator and Generator are learning during an adversarial game between them. Discriminator learns the boundaries of the reality (distribution of the reality samples in mathematical terms, or future decision space). Generator challenges the Discriminator with generated fake samples of the reality and learns how to make fake samples to look realistic enough to fool the Discriminator (i.e., Generator continuously discovers the gaps within the Discriminator's current learning outcomes). Both players are getting feedback (loss function value as a measure for the punishment) after each iteration. It is used to improve the parameters of the correspondent neural networks. Discriminator is punished if it fails to distinguish between the real or the fake inputs (either when being fooled by Generator or when it takes real inputs for fake ones). Generator is punished if Discriminator uncovers the generated fakes. This game is a continuous process of Discriminator`s and Generator`s co-evolution towards the maximal achievable performance in the conflicting skills they learn.

Since our task is cognitive cloning, we need an extra player in the GAN architecture as well as some changes in the game logic. The extra player is Trainer, i.e., a human donor (or any system which provides some target cognitive skill to be copied by the clone). Trainer becomes an integral part of a new type of Discriminator – TURING Discriminator (TD), and the whole new family of corresponding TD-enabled GAN architectures named accordingly as TURING-GAN (T-GAN). TD is a Trainer-Trainee (or donor-clone) pair (see Figure 1), where Trainee (an

artificial neural network driven potential clone) learns to copy some target decision-making skill from the Trainer (choosing appropriate class or action label for some input).

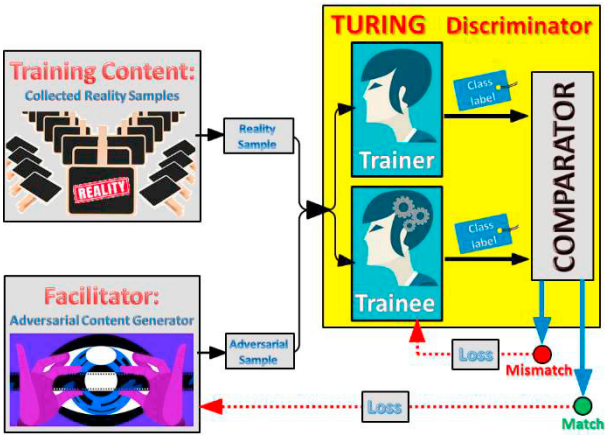

Fig.1. Basic architecture of TURING-GAN used for training the clone (Trainee) to classify the same way as its human donor (Trainer)

The rules of the new game are the following. Both Trainer and Trainee are addressing the same inputs independently. The task of Trainee is to guess the outputs from Trainer, i.e., to minimize the difference in Trainer-Trainee cognitive behavior in same decision-making situations. The difference (mismatch) between the Trainer and Trainee outcomes is measured by Comparator and provided as a feedback (the loss function value) for Trainee. The task of Generator in T-GAN architectures is exactly the opposite. Generator learns to generate such adversarial examples (i.e., sophisticated corner cases), which are differently interpreted by Trainer and Trainee. The closer are the outcomes from Trainer and Trainee the more punishment (the loss function value) Generator gets as a feedback. As a result, Generator learns how to generate the best adversarial training content for Trainee and, therefore, facilitates (catalyzes) the training (cloning) process dramatically. Consequently, Trainee iteratively improves its imitation performance up to becoming a matching clone for the targeted Trainer.

6. Bottom-up AI training for the artificial decision-makers

Experiments on the utility, quality, and efficacy of the bottom-up development of Pi-Mind clones have been implemented within the beforementioned ongoing security project. The idea behind the project is to enhance civil and military security infrastructures with the new proactive intelligent components capable of autonomous decision-making in adversarial situations. Such components take over the control on certain operations and act as artificial security officers working together with human decision makers. Increasingly sophisticated security systems and advanced cyber-attacks require not only good reactive behavior in well-known situations (business-as-usual) but also the preparedness to address new unforeseen threats autonomously (critical decision-making). That is where artificial intelligence can outpace human intellectual abilities. Continuous adversarial learning in the training environment of deep university [24] can prepare trainees (Pi-Mind clones) for that.

