JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Terziyan, Vagan; Gavriushenko, Mariia; Girka, Anastasiia; Gontarenko, Andrii; Kaikova, Olena

Title: Cloning and training collective intelligence with generative adversarial networks

Year: 2021

Version: Published version

Copyright: © 2021 The Authors. IET Collaborative Intelligent Manufacturing published by Joh

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Terziyan, V., Gavriushenko, M., Girka, A., Gontarenko, A., & Kaikova, O. (2021). Cloning and training collective intelligence with generative adversarial networks. IET Collaborative Intelligent Manufacturing, 3(1), 64-74. https://doi.org/10.1049/cim2.12008

Received: 27 April 2020 | Revised: 28 October 2020 | Accepted: 11 November 2020 | IET Collaborative Intelligent Manufacturing

DOI: 10.1049/cim2.12008

ORIGINAL RESEARCH PAPER

Cloning and training collective intelligence with generative adversarial networks

Vagan Terziyan | **Mariia Gavriushenko** | **Anastasiia Girka** | **Andrii Gontarenko** | **Olena Kaikova**

Faculty of Information technology, University of Jyväskylä, Jyväskylä, Finland

Correspondence
Vagan Terziyan, Faculty of Information technology, University of Jyväskylä, Jyväskylä, Finland.
Email: vagan.terziyan@jyu.fi

Abstract

Industry 4.0 and highly automated critical infrastructure can be seen as cyber-physical-social systems controlled by the Collective Intelligence. Such systems are essential for the functioning of the society and economy. On one hand, they have flexible infrastructure of heterogeneous systems and assets. On the other hand, they are social systems, which include collaborating humans and artificial decision makers. Such (human plus machine) resources must be pre-trained to perform their mission with high efficiency. Both human and machine learning approaches must be bridged to enable such training. The importance of these systems requires the anticipation of the potential and previously unknown worst-case scenarios during training. In this paper, we provide an adversarial training framework for the collective intelligence. We show how cognitive capabilities can be copied ("cloned") from humans and trained as a (responsible) collective intelligence. We made some modifications to the Generative Adversarial Networks architectures and adapted them for the cloning and training tasks. We modified the Discriminator component to a so-called "Turing Discriminator", which includes one or several human and artificial discriminators working together. We also discussed the concept of cellular intelligence, where a person can act and collaborate in a group together with their own cognitive clones.

1 | INTRODUCTION

Many areas of the human life are becoming more and more affected by the artificial intelligence (AI). The benefits of the AI in solving the actual applied problems are undeniable. For example, expert systems can help in providing efficient decision-making services based on the formalised explicit human expertise; computational intelligence enables automated learning of implicit expertise hidden within data or experimental observations; autonomous smart devices can help in exploration (directly on the spot) of environments harmful to human's health or life, and so on. AI can even replace analysts and managers. Intuition, experience, and manual labour can no longer cope with processing a large flow of information. That is why businesses are currently optimising their work with the help of various AI tools.

AI is also a driver of a popular digital transformation trend of the modern industry. Current COVID-19 crisis, surprisingly, played the role of a catalyst for the evolution of the AI component in digital transformation. According to [1], the importance of the smart online services and corresponding (new) customer experience is now as high as never before because of worldwide lockdown. This will drive the focus of the future investments to the new technologies.

For the modern AI systems, training is the major need, which humans can address at the current stage of the AI evolution. In the human world, this need is covered by education. We argue that (deep) learning for a machine is a dynamic, evolutionary process, very similar to a traditional higher education, however, with some new challenges and features. It facilitates comprehensive acquisition of different skills at all the major cognitive levels, leveraging on the collaboration in

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Collaborative Intelligent Manufacturing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

creative, dynamically changing ecosystems, similar to those built around the universities. The most powerful weapon in the IT business today is the alliance between the AI, or analytical skills of self-learning machines, and the imaginative human intellect of great leaders. Together they make collective intelligence (CI), which is the major business model of the future [2].

This study is an extended version of an article presented at the International Conference on Industry 4.0 and Smart Manufacturing (ISM 2019) [3]. The main research questions this article addresses are (1) what is the added value of the CI concept if applied to secure digital transformation of various business processes in the industry; (2) how to design digital cognitive clones of a human CI to automate business processes; and (3) what kind of machine learning (ML) architectures could be appropriate for such cloning.

The rest of the article is organised as follows: Section 2 discusses the role of the CI within the digital transformation; Section 3 provides basic approaches and architectures for cloning human intelligence and training the CI; Section 4 describes the use of cloning for cellular CI (human + digital assistants); and the study is concluded in Section 5.

2 | COLLECTIVE INTELLIGENCE AS A DRIVER OF DIGITAL TRANSFORMATION

Development of technologies influences the way companies are doing their businesses [4]. Authors in [5] explain why and how the digital transformation and Industry 4.0 can change numerous business models and organisations.

We believe that AI in general is the main enabler for digital transformation of a variety of processes within the Industry 4.0. However, the role of humans capable of using AI smartly within these processes remains an important success factor. In [6], Vial defined digital transformation as a process that aims to 'improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies'. The definition is not organisation-centric and refers to the broader term 'entity'. We would like to focus on a human (an employee, a specialist, a professional, and also a customer) as an entity (a subject) of digital transformation. Therefore, we study the concept of collective (collaborative) intelligence (CI) as a (human + autonomous AI) collaborative resource in managing complex business processes. It will be a compromise between the bottom-up statistical AI approaches (computational intelligence, deep learning, etc.) and top-down symbolic AI approaches driven by explicit human knowledge and decision-making (including intellectual clones of humans). We need a kind of *responsible artificial intelligence* (RAI) as a compromised (academy–industry) solution framework or such a layer of the CI, which will be trusted and adopted in various business ecosystems as well as preserve a human-centric nature of everyday business processes. The compromise will be the meeting point of the two mutually oriented processes: (1) ongoing AI-like digital transformation of humans and human-centric infrastructure (H2AI) and (2) emerging human-like transformation of the AI itself and AI-driven smart autonomous industrial and business infrastructure (AI2H). RAI must be (a) explainable (XCI: *explainable CI*—the results of the CI solutions can be understood both by human and artificial 'experts') plus (b) operational (OCI: *operational CI*—bridging the gap between the CI research, promises, and expectations, and the reality needs, challenges, and problems). On one hand, RAI must be capable of benefiting from the ML techniques (adversarial, supervised, unsupervised, semi-supervised, reinforcement, deep learning, etc.) for capturing the behaviour, knowledge, and decision-making models of humans (end-users, customers, experts, managers, etc.). On the other hand, based on the autonomous agents' technology, it enables creation of *digital (cognitive) clones* and embedding them into a simulated or real business environment for playing a similar role to their original human-twins. In such cases, humans will share their existing responsibilities and capabilities with their digital clones, and vice versa.

