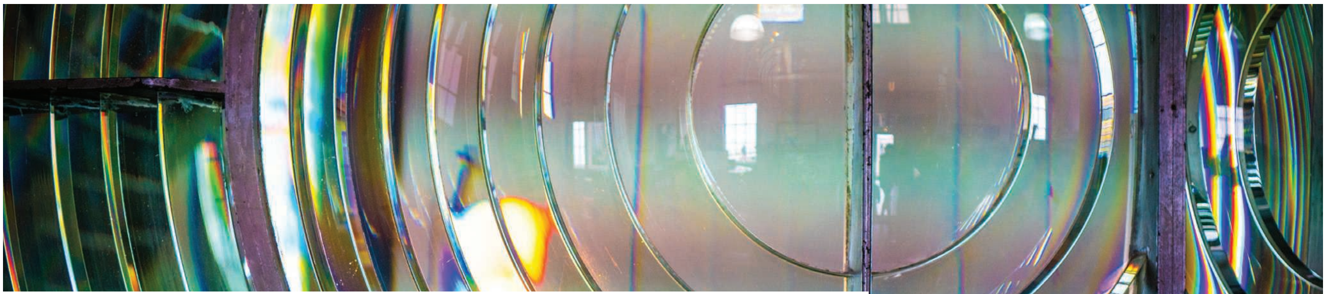
EDITOR MARK CAMPBELL
EVOTEK; mark@evotek.com



Generative Artificial Intelligence: Trends and Prospects

Mlađan Jovanović, Singidunum University
Mark Campbell, EVOTEK

Generative artificial intelligence can make powerful artifacts when used at scale, but developing trust in these artifacts and controlling their creation are essential for user adoption.

Generative modeling is an artificial intelligence (AI) technique that generates synthetic artifacts by analyzing training examples; learning their patterns and distribution; and then creating realistic facsimiles. Generative AI (GAI) uses generative modeling and advances in deep learning (DL) to produce diverse content at scale by utilizing existing media such as text, graphics, audio, and video.[1,2] While mainly used in research settings, GAI is entering various domains and everyday scenarios. This article sheds light on the unique practical opportunities and challenges GAI brings.

Digital Object Identifier 10.1109/MC.2022.3192720
Date of current version: 26 September 2022

GAI TECHNIQUES

Although there are many forms of GAI, we will look at four of the most common techniques being leveraged today.

Generative adversarial networks

Generative adversarial networks (GANs) are the most prevalent GAI technique being used today.[3] A GAN uses a pair of neural networks. One, known as the *generator*, synthesizes the content (for example, an image of a human face). The second, known as the *discriminator*, evaluates the authenticity of the generator's content, (that is, whether the face is natural or fake). The networks repeat this generate/discriminate cycle until the generator produces content that the discriminator cannot discern between real and synthetic.

Generative Pre-trained Transformer

Generative Pre-trained Transformer (GPT) models generate text in different languages and can create human-sounding words, sentences, and paragraphs on almost any topic and writing style—from convincing news articles and essays to conversations in customer-service chatbots or characters in video games.[4] These have matured over several generations, each with an increased parameter set trained on a more extensive online textual corpus than the previous. One recent example is OpenAI's GPT-3, which stunned the AI world by writing, without human assistance, a convincing article about scientists discovering a herd of unicorns in the Andes.[5]

The generative diffusion model

The generative diffusion model (GDM) synthesizes content by taking a training data distribution, gradually adding noise, and learning how to recover the data as a reversal of the noise addition process.[6] This way, data are generated from randomly sampled noise through the learned denoising process.

Geometric DL

Geometric DL (GDL) attempts to understand, interpret, and describe AI models in terms of geometric principles. These principles have already been extensively studied over domains such as grids; transformations in homogeneous spaces; graphs; and vector bundles.[7] Petar Veličković,[8,9] a staff research scientist at DeepMind and Affiliated Lecturer at the University of Cambridge, describes it as

> "…building machine learning (ML) models that respect variances and symmetries inside data. For example, if you know your data lives on a grid and it should be translation symmetric, you should end up driving something like a convolutional neural network [CNN]. So, our blueprint that consists of the chosen data domain and symmetry group you want to be resistant can guide different ML architectures such as CNNs, GNNs, transformers, or recurrent models. For example, transformers are a special case of attentional graph neural networks over the complete graph."