In order to prove the concept of artificial decision-makers' efficiency in cyber-physical environments of Industry 4.0, we use capacities of a real logistic system using an interroll cassette conveyor. The conveyor is used for the simulation of an airport luggage system and for the automated inspection at the security checkpoint. The sensors and actuators of the conveyer are interconnected via the respective programmable logic controllers under supervision of a Supervisory Control and Data Acquisition System. Decisions about the distribution of the cassette loads ("bags") are made by artificial intelligent components – Pi-Mind security officers ("airport workers") based on the load's safety aiming to prevent any potentially danger caused by the items in the load. The correspondent decision depends not only on the image recognition quality but also on the personal judgement and intuition of the security officers.

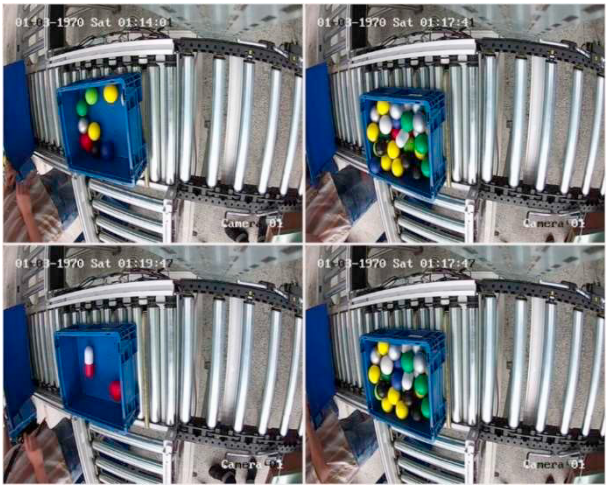

Fig. 2. Image samples representing various configurations of cassettes from Conveyer-V2 Data Set

To train Pi-Mind clones to distribute and dispatch the cassettes, an airport inspection procedure held by three different human experts is simulated. The Conveyer-v2 dataset created for training contains 2198 images of cassettes taken by the cameras installed in the critical distribution points of the conveyer. The images present various configurations of the cassettes and items in them (see Figure 2). The experts label each image either as “dangerous” if it is suspected to contain some potentially unsecure object(s) or as “not dangerous” otherwise. For the first experiments we imitate non-hazardous items with plastic balls and pills of different colors and hazardous items with red hearts which are considered to be a “bomb”.

6.1. Case 1: Airport luggage inspection (business-as-usual)

Three classifiers based on a deep convolutional neural architecture mimic decision behavior of human inspectors. The transfer learning technique is used for the optimization purposes. The image feature extraction module uses the Inception V3 [25] architecture trained on ImageNet (image-net.org). The accuracy of the trained artificial security officers is evaluated with respect to four parameters: (1) the actual correctness of the classification obtained on the test and validation sets (see Table 1); (2) correlation between the artificial inspectors’ and the human experts’ decisions (see Table 2); (3) the actual correctness of the decisions from human experts (see Table 3); (4) the correctness of the human experts in case of artificial decision advice (see Table 4).

Table 1. The correctness of the classification of artificial security inspectors in business-as-usual processes

Artificial decision-maker	Test set	Validation set
Classifier_1	92.05%	87.5%
Classifier_2	97.95%	96.5%
Classifier_3	95.23%	94%

Table 2. The correlation between artificial and human security inspectors in business-as-usual processes (Turing test)

Decision-makers	Test set	Validation set
Classifier_1 vs DM_1	92.05%	92.1%
Classifier_2 vs DM_2	99.53%	99.48%
Classifier_3 vs DM_3	95.66%	94.94%

Table 3. The actual correctness of the human experts in business-as-usual processes

Human decision-makers	Test set	Validation set
DM_1	95.45%	95%
DM_2	98.41%	97%
DM_3	99.55%	99%

Table 4. The correctness of the human experts provided an advice from an artificial inspector

Human decision-makers	Test set	Validation set
DM_1	97.27%	95%
DM_2	98.86%	98%
DM_3	99.77%	97.5%

Although human decision-makers show better performance in threat recognition, the results are promising due to the high accuracy of the artificial predictions and the high human-clone correlation of the decisions. Moreover, artificial experts appear to be capable of valuable advice since the human donors improve their accuracy after getting an advice from the artificial inspectors.

6.2. Case 2: Airport luggage inspection under adversarial attacks (critical decision-making)

An artificial security officer is trained to be prepared to new tasks and complex contexts. A disrupted reality and unexpected confusing conditions are simulated by poisoned images coming from the camera so that they are expected to be misclassified by the artificial predictors. Two attack scenarios are considered:

- Attack Scenario 1: there is “a bomb in the bag” but the image is “poisoned” to be potentially misclassified as being “not dangerous”. The intent of this attack is to allow dangerous load coming undetected “on board”.