One can see the generic schema of the typical interconnected business processes, digitalisation of which can benefit from the CI as a managing component (see Figure 1).

According to its managing role in the process, CI is a digital innovation that would provide digital transformation of the organisational infrastructure [7].

Assume we have some complex industrial system (system of systems) and the overall goal is to upgrade this system by applying an AI-driven digitalisation meaning smart digital transformation of business processes within the system. We will focus on several major aspects of such digitalisation where the CI is assumed to be the key component:

1. *Smart data collection*. This component may work similarly to an 'autonomic nervous system' of a human. The main issues here are how to identify the need vs. availability of the data and its location; how to assure the quality and integrity of the collected data during the real-time data collection process; how to recognise and proactively neutralise the factors (both internal and external) which negatively affect the quality of the data; how to prepare the data for the potential ML process over it; how to address the issue of data privacy and anonymisation.
2. *Smart data integration and representation*. This component is responsible for structuring and integrating the data together with metadata and available knowledge to enable seamless integration and interoperability of various subsystems, tools, and algorithms working with this data. The main issues here are what would be a suitable ontology as an umbrella on top of many diverse data sources, data types, and formats; how to apply semantic technologies and enable linked data; how to enable smart query engine for various applications (e.g. SPARQL endpoint); how to prepare such semantic data storage infrastructure (semantic 'data lake') to efficiently store the knowledge, which will be discovered by the AI-driven analytics.

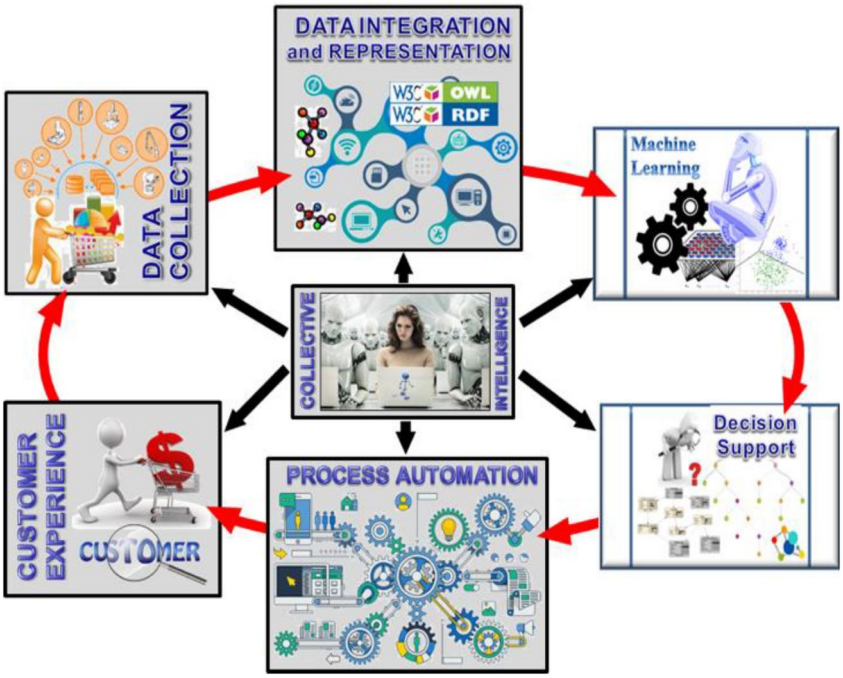

FIGURE 1 Digitalisation of industrial processes managed by collective intelligence

3. *Machine learning*. This component is responsible for discovering implicit knowledge (models) on the basis of data using a variety of AI/ML techniques and for a variety of potential intelligent tasks (control, decision-making, prediction, diagnostics, etc.). Consider, for instance, the sales and demand forecast problem, which has highly significant for many businesses, as the impact of its accuracy can be dramatic. Currently it is a visible trend towards shifting from traditional sales forecast approaches (human-expert-driven survey and statistical methods) to the use of AI/ML-driven predictive analytics, which combines, for example, deep regression analysis with recurrent (LSTM) neural networks. CI enables additional opportunities for smart forecasting: (a) merging predictive analytics with the context discovery; (b) combining the 'black-box' deep learning methods with the top-down explainable AI; (c) utilising the concepts of 'digital customer' and 'digital competitor' to apply autonomous AI and 'cognitive cloning' algorithms for making predictions on the basis of simulation, proactive analytics, adversarial and reinforcement learning approaches.
4. *Smart decision-support*. This component may work similarly to a central nervous system of a human. It will use the trained models to automate the decision-making process on the basis of available information. It will be capable of making a variety of control decisions; evaluating/classifying/recognising various inputs; diagnosing the assets; making predictions regarding the potential issues (e.g. faults or breaks), regarding evolution of external factors, or regarding the behaviour of the customers.
5. *Smart process automation and customer experience using autonomic computing*. This component will benefit from using autonomous and self-managed software agents, proactive digital twins, digital clones, digital assistants, digital advisors, and so on, and will be capable of (partially or completely) automating certain critical business processes. Important aspect of this component would be the essential breakthrough within the customer experience if the 'digital customer' approach is applied.
6. *Collective intelligence platform*. This is the main component and our main objective. It is an enabler of smart digital transformation for the variety of industrial and business processes within a data/knowledge management cycle. It provides autonomous AI support for the processes to enable (like in the Industry 4.0) self-management (self-configuration, self-optimisation, self-protection, self-healing, etc.) of critical industrial systems and assets. The specific of the CI platform is that it finds the best compromise between completely autonomous and human-driven processes by enabling collective/collaborative intelligence. The systems under the CI platform surveillance are expected to be more efficient, robust, fault-tolerant, and resilient.