USING GAI

These and other GAI techniques are being used in a host of applications, including the following.

Natural language and music

GPTs can be readily applied to natural language (NL) text generation. GPT-3,[4] mentioned previously, has been successfully scaled to 175 billion learnable parameters and trained on global-scale corpora of textual exemplars. Aside from showing high performance on a variety of NL processing (NLP) tasks, such as translation and question answering, it is also a competent text generator producing eerily human-like textual content.[10]

For example, a GPT-3 program wrote an entire student essay from a simple prompt ("The construct of 'learning styles' is problematic because…").[11] The narrative flowed as if it was written by a human—plagiarism software did not detect copying, and a Google search showed that each sentence was original. The authors highlighted several potential uses of GAI tools in education, such as facilitating creative writing (for example, students and AI write paragraphs intermittently to explore alternatives and overcome writer's block) or developing a critical analysis of academic writing (for example, AI generates texts on a topic, and students critique and revise them).

Language Model for Dialogue Applications (LaMDA) is another example. This generative, textual conversational agent mimics human conversations,[12] but unlike GPT models trained on text corpora, it is trained on dialog corpora. Objective-Reinforced GAN (ORGAN) is another example that produces time series artifacts in sequential media such as music.[13]

> While mainly used in research settings, GAI is entering various domains and everyday scenarios.

Computer graphics

AlphaFold is a neural network that creates highly accurate 3D protein structures[14] by modeling and predicting protein structures as a graph inference problem in 3D space where nearby residues define the edges of the graph. The pair representation is encoded as a directed edge in a graph (that is, the connection between the residues). The NVIDIA Canvas application GauGAN transforms a textual phrase like "ocean waves hitting rocks on the beach" into virtual landscape images in real time. When adding adjectives like "sunset at a rocky beach" or swapping "sunset" for "afternoon" or "rainy day," the model modifies the picture instantly.[15] Similarly, DALL•E is a compiled version of GPT-3 that produces images from text descriptions for concepts expressed in NL, taking text/image pairs as input.[16] The latest GDM-based approaches for text-to-image generation are DALL•E 2[16,17] and Imagen,[18] capable of producing diverse, high-quality artistic and realistic images, respectively. 3D-GAN creates 3D shapes[19] that can be manipulated in 3D spaces (geometric transformation) and then scaled down to 2 D image representations.

Computer vision

Using semantic label maps as an input, conditional GANs (CGANs) can produce images of high-fidelity urban scenes containing objects. Changing labels modifies scenes concerning individual objects, such as replacing trees with buildings or changing colors or textures.[20] TediGAN (Text-Guided Diverse Face Image Generation and Manipulation) creates human portrait drawings from facial photos with random changes to facial attributes.[21] SinGAN[22] is a single-image generative model that synthesizes realistic textures of arbitrary size and aspect ratio with significant variability.

Motion and Content decomposed GAN (MoCoGAN)[23] conducts video synthesis with two modalities separated. The content can have different movements. Conversely, it can apply the same action to varying content. Enhanced Super-Resolution GAN (ESRGAN) improves the quality of the media through super-resolution, a group of ML methods that upscale low-resolution image or video media to a higher resolution.[24]

Figure 1 illustrates the accuracy/complexity tradeoff for some typical GAI products. The *accuracy* refers to the intended performance of synthesized artifacts, while the *complexity* denotes their richness in media content and structure.

EVALUATION

Metrics to verify the GAI efficiency and effectiveness of a generation task can reuse current objective ML model evaluation techniques.[1,2] For instance, GANs' functional quality metrics focus on their inner workings by evaluating outputs such as image quality; resolution; inception score (that is, the realism of generated images); and training time reduction.[25] Nonfunctional evaluation, privacy, and security are also significant concerns for GANs.[26] Some GAI evaluations, like a personal rating of the output for utility, aesthetics, clarity, or similarity to real-world content, are inherently subjective and difficult to evaluate.

Currently, there are no standard means to determine if a GAI is as realistic to a user as a non-GAI application. Respectively, reliable and consistent measurements of the effects of GANs are still undetermined. Relatedly, Veličković[9] notes the following:

> "Geometric approaches are going to be very important for generative modeling because the data you are generating will need to respect some kind of geometry if you want your solutions to be well constrained. Therefore, models that are mindful of geometry are, in my opinion, more likely to succeed in the long run."