- Attack Scenario 2: there is “no bomb in the bag”, but the image is “poisoned” to be potentially misclassified as being “dangerous” causing a false alarm. The intent of this scenario is to cause disruption in the normal operation of the system, causing delays and panic.

Applying different types of adversarial noise during the so-called white-box attacks we are able to compromise the work of the deep-learning convolutional classifiers completely and decrease their accuracy to 0%. Experiments show that both artificial and human workers tend to misclassify poisoned images (see Table 5).

Table 5. The accuracy of the tampered images recognition by Pi-Mind clones vs human experts

Human decision-makers	Pi-Mind clone
75%	70.05%
90%	71%
85%	65.5%

To foster resilience of decision-making in the logistic system under cyber-attacks, the clone is trained to develop a new capability – an artificial cognitive immunity against potential adversarial attacks.

Experiment A. Training a powerful discriminator to recognize tampered images based on StyleGAN2 architecture.

Using data-driven unconditional generative image modeling based on StyleGAN2 architecture [26] configured as shown in Table 6, we pursue two objectives: to train a powerful discriminator capable of detecting a tampered image; and to generate a pool of new high quality images, which are deliberately designed as adversarial samples aka “digital vaccine” (see Figure 3) for smart vaccination during comprehensive retraining of the Pi-Mind clones.

Table 6. The parameters of the StyleGAN2 transfer learning with Conveyer-v2 dataset

Parameter	Value
Configuration	config-a
Resolution	1024x1024
Total number of images	25000
GPU	2
Duration	45d23h
GPU Memory	7.5 Ghz

From 3000 generated images a batch of 300 samples, the most challenging ones for both human and artificial security officers, is selected as the samples of the disrupted reality used to retrain Pi-Mind clones (see Fig.3).

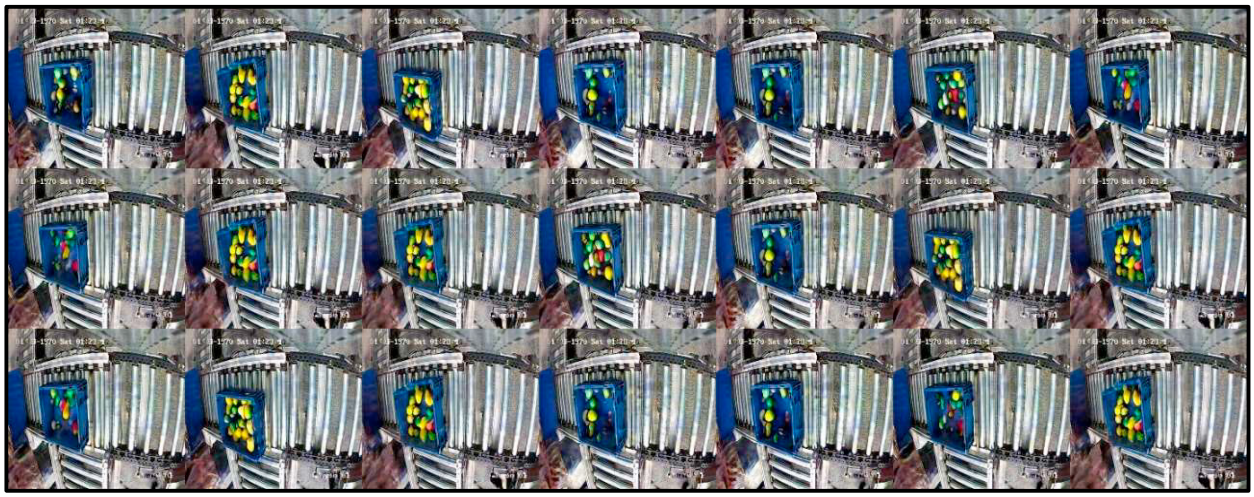

Fig. 3. A pool of artificially generated images used for smart vaccination

The experiments show an increase of the discrimination accuracy after retraining/vaccination (see Table 7).

Table 7. The accuracy of the tampered image recognition by Pi-Mind clones after retraining vs human experts

Human decision-makers	Pi-Mind clone
72%	80.05%
92%	75%
80%	70%

By using unconditional generative models, we can increase the detection of poisoned images, but we can't ensure the correctness of their classification.

Experiment B. Training a powerful discriminator to recognize adversarial images with dangerous items.