The CI platform opens the opportunity of creating an innovative type of businesses, which will enable, support, and facilitate selling and buying digital proxy/advisors/assistants/twins/clones, etc.; technologies for their design and training; licenses and patents for their use; digital spaces (platforms) for their execution and coordination support; models for the efficient human–AI collaboration; practical implementation of the collaborative-intelligence-driven business processes for a variety of industries worldwide, and so on. We believe that there is no other than the CI way to proceed because the gap between the challenging and evolving environment and the capabilities of the human processes as such is continuously growing. To keep and improve the quality of industrial

processes and efficiency of human activity within them, one needs the next generation of artificial, autonomous, and smart labour force that must be naturally integrated into the existing processes. Taking into account that the recent status of AI science, solutions and tools make it possible to design such autonomous enhancement (and not at all replacement!) for humans, we must explore this opportunity and make radical changes to the quality of life for all humans.

3 | BRIDGING THE GAP BETWEEN HUMAN AND 'AI' Learning

The need for training the autonomous AI systems in the same way as humans (in addition to traditional ML) was recently discussed in [8]. The authors suggested the never-ending learning paradigm for the ML, according to which the intelligent agents will learn and generalise many types of knowledge, continuously over many years to become better learners over time.

According to the Asilomar Principles [9] signed by the majority of leading AI scientists, the goal of the AI research should be to create beneficial intelligence but not undirected intelligence and, therefore, the AI systems are designed to recursively self-improve or self-replicate under strict human control.

While admiring the computer simulations for the experience, it provides for the human learners, [10] also points out their drawback in a lack of pedagogical ability that appears in the absence of a feedback. They consider the intelligent tutoring system (ITS) as a solution that addresses the pedagogical issues, since it is supposed to provide hints, guidance, and feedback.

The ITS is a long-standing concept, its history begins from the first teaching machines in the mid-1920s [11]. The synergy of the ITS and education came up with different (from ones presented here) approaches including the learning by teaching approach that is based on teachable agents, for example, Betty's Brain [12]. The biggest advantage of ITS is reducing dependencies on human resources.

Nevertheless, all existing researches on the ITS had a focus on how to teach humans with the ITS (pedagogical issues) and how to represent knowledge within the ITS (AI issues). Whereas our research is focused on how to train (teach) a digital learning assistant based on neural network as an autonomous artificial cognitive system within the concept of the University for Everything [13]. Thus, it is a shift from the traditional ITS concept that serves humans to the University for Everything that is able to teach neural networks among others.

It is still an open problem, how to encode the knowledge into the software [10], since it is a significantly resource-consuming task while developing an ITS instance. Our approach with the application of adversarial training of neural networks may contribute to this problem.

3.1 | Collective intelligence in decision-making process in Industry 4.0

In previous research, we introduced briefly the concept of the collaborative intelligence and the University of the Future, as well as the concept of digital clone, which will benefit in creation of digital learning assistants [3]. The concept of 'digital twins or clones' was first introduced in 2003 [14].

The authors of [15] proposed 5C architecture of a cyber–physical system (CPS), that is, connection, conversion, cyber, cognition, and configure, the third layer of which is 'cyber', and the concepts of digital twins and clones belong to it. CPS is a key technological concept of the Industry 4.0 [16]. However, the key role of CI comes from the definition of the Industry 4.0 as a 'trend related to smart factories, which are cyber-physical spaces populated and controlled by the CI for the autonomous and highly flexible manufacturing purposes' [17].

An extensive literature analysis in [18] revealed that currently digital clones are mostly utilised in terms of smart manufacturing, production equipment maintenance and optimisation, rather than as twins of the product itself, which could be useful during the whole product lifecycle (also after the production). The authors of [19] describe the approach to creating a smart digital clone of a manufacturing process enhanced with AI technologies aimed at integrating the digital twin of the product itself and a twin of the product's development process. In [20], the authors describe the application of digital clones as a service provider in the manufacturing industry. Such an application would help to shift the current 3.0 digitised factory to a 4.0 smart factory.

As spotted in [21], the dynamic Industry 4.0 environments are full of uncertainties, complexities, and ambiguities, and, therefore, they demand faster and more confident decisions. However, as the authors of [21] have noticed, there is still no survey study that would show how to support decision-making in organisations in the context of the Industry 4.0. In this study, we suggest the CI as a powerful decision-making tool to manage complexity and uncertainty within the Industry 4.0 processes.

The authors in [22] consider the Industry 4.0 as a socio-technical system that has an impact on people, infrastructure, technology, processes, culture, and goals. We suggest expanding the social aspect of such an integrated ecosystem also with the smart autonomous AI and particularly with the CI.

This study tries to bridge the gap between human and AI learning addressing human–machine co-working that is inevitable for the knowledge management in Industry 4.0 [23]. Small and medium businesses lack affordable solutions to benefit from the Industry 4.0 technologies. A recent research based on literature review has discussed this problem and proposed research framework, where CI plays the crucial role in the decision-making process [24].

In this study, we focus more on technical description of the architecture for digital cloning utilising Generative Adversarial Networks (GAN) in terms of business processes.

3.2 | Adversarial training of the collective intelligence

To perform its mission in challenging and constantly changing environments, the CI cannot be hardcoded; it must be trained [3]. It would be naïve to assume that we could anticipate all the future challenges that the CI might potentially face and adapt the training process and learning content accordingly. Therefore, to make the training efficient with limited resources, we have to train the CI in an 'aggressive' (adversarial) environment.

Hence, we suggest using adversarial learning as a popular ML technique and, in particular, the concept and architecture of the GAN. During such training, an artificial adversary discovers the learning gaps within the target component skills such as fuzzy unreliable decision boundaries, weak spots or 'grey zones' within the training data [25]. It then attacks the target component accordingly forcing it to learn faster to adapt. Another advantage of adversarial training is that it enables 'cognitive cloning' of humans to design artificial CI teams for various processes, for example, Industry 4.0 [17].

GAN is a kind of game model of the two competing neural networks, a generator and a discriminator. These two components come together in the network and work as adversaries, pushing the performance of one another. Adversarial learning in general and GAN in particular has recently become a popular type of deep learning algorithms producing realistically looking images [26].

Discriminator gets samples from two sources: the real world and the fake generator. It then trains to distinguish between the fake and the real. Assume $\{x^{(1)}, \ldots, x^{(m)}\}$ is a sample minibatch of real samples with probability distribution $p_{data}(x)$. Assume also $\{z^{(1)}, \ldots, z^{(m)}\}$ is a sample minibatch of latent vectors and corresponding fake samples generated by the generator $\{G(z^{(1)}), \ldots, G(z^{(m)})\}$ with probability distribution $p_z(z)$. Generator tries to generate samples from the scratch (latent vector) aiming the same distribution as that of the real samples. It trains to capture the real samples distribution and therefore to fool the discriminator.