EXPECTATIONS

Along with algorithmic improvements,[25,26] future GAI solutions must meet the following expectations to gain user trust and adoption.

Efficiency

Training and deploying GAI models leave a significant carbon footprint and high computation costs. For example, GDMs naturally lag in sampling speed.[27] Creating a fine-tuned, downsized model for input data and parameter space is a cost-efficient approach for researchers and practitioners.

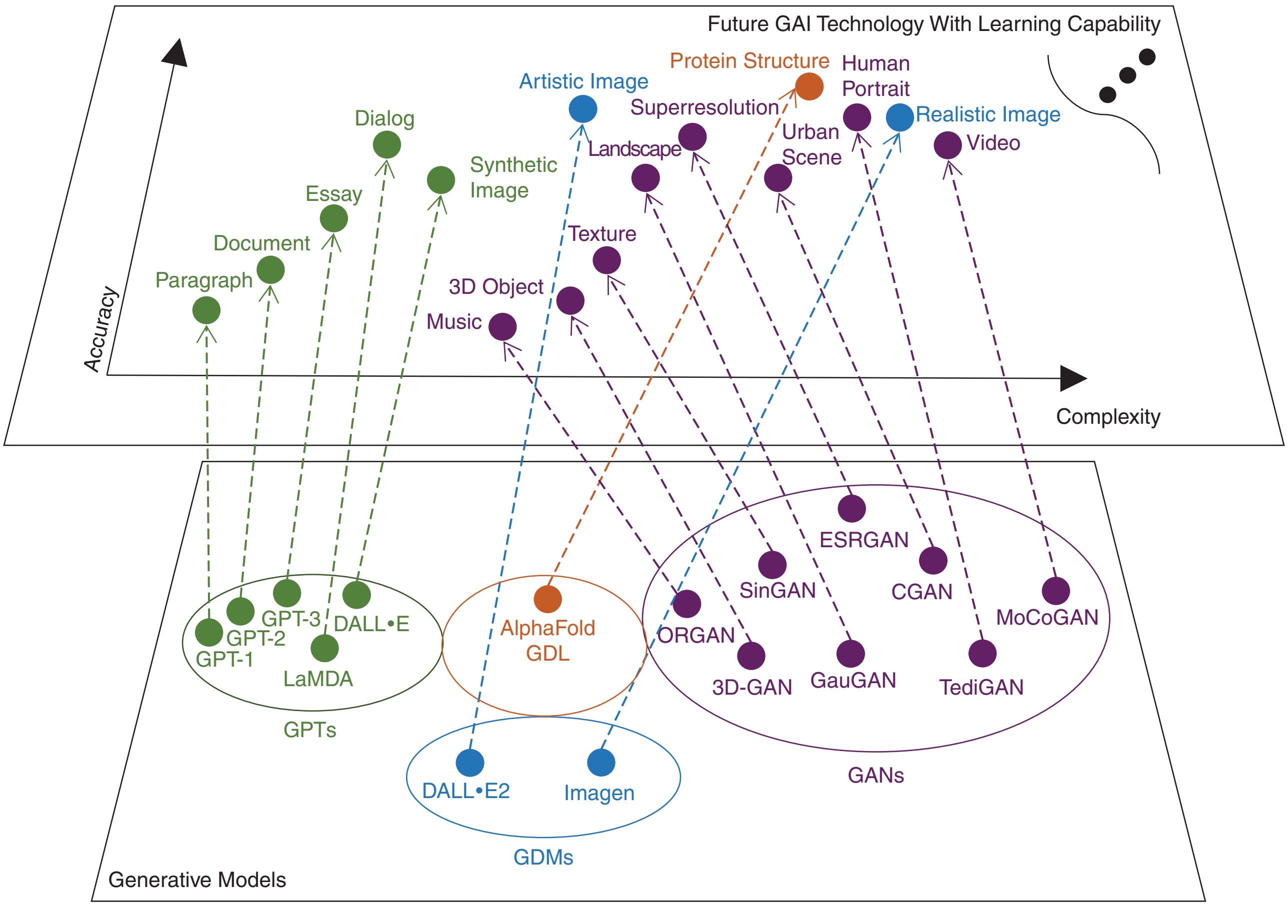


FIGURE 1. The GAI landscape: generative models and artifacts.