Experiments with different conditional GAN architectures are focused on the creation of the model capable of accurate generation of images with specific semantic attributes, such as class "dangerous/not dangerous" membership. The trained generator is used as an adversary within the learning model of Pi-Mind clones and, therefore, as a facilitator of their further retraining.

Experiments with several conditional GANs architectures [27, 28] show that the quality and the resolution (256x256 maximum) of the generated images are not sufficient for real-world application: the generated images are easily classified as tampered by human decision-makers and cannot be used for Pi-Mind clones retraining.

7. Conclusions

We present an adversarial learning environment, which facilitates the process of making digital cognitive clones of decision-makers. Several experiments within the environment are aimed to check the performance of digital clones both for business-as-usual decision situations and for critical decision-making. We experiment on image classification skills' cloning process complicated by adversarial machine learning attacks on the images. Although human decision-makers still show better performance in threat recognition, the results of the experiments are promising due to the high accuracy of the artificial predictions and the high level of the human-clone correlation with respect to the decisions. Moreover, experiments with different types of adversarial attacks show that both artificial and human workers tend to misclassify threatening objects in adversarial settings but artificial agents can be trained to develop a new capability – an artificial cognitive immunity, and help improving the accuracy of the human predictions by giving artificial advice.

We use the design science research methodology [9] regarding the Pi-Mind adversarial learning environment as a design artifact and we experimentally demonstrate its efficacy and efficiency for training digital cognitive clones of the decision makers restricted by one cognitive skill, i.e., image classification in normal and adversarial conditions.

Invention of several new GAN architectures better fitting the concept of cloning-as-an-immunity-training is a contribution to the theory of neural networks, particularly, GANs. The first experiments with the new architectures demonstrate that they are capable to generate a certain type of evolving adversarial and challenging decision situations (aka dynamic vaccine), which help the clone to learn how to make decisions in adversarial conditions. Our experiments have been performed on the premises of a real laboratory with real yet simple objects. However, the results can prove the initial concept and the new architectures can be recommended as a basic cognitive cloning technique as well as a protection mechanism to a variety of critical systems.

Due to the impact of the use of AI and automation within the industrial processes, one may think that the supervisory role of humans in Industry 4.0 is decreasing. However, we consider cloning as a means to preserve the leading role of humans (via their digital twins) in different industrial processes (in addition to the use-case scenarios presented in [4]) from manufacturing to customer experience management for, e.g.:

- Manufacturing process automation: digital cognitive clones of real workers, i.e. operators and decision makers, preserve the human-centric nature of manufacturing processes in industry 4.0 and are able to address challenges, such as, e.g., highly personalized products' manufacturing, future-proofing manufacturing automation scenarios, etc.

- Industrial product exploitation: use of self-driving cars is seriously influenced by the ethical dilemma related to the decision-making of the autonomous driver in critical situations (hard choices). In our opinion, giving the autonomous control over the vehicle to a trained AI agent (a cognitive clone) that makes choices similar to the donor would increase the customer`s trust and responsibility for the autonomous decision-making.

- Customer experience, supply chain and asset management: simulations with the self-managed "digital customers" and digitalized processes can be used to obtain the optimal settings in real-world environments and improving customer`s involvement into the design and manufacturing processes. It would allow collective intelligence (integrated digital customers and humans) interacting with real services and products via their digital twins in cyber-physical environments.

References

[1] Saddik A.E., 2018. Digital Twins: The Convergence of Multimedia Technologies. IEEE MultiMedia, 25 (2), 87–92.

[2] Terziyan V., 2008. SmartResource – Proactive Self-Maintained Resources in Semantic Web: Lessons learned. International Journal of Smart Home, 2(2), 33-57.

[3] Katasonov A., Terziyan V., 2008. Semantic Agent Programming Language (S-APL): A Middleware Platform for the Semantic Web. Proceedings of the 2nd IEEE Intern. Conf. on Semantic Computing, pp. 504-511. August 4-7, Santa Clara (CA, USA).

[4] Terziyan V., Gryshko S., Golovianko M., 2018. Patented Intelligence: Cloning Human Decision Models for Industry 4.0. Journal of Manufacturing Systems, 48 (Part C), 204-217.

[5] Terziyan V., Golovianko M., Shevchenko O., 2015. Semantic Portal as a Tool for Structural Reform of the Ukrainian Educational System. Information Technology for Development, 21(3), 381-402.

[6] Gavriushenko M., Kaikova O., Terziyan V., 2020. Bridging Human and Machine Learning for the Needs of Collective Intelligence Development. Procedia Manufacturing, 42, 302-306.