Discriminator loss (provided as a feedback for the update of the discriminator) takes into account the own misclassification error and therefore the generation success of the generator as follows in (1):

$$Loss(D) = \frac{1}{m} \cdot \sum_{i=1}^{m} \log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \quad (1)$$

where $D(x^{(i)})$ is discriminator output for the real data sample $x^{(i)}$; $D(G(z^{(i)}))$ is discriminator output for the generated fake data sample $G(z^{(i)})$.

The first term within the sum operand of the loss function above corresponds to the aim of optimising the probability that the real data is rated highly. The second term corresponds to optimising the probability that the generated data is rated poorly.

Generator loss (provided as a feedback for the update of the generator) takes into account the own generation error (uncovered samples) and therefore the discrimination success of the discriminator is as follows in (2):

$$Loss(G) = \frac{1}{m} \cdot \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \quad (2)$$

The term within the sum operand of this loss function corresponds to the aim of optimising the probability that the generated data is rated highly.

Taking into account that we consider the CI as a collaborative and hybrid (human plus machine) intelligence, we have to update the basic GAN concept by finding the place for a 'human' component as well. For that purpose, we suggest the new type of a discriminator—a '*Turing discriminator*', which will be considered as a kind of a 'mixer' for human and machine intelligence.

We name the GAN architecture with such discriminator as T-GAN (see Figure 2). TD has very different semantics comparably to a traditional discriminator. TD is actually a kind of 'CI', which includes at least one 'banner' component considered already trained, e.g. a 'human' (H) and at least one traditional learnable neural discriminator (D). The generator (G), which plays against TD (H+D team) aims to generate such samples that would maximise the difference between H and D's reaction to those samples. Both H and D are differentiating among inputs (i.e. 'real' or 'fake' for simple GAN). The TD outputs the probability distribution between 'match' and 'no match' options of the H and D opinions, which is used as a loss function to train both D and G. The D player within TD tries to learn to synchronise its own opinions on the inputs with the H's opinions. Such schema allows trained D to capture the hidden discriminative logic of the 'ideal' H. Sometimes it is possible to use 'strong and trained artificial classifier' AI–H instead of H, and then D tries to learn to the level of such AI–H.

If to compute a mismatch between H and D outputs regarding some input q as it defined in (3):

$$\Delta HD(q) = |D(q) - H(q)|, \quad (3)$$

then the loss functions for training TD and G will be the ones presented in (4) and (5), respectively:

$$Loss(TD) = \frac{1}{m} \cdot \sum_{i=1}^{m} \left\{ \log \Delta HD\left(x^{(i)}\right) + \log \Delta HD\left(G\left(z^{(i)}\right)\right) \right\}; \quad (4)$$

$$Loss(G) = \frac{1}{m} \cdot \sum_{i=1}^{m} \left\{ \begin{array}{l} \log 1 - \Delta HD\left(x^{(i)}\right) \\ + \log 1 - \Delta HD\left(G\left(z^{(i)}\right)\right) \end{array} \right\} \quad (5)$$

GANs have been modified to enable not only the fake detection capabilities of the discriminators but also generic classification skills. For example, semi-supervised GAN (SGAN) [27] is such an extension of a generic GAN

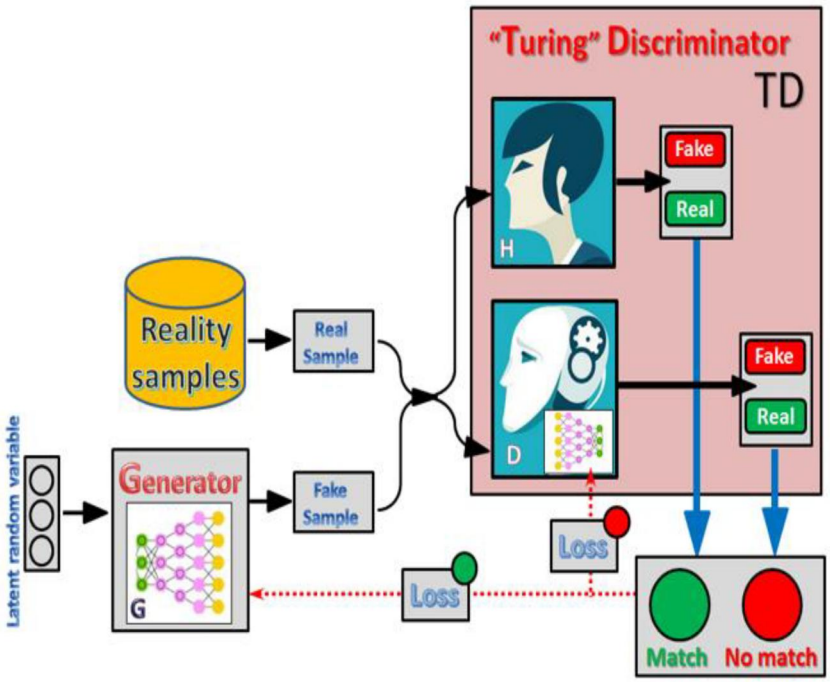

FIGURE 2 T-GAN architecture, where traditional GAN discriminator is replaced with the 'Turing discriminator'

architecture towards a semi-supervised context by forcing the discriminator network to output class labels. Generator and discriminator are trained on a dataset with inputs belonging to one of N classes. Trained discriminator is assumed to predict, which of $N+1$ classes the input belongs to, where an extra 'fake' class is added to correspond to the outputs of the generator. This method appears to be capable of creating more data-efficient classifiers and at the same time it allows generating higher quality samples than a regular GAN. For instance, [28] utilise SGAN as a semi-supervised learning architecture to address such problems as labelled data scarcity and data domain overfitting. For cardiac abnormality classification in chest X-rays, they demonstrated that significantly less data is required with SGANs than with conventional supervised learning convolutional neural networks.

We can enhance the SGAN architecture with the TD the same way we did for the traditional GAN. See Figure 3, where appropriate T-SGAN architecture is shown. This way we can get the learning D component of the TD as a kind of 'clone' of a human, and this clone will be capable of classifying the real samples the same way as that particular human.