Explainability

Explaining the mechanisms of a deep neural network involves analyzing the input data properties (also known as *features*) to determine which affect the outcomes and infer what happens inside the black box. However, determining which neurons affected the synthesis of which output objects remains problematic today. Moreover, in the case of GANs, quantifying the mutual behavior of a pair of networks is currently an intractable problem. A deeper understanding of GAI models' inner workings and explanations tailored to specific user groups are both still missing. Accordingly, Veličković[9] notes the following:

> "Explainability will be very, very important for generative modeling. In our recent research,[28] mathematician collaborators took the explanations produced by our models and tried to prove a mathematical theory based on these explanations, which is significantly more rigorous than the original explainer's output. What came out from these models was, to us, quite unintelligible. The mathematicians were able to spot the signal hidden inside the tons of noise after looking at sent explanations for two weeks. So, explainability methods are rudimentary, and if we truly want to make it accessible to everyone, we really need explainability techniques that have a far better signal-to-noise ratio."

GAI has become a key technology in synthesized new virtual artifacts or enhanced semisynthesized augmented artifacts.

Fairness

Although generative language models such as GPTs offer marked improvements over various NLP tasks, they require massive amounts of unfiltered online text. Consequently, they can generate synthetic language with bias, stereotypes, and harmful content.[29] To manage these risks, GAI providers should offer tools for preprocessing and curating training data; monitoring and moderating the media generation processes; and developing guidelines for responsible deployment models.

Ethics

GAI models can immediately synthesize artifacts at scale for many different contexts, from education to medical decision making. However, before diving into production deployment, model creators should clearly define their goals; identify beneficiaries; and confirm usage scenarios with target users to prevent unintended unethical product behavior. This requires that all affected stakeholders—GAI scientists, AI engineers, domain experts, regulatory authorities, and target users—are identified and actively participate.

Accountability

Prospective users must weigh GAI products' benefits against their risks. Organizations that are creating, training, and deploying GAI systems must diligently strive to reduce model behavior risks. This requires teams to be thorough, transparent, and proactive in communicating identified threats, blind spots, and areas where risks are unknown when highlighting GAI system benefits.

GAI WORKFLOW

To address these challenges, we envision a future GAI deployment workflow as in Figure 2. This proposed workflow can provide benefits beyond GAI models. Because synthetic data creation increases the data set size for a particular GAI model, it can become part of the model's iterative and incremental development and deployment cycle to continually improve performance. Specifically, if a GAI model works well overall but performs poorly on certain features, more data can be generated for those critical categories to help detect and correct errors.

Moreover, traditional ML workflows take good testing results on used data sets as an indicator of modeling input data distribution. However, physical- and digital-world data sets collected over several model generations will change their class structure and internal connections. Newly synthesized objects or artifacts, such as a language, an environment, or a (human) being, are examples of the data sets. When observing and evaluating such changes in deployment, the straightforward way to expand generative capacity is to retrain the model with modified class structure data.

GAI has become a key technology in synthesized new virtual artifacts or enhanced semisynthesized augmented artifacts. GAI's achievements in many fields are paving the way for synthetic sciences to combine AI with basic disciplines such as engineering, biology, medicine, and environmental science. Although current advancements within a discipline are seldom connected with developments in others due to various sociotechnical factors like communication norms, cultural differences, AI models, data, and

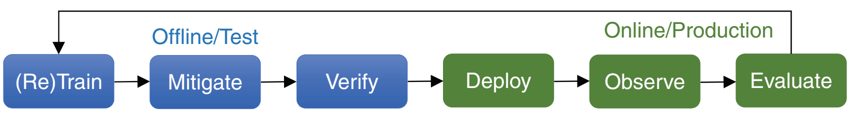


FIGURE 2. The future GAI model deployment workflow as a closed system with a feedback loop.

procedures, the metaverse is a promising global, interdisciplinary testbed for resolving these obstacles. GAI would diminish the distinction between real and virtual artifacts to meld human experience and behavior across the virtual and physical worlds.