[7] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014. Generative Adversarial Nets. In: Z. Ghahramani et al., eds. Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2672–2680.

[8] Terziyan V., Golovianko M., Gryshko S., 2018. Industry 4.0 Intelligence under Attack: From Cognitive Hack to Data Poisoning. In: K. Dimitrov, ed. Cyber Defence in Industry 4.0 Systems and Related Logistics and IT Infrastructure (Information and Communication Security, Vol. 51). Amsterdam, Netherlands: IOS Press, 110-125.

[9] Peffers K., Tuunanen T., Rothenberger M.A., Chatterjee S., 2007. A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), 45-77.

[10] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Advances in neural information processing systems (pp. 6626-6637).

[11] Bruzzone A., Massei M., Longo F., Poggi S., Agresta M., Bartolucci C., Nicoletti L., 2014. Human Behavior Simulation for Complex Scenarios based on Intelligent Agents. Proceedings of the 2014 Annual Simulation Symposium, pp. 1-10. Tampa (USA).

[12] Petrillo A., Felice F.D., Longo F., Bruzzone A., 2017. Factors Affecting the Human Error: Representations of Mental Models for Emergency Management. International Journal of Simulation and Process Modelling, 12(3-4), 287-299.

[13] Longo F., 2012. Supply Chain Security: an Integrated Framework for Container Terminal Facilities. International Journal of Simulation and Process Modelling, 7(3), 159-167.

[14] Padovano, A., Longo, F., Nicoletti, L., Mirabelli, G., 2018. A Digital Twin Based Service Oriented Application for a 4.0 Knowledge Navigation in the Smart Factory. IFAC-PapersOnLine, 51(11), 631-636.

[15] Ren K., Zheng T., Qin Z., Liu X., 2020. Adversarial Attacks and Defenses in Deep Learning. Engineering, 6(3), 346-360.

[16] Zhang M., Zhang Y., Zhang L., Liu C., Khurshid S., 2018. DeepRoad: GAN-Based Metamorphic Testing and Input Validation Framework for Autonomous Driving Systems. In: M. Huchard et al., eds. Proceedings of the 33rd ACM/IEEE Intern. Conf. on Automated Software Engineering. NY, USA: ACM, 132-142.

[17] Kietzmann J., Lee L.W., McCarthy I. P., Kietzmann T.C., 2020. Deepfakes: Trick or Treat? Business Horizons, 63(2), 135-146.

[18] Pei K., Cao Y., Yang J., Jana S., 2019. DeepXplore: Automated Whitebox Testing of Deep Learning Systems. Communications of the ACM, 62(11), 137-145.

[19] Terziyan V., Nikulin A., 2019. Ignorance-Aware Approaches and Algorithms for Prototype Selection in Machine Learning. arXiv:1905.06054

[20] Zhang C., Zhao Q., Wang Y., 2020. Hybrid Adversarial Network for Unsupervised Domain Adaptation. Information Sciences, 514, 44-55.

[21] Yinka-Banjo C., Ugot O.A., 2019. A Review of Generative Adversarial Networks and its Application in Cybersecurity. Artificial Intelligence Review, 53, 1721–1736.

[22] Wang Z., She Q., Ward T.E., 2019. Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. arXiv:1906.01529

[23] Gui J., Sun Z., Wen Y., Tao D., Ye J., 2020. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. arXiv:2001.06937

[24] Golovianko M., Gryshko S., Terziyan V., 2018. From Deep Learning to Deep University: Cognitive Development of Intelligent Systems. In: J. Szymański, Y. Velegrakis, eds. Semantic Keyword-Based Search on Structured Data Sources IKC 2017 (Lecture Notes in Computer Science, Vol. 10546). Springer, 80-85.

[25] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z., 2016. Rethinking the Inception Architecture for Computer Vision. Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2818-2826, June 27-30, Las Vegas (NV, USA).

[26] Karras T., Laine S., Aittala M., Hellsten J., Lehtinen J., Aila T., 2019. Analyzing and Improving the Image Quality of Stylegan. arXiv:1912.04958

[27] Odena A., 2016. Semi-Supervised Learning with Generative Adversarial Networks. arXiv: 1606.01583

[28] Odena A., Olah C., Shlens J., 2017. Conditional Image Synthesis with Auxiliary Classifier Gans. Proceedings of the 34th Intern. Conf. on Machine Learning, pp. 2642-2651. Sydney (Australia).