Variations of T-GAN and T-SGAN architectures may also include a more generic version of the TD, in which several 'humans' and several trainable discriminators (or clones) can be involved together. The basic architecture of such a 'Turing group discriminator' (TgD) is shown in Figure 4. Such discriminator includes N different non-trainable 'human' (H) components and the same number N of trainable discriminators (D) or potential 'clones' of corresponding Hs. Likewise T-SGAN, within such TgD, each individual discriminator D_i is trained to copy classification capability of the corresponding 'human' H_i so that (after training) the group of artificial-CI containing $D_1, D_2,..., D_N$ may replace human–CI containing $H_1, H_2,..., H_N$ within some decision-making process (e.g. in Industry 4.0). To enable this, the 'individual loss' is applied to each Di as a feedback for correctness of guessing the outcomes from corresponding H_i. Special feature of this architecture (Figure 4) is that, in addition to the capability of guessing an individual outcome, each D_i will be also trained to bias the compromised decision. The 'Compromise' component of the architecture collects outputs from each 'human' individual ($H_1, H_2,..., H_N$) and outputs the compromised (e.g. most supported) class label. During training, the outputs of each artificial discriminator D_i are compared with the compromised class label and the mismatch with yet another loss function ('compromise loss') are used as a feedback for D_i. Therefore, due to the 'individual loss', the D_i is trained to copy classification skills of corresponding H_i, but at the same time, due to the 'compromise loss', the D_i is trained to find a compromise with the others. Finally, after training, the artificial–CI group $D_1, D_2,..., D_N$ will preserve the hidden individual decision-making logic of the human–CI group $H_1, H_2,..., H_N$ and may not only simply replace it but also will be more mutually tolerant in finding a compromise in decision-making.

Classification is a type of decision-making problem, which involves a choice of an option (particular class label) from the finite number of the available ones. Similarly, one can consider someone's behaviour as a decision-making problem, i.e. choosing particular action from the available ones. This similarity allows adapting and using our GANs modifications (T-GANs, T-SGANS) not only for classification skills cloning but also for more generic behaviour-policies-cloning. Loss functions can be replaced with the award/punishment environmental feedback like in the reinforcement learning. For instance, the Inverse Reinforcement Learning (IRL) [29] idea has some similarities with the architectures around our TD.

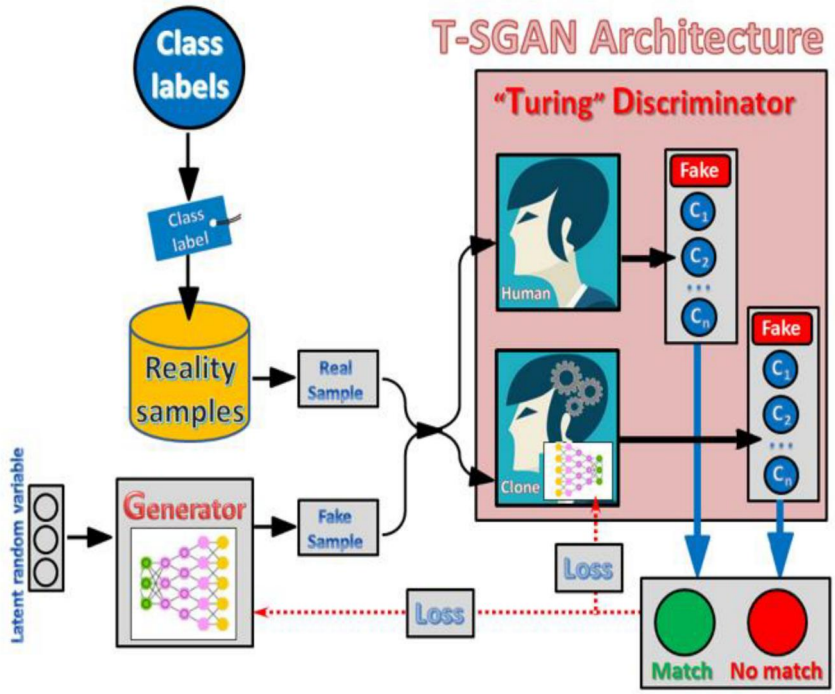

FIGURE 3 T-SGAN architecture, where an artificial 'clone' component learns to classify real or fake inputs the same way as a 'human'

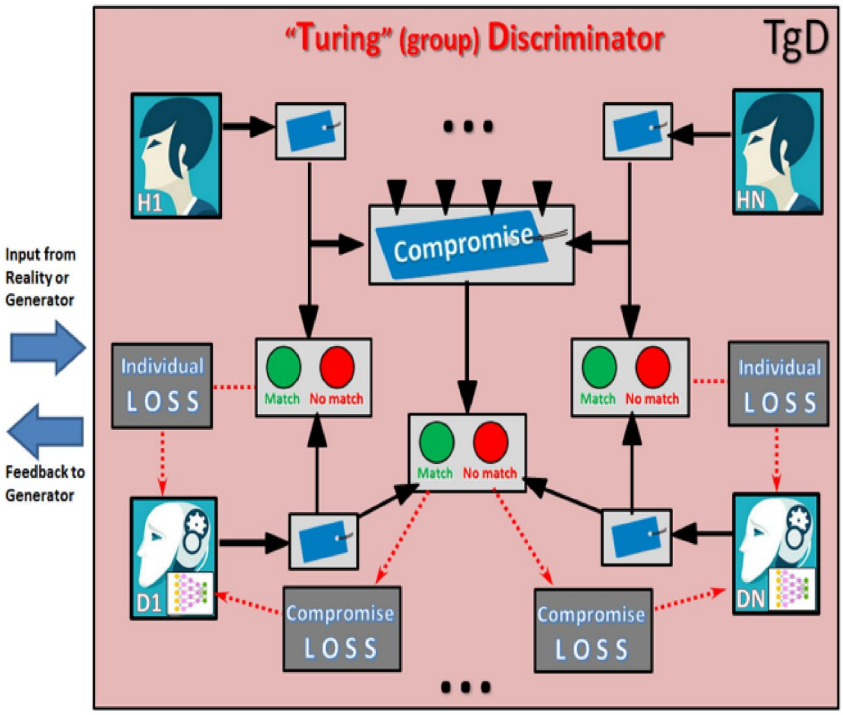

FIGURE 4 Architecture of a 'Turing group Discriminator', which can be used as a complex 'discriminator' component in various GAN architectures, particularly in T-SGAN

While ordinary reinforcement learning involves using rewards and punishments to learn behaviour, in IRL the direction is opposite, and an artificial agent observes a person's behaviour to figure out what goal that behaviour seems to be trying to achieve. In such learning, no reward function is given. Instead, the reward function is computed (inferred) given an observed behaviour from an intelligent target actor (e.g. an expert). The idea is to mimic observed behaviour.

In terms of IRL, the TD enables addressing the following problem. Given: (a) model of the reality (observable environment); (b) an observable target actor (e.g. a human) acting in the environment, and which (the target) behaviour is a subject of learning and cloning; (c) measurements of the data coming as an input to the target actor (sensory inputs) and the data coming as an output from the target actor (actuators' outputs), which are the measurements of a target (query–response)

behaviour over time, under a variety of circumstances. The goal of IRL process here would be to determine the reward function that the target is optimising and to use this function to reinforce the training of an artificial clone of the target.

As we have shown above, a TD (particularly TgD) supports not only one-to-one (target-clone) training but also more generic group-to-group training. The latter training option has some similarities with the multi-agent IRL. See, e.g., how in [30] authors extended the concept of the IRL to be an instrument for learning a group behaviour. They introduced the problem so that the reward functions of multiple agents are learned by observing their uncoordinated behaviour. After that, an abstract controller learns to coordinate behaviour of the group by optimising the weighted sum of the individual reward functions.

We believe the new suggested architectures (T-GAN, T-SGAN) and their TgD modification enable training efficient, responsible (due to human-like nature), and collaborative (biased to the compromises) artificial CI for a number of potential business processes within the Industry 4.0.

4 | CELLULAR COLLECTIVE INTELLIGENCE

What would be a reasonable size of a minimal CI team? In this article, we assume a human-centric nature of a CI meaning that in each CI team the leading role must remain for the humans. Therefore, the minimal CI team would include just one human and several autonomous AI components. We name such a team as a COIN 'cell'; the human there would be a cell-master; and the artificial components of the cell would be personal digital assistants of the cell-master.

Each COIN cell is designed as follows: (a) the cognitive clone of some human (potential cell-master) is created. The clone would contain the digital copy of the basic cognitive skills of the cell-master as pre-trained neural network models (see Section 3.2). The clone will be used as a basis for training some additional skills on top as personal digital assistants (transfer learning); (b) several digital personal assistants are trained simultaneously and each of them is a clone of the cell-master enhanced with some new specific extra skill; (c) the whole team (the master and the assistances) will be trained together on an adversarial environment (Figure 4) to learn to compromise while making decisions. As a result, the human will have a team of assistants around, who, on one hand, inherit the basic characteristics of the master and, on the other hand, have some extra capabilities each and these proactive capabilities are ready for a compromise decision-making.

The digital learning assistant begins its lifecycle as a digital clone of a given human and then develops itself on its own obtaining new cognitive skills that the human desires but does not possess, and, therefore, potentially enhancing (as a team player with the well-defined role) not only itself but also the cell-master. The assistant effectively addresses the challenges of natural human limitations: lack of memory and time. It will be always available to assist in decision-making processes, based on its own domain knowledge and its human counterpart's personality.

Besides the help in decision-making, the proposed digital learning assistant as an autonomous artificial cognitive system would track, analyse, and categorise relevant and useful content. As a result, it would keep up-to-date industry development, fill its own knowledge gaps, get professional top-level skills, and reinforce skills.

As Figure 5 shows, these assistants are complementing the human (shown as the central yellow cell) capacity, creating a new entity: a human enforced with their digital assistants—COIN cell. If you assume many persons have designed such COIN cells for themselves and appropriate infrastructure is available to support the intra-cell and inter-cell communication (like an agent platform), then these cells can together take part in very complex activities. Groups of collaborating cells we name as cellular CI (Figure 5), which can be a great flexible resource to enhance modern industry and various businesses.

It is important to mention that humans are constantly training and updating their skills in other ways than the AI does. How much synchronisation the training processes of humans, clones, and assistants would require? One of our studies [11] shows that there is a need for a common training place, University for Everything where humans, their clones and assistants will learn their complementary professional skills synchronously as a team. Imagine a situation when a graduate from such university, in addition to some certificate with list of courses and grades, will also obtain the personal digital team (completely trained COIN cell as an additional digital autonomous skillset) capable to help the graduate to perform his/her further professional activity. The pilot for such a

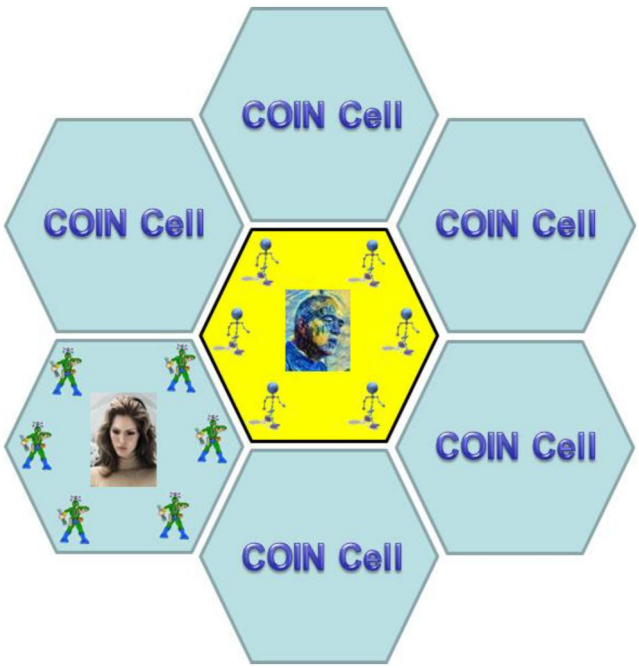

FIGURE 5 Cellular collective intelligence

'university' has been launched as an International Master Program on Cognitive Computing and Collective Intelligence [31] where we are supposed to combine traditional learning with ML to enable students to train their own COIN cells while learning themselves.

Security is an important concern for cellular CI as it contains specific vulnerabilities for both human and AI components. Especially critical would be protecting the training process from various data poisoning and evasion attacks, which are the major threats for the AI today. In [32], these attacks and their potential impact to critical Industry 4.0 infrastructure were discussed together with appropriate protection (e.g., artificial immune system).

5 | TESTING CLONED COLLECTIVE INTELLIGENCE IN ACTUAL SCENARIOS

The technology for cloning CI has been tested in three actual scenarios from private and public sectors: (1) for secure supply-chain and logistics within a real laboratory; (2) as a component of a middleware for the internet of things; and (3) for collaborative work management at the academic portal.

Within the first scenario, cloning experiments were performed and are ongoing in the framework of the NATO SPS project 'Cyber-Defence for Intelligent Systems' (http://recode.bg/natog5511) in the real logistics laboratory environment, where various kinds of adversarial attacks are generated to challenge the supervisory CI-driven AI systems [32]. We applied there special innovative GAN architecture presented in this article where an artificial adversary generates continuously evolving situations aiming to destroy the coordination among different players (by confusing the automated autonomous smart entities), who are taking care of secure logistics, supply chain, and delivery. Trained CI groups [33] are capable of coordinating their activities in adversarial situations and respond proactively to the new threats.

Within the second scenario, we extended the capabilities of UBIWARE, which is a middleware for the internet of things [34]. The middleware is based on the proactive digital twins of various industrial objects and processes [35,36]. Now, after we added the group cognitive cloning technology, described in this article as a feature of UBIWARE, the middleware become capable of coordinating groups of people with their digital cognitive clones and digital twins of smart industrial devices within the Industry 4.0 processes.

Within the third scenario, the group cloning experiments have been performed at the TRUST-Portal [37], which is an academic digital space for collaborative work of the humans and their digital clones [38]. Group cloning techniques presented in this article enabled automation of the collaborative cognitive work at the Portal (collaborative decision-making, co-supervision, collaborative recruitment, assessment, design, etc.) as well as managing compromises between individual and collective choices in various academic processes.

6 | CONCLUSIONS

In his recent masterpiece [39], Harari (the author of famous 'Sapiens' and 'Homo Deus') noted that it was a groupthink that allowed us (sapiens) to become the masters of the planet. Human views are largely shaped by collective mind (conscience and intelligence), and not by the individual rationality. It is just because of the ability for group and compromises, and this ability allows the humanity to surpass all other living species in its success [40]. Group decision-making is especially important in dynamic, uncertain, and contradictory situations, which are often taking place in the modern industry.

AI is used to automate many processes enabling step-by-step digital transformation of industrial processes towards the demands of the Industry 4.0. Autonomous (software) robots are capable of automating many of the decision-making processes. However, there is still the lack of solution on how to automate and enable that compromised coordination effect ('groupthink') as a special way of collaborative decision-making. If in [3] we have shown how to model the CI by training (using adversarial ML) digital cognitive clone for each of the individuals separately, in this article, we present a novel model of how to design and train the digital cognitive clones of the groups capable of the groupthink. We train the group clone as a compromise: on one hand, keeping as much as possible of the human individual features (donors of the individual digital clones) and, on the other hand, we train the capability of each group member to find reasonable compromises in making reasonable group decision from the individual expert opinions. We have also studied how to put trained (by adversarial ML) individual decision models (neural networks) into the shell of autonomous agents making these models proactive digital replica of humans; and we have developed a framework (cellular CI) for the enabling environment for the interaction and coordination of such smart personal digital assistants.

In this article, we significantly expand [3] addressing the concept of CI, also from the business point of view, meaning that CI is also a driver for digital transformation and is capable of managing complex business processes. In addition, we introduced the concept of RAI, which enables the creation of digital clones for simulated or real business environments. This framework has a chance to become trusted and adopted in various digitalised business ecosystems, and at the same time it will preserve the human-centric nature of the processes.

We also explain how to 'inject' the digitalised human intelligence (in the form of 'cognitive clones' of humans) into automated business processes and how to train such clones in artificial adversarial environments. We focused on several major aspects of such digitalisation where the CI is the key component including Smart Data Collection, Smart Data Integration and Representation, Machine Learning, Smart Decision-Support, Smart Process Automation and Customer Experience and Collective Intelligence Platform.

We suggested new architectures for GAN, which can help with individual and group cloning with the capability to find reasonable compromises in the decisions.

Well-organised CI is supposed to make better decisions than uncoordinated individuals. CI can work either as a group of independent players (human and artificial) or as a COIN cell (several digital assistants around some person), or even as an integrated group of COIN cells (cellular intelligence). In all the cases, an artificial player needed for the CI cannot be simply hardcoded, they must be trained. If, in addition to the 'business-as-usual' decisions, we want these entities to address new challenging situations (like COVID-19, for example), we must train them in complex and adversarial conditions.

We added (to the architecture of GAN) a human component, which allows the AI to make decisions synchronously with humans. We show how the modified GAN architectures can be used to train individual clones and groups of them so that they can take some responsibilities from the humans in making decisions and finding compromises in complex situations.

The proposed adversarial training framework and architecture of the CI are applicable to any situation within the Industry 4.0 when there is a need for collaborative and automated decision-making.

By studying the concept of CI training and introducing some new architectures for GAN, this article establishes the basis for future practical research and experiments on 'cognitive cloning'. Future work also includes further development of the CI training platform capable of supporting a wider scope of industrial applications.

ORCID

Vagan Terziyan https://orcid.org/0000-0001-7732-2962
Mariia Gavriushenko https://orcid.org/0000-0003-3440-7389
Anastasiia Girka https://orcid.org/0000-0002-6960-7420

REFERENCES

1. Forbes: 10 Ways AI Can Improve Digital Transformation's Success Rate. https://www.forbes.com/sites/louiscolumbus/2020/04/15/10-ways-ai-can-improve-digital-transformations-success-rate/#236898e65c43. Accessed 15 April 2020
2. Jarrahi, M.H.: Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Bus. Horizons. 61(4), 577–586 (2018)
3. Gavriushenko, M., Kaikova, O., Terziyan, V.: Bridging human and machine learning for the needs of collective intelligence development. Procedia Manuf. 42, 302–306 (2020)
4. Preindl, R, Nikolopoulos, K, Litsiou, K: Transformation strategies for the supply chain: the impact of industry 4.0 and digital transformation. Supply Chain Forum Intl. J. 21(1), 26–34 (2020) Jan 2
5. Buyuzkozkan, G., Gocer, F.: Digital supply chain: literature review and a proposed framework for future research. Comput. Indus. 97, 157–177 (2018)
6. Vial, G.: Understanding digital transformation: a review and a research agenda. J. Strateg. Inform. Syst. 28(2), 118–144 (2019)
7. Hinings, B., Gegenhuber, T., Greenwood, R.: Digital innovation and transformation: an institutional perspective. Inform. Organ. 28(1), 52–61 (2018)
8. Mitchell, T., et al.: Learning, never-ending. Commun. ACM. 61(5), 103–115 (2018)
9. Asilomar, A.: AI Principles, Future of Life Institute. https://futureoflife.org/ai-principles/. Accessed 26 September 2019
10. Ong, J., Ramachandran, S.: Intelligent tutoring systems: using AI to improve training performance and ROI. Netw. Newslett. 19(6), 1–6 (2003)
11. Pressey, S.L.: A simple apparatus which gives tests and scores-and teaches. Sch. Soc. 23, 373–376 (1926)
12. Leelawong, K., Biswas, G.: Designing learning by teaching agents: the Betty's Brain system. Int. J. Artif. Intel. Educ. 18(3), 181–208 (2008)
13. Golovianko, M., Gryshko, S., Terziyan, V.: From deep learning to deep university: cognitive development of intelligent systems. In: Szymański, J., & Velegrakis, Y. (eds.) Semanitic Keyword-based Search on Structured Data Sources, pp. 80–85. Springer, Cham (2017)
14. Grieves, M.W.: PLM beyond lean manufacturing. Manuf. Eng. 130(3) (2003)
15. Lee, J., Bagheri, B., Kao, H.A.: A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manuf. Lett. 3, 18–23 (2015)
16. Rajkumar, R., et al.: Cyber-physical systems: the next computing revolution. In: Design automation conference, pp. 731–736. Anaheim, CA (2010)
17. Terziyan, V., Gryshko, S., Golovianko, M.: Patented intelligence: cloning human decision models for Industry 4.0. J. Manuf. Syst. Part C. 48, 204–217 (2018)
18. Zhang, H., et al.: Digital twin in services and industrial product service systems: review and analysis. Procedia CIRP. 83, 57–60 (2019)
19. Schutzer, K., et al.: Contribution to the development of a Digital Twin based on product lifecycle to support the manufacturing process. Procedia CIRP. 84, 82–87 (2019)
20. Padovano, A., et al.: A digital twin-based service-oriented application for a 4.0 knowledge navigation in the smart factory. IFAC-PapersOnLine. 51(11), 631–636 (2018)
21. Souza, M.L.H., et al.: A survey on decision-making based on system reliability in the context of Industry 4.0. J. Manuf. Syst. 56, 133–156 (2020)
22. Sony, M., Naik, S.: Industry 4.0 integration with socio-technical systems theory: a systematic review and proposed theoretical model. Technol. Soc. 101248 (2020)
23. Fazel, A.: Knowledge Management 4.0: theoretical and practical considerations in cyber physical production systems. IFAC-PapersOnLine. 52, 1597–1602 (2019)
24. Lopez, C.P., Segura, M., Santórum, M.: Data analytics and BI framework based on collective intelligence and the Industry 4.0. In: Proceedings of the 2019 2nd International Conference on Information Science and Systems. Tokyo, Japan (2019)
25. Terziyan, V., Nikulin, A.: Ignorance-Aware Approaches and Algorithms for Prototype Selection in Machine Learning (2019) arXiv preprint arXiv:1905.06054
26. Goodfellow, I., et al.: Generative adversarial nets. Adv. Neural Inform. Process. Syst. 2, 2672–2680. (2014)
27. Odena, A.: Semi-Supervised Learning with Generative Adversarial Networks (2016) arXiv preprint arXiv:1606.01583
28. Madani, A., et al.: Semi-supervised learning with generative adversarial networks for chest X-ray classification with ability of data domain adaptation. In: Proceedings of the 15th IEEE international symposium on biomedical imaging, pp. 1038–1042. IEEE, Washington, DC (2018)
29. Ng, A.Y., Russell, S.J.: Algorithms for inverse reinforcement learning. In: Proceedings of the 17th International Conference on Machine Learning, vol. 1, pp. 663–670. Morgan Kaufmann Publishers Inc., Stanford, CA, USA (2000)
30. Natarajan, S., et al.: Multi-agent inverse reinforcement learning. In: Proceedings of the 2010 Ninth International Conference on Machine Learning and Applications, pp. 395–400. IEEE Computer Society, Washington, DC (2010)
31. COIN: Cognitive Computing and Collective Intelligence: International Master Program. Faculty of Information Technology, University of Jyvaskyla. http://www.jyu.fi/coin. Accessed 26 September 2019
32. Terziyan, V., Golovianko, M., Gryshko, S.: 'Industry 4.0 Intelligence under Attack: From Cognitive Hack to Data Poisoning', Cyber Defence in Industry 4.0 Systems and Related Logistics and IT Infrastructure. NATO Sci. Peace Secur. Series D. 51, 110–125 (2018)
33. NATO G5511 Project, http://recode.bg/natog5511
34. Katasonov, A., et al.: Smart semantic middleware for the internet of things. In: Filipe, J., Cetto, J.A., Ferrier, J.-L. (eds.) Proceedings of the 5th

International Conference on Informatics in Control, Automation and Robotics, vol. 1, pp. 169–178. SciTePress. https://doi.org/10.5220/0001489001690178
35. Terziyan, V., Zharko, A.: Semantic web and peer-to-peer: integration and interoperability in industry. Intl. J. Comput. Syst. Signal. 4(2), 33–46
36. Terziyan, V., & Katasonov, A.: Global Understanding Environment: Applying Semantic and Agent Technologies to Industrial Automation. In: Lytras, M.D., & De Pablos, P.O. (eds.) Emerging Topics and Technologies in Information Systems, pp. 55–87. IGI Global. https://doi.org/10.4018/978-1-60566-222-0.ch003
37. TRUST portal. (2016). http://portal.dovira.eu
38. Terziyan, V., Golovianko, M., Shevchenko, O.: Semantic portal as a tool for structural reform of the Ukrainian Educational System. Inform. Technol. Dev. 21(3), 381–402 (2005)
39. Harari, Y.N.: 21 Lessons for the 21st Century. Random House. New York (2018)
40. Sloman, S., Fernbach, P.: The Knowledge Illusion: Why we Never Think Alone. Penguin. New York (2017)

How to cite this article: Terziyan V, Gavriushenko M, Girka A, Gontarenko A, Kaikova O. Cloning and training collective intelligence with generative adversarial networks. *IET Collab. Intell. Manuf.* 2021;1–11. https://doi.org/10.1049/cim2.12008