GAI models and the synthetic artifacts they create are ever increasing in prevalence, adoption, and sophistication. As verification, risk mitigation, and cross-disciplinary deployment techniques evolve, GAI solutions will become a foundational feature of any data architecture. Challenges like model explainability, cultural acceptance, and sociotechnical issues remain but will be surmounted as GAI applications evolve into an escalating arena of deployment. ■

ACKNOWLEDGMENT

We thank Petar Veličković for his constructive suggestions and helpful comments.

REFERENCES

1. J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *IEEE Trans. Knowl. Data Eng.*, early access, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9625798/authors#authors, doi: 10.1109/TKDE.2021.3130191.
2. M. Abukmeil, S. Ferrari, A. Genovese, V. Piuri, and F. Scotti, "A survey of unsupervised generative models for exploratory data analysis and representation learning," *ACM Comput. Surv.*, vol. 54, no. 5, pp. 1–40, 2021, doi: 10.1145/3450963.
3. I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, vol. 27, pp. 1–9. [Online]. Available: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
4. T. Brown *et al.*, "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 1877–1901. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
5. "GPT-3 discovers unicorns." GPT-3. https://www.buildgpt3.com/post/88/ (Accessed: Jul. 1, 2022).
6. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf
7. M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," 2021, *arXiv:2104.13478*.
8. C. K. Joshi. "Transformers are graph neural networks." The Gradient. https://thegradient.pub/transformers-are-graph-neural-networks/ (Accessed: Jul. 15, 2022).
9. P. Veličković, private communication, Jul. 5, 2022.
10. "Text generation examples." OpenAI. https://beta.openai.com/examples?category=generation (Accessed: Jul. 15, 2022).
11. M. Sharples and R. y. Pérez, *Story Machines: How Computers Have Become Creative Writers*. London, U.K.: Routledge, 2022. [Online]. Available: https://www.routledge.com/Story-Machines-How-Computers-Have-Become-Creative-Writers/Sharples-Perez/p/book/9780367751975
12. R. Thoppilan *et al.*, "LaMDA: Language models for dialog applications," Feb. 10, 2022. Accessed: Jul. 15, 2022. [Online]. Available: https://arxiv.org/abs/2201.08239
13. G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias, and A. Aspuru-Guzik, "Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models," May 30, 2017. Accessed: Jul. 15, 2022. [Online]. Available: https://arxiv.org/abs/1705.10843
14. J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021, doi: 10.1038/s41586-021-03819-2.
15. T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "GauGAN: Semantic image synthesis with spatially adaptive normalization," in *Proc. SIGGRAPH 2019 Real-Time Live!* New York, NY, USA, 2019, p. 1, doi: 10.1145/3306305.3332370.
16. A. Ramesh *et al.*, "Zero-shot text-to-image generation," in *Proc. 38th Int. Conf. Mach. Learn.*, 2021, pp. 8821–8831. [Online]. Available: https://proceedings.mlr.press/v139/ramesh21a.html
17. A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," Apr. 13, 2022. Accessed: Jul. 15, 2022. [Online]. Available: https://arxiv.org/abs/2204.06125
18. C. Saharia *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding," May 23, 2022. Accessed: Jul. 15, 2022. [Online]. Available: https://aps.arxiv.org/abs/2205.11487
19. J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum, "Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling," in *Proc. Adv. Neural Inf. Process. Syst.*, Barcelona, 2016, pp. 82–90, doi: 10.5555/3157096.3157106. [Online]. Available: https://papers.nips.cc/paper/2016/hash/44f683a84163b3523afe57c2e008bc8c-Abstract.html
20. T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, 2018, pp. 8798–8807. [Online]. Available: https://ieeexplore.ieee.org/document/8579015, doi: 10.1109/CVPR.2018.00917.
21. W. Xia, Y. Yang, J.-H. Xue, and B. Wu, "TediGAN: Text-guided diverse face image generation and manipulation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*,

Nashville, 2021, pp. 2256–2265. [Online]. Available: https://ieeexplore.ieee.org/document/9578577, doi: 10.1109/CVPR46437.2021.00229.
22. T. R. Shaham, T. Dekel, and T. Michaeli, "SinGAN: Learning a generative model from a single natural image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, 2019, pp. 4570–4580. [Online]. Available: https://ieeexplore.ieee.org/document/9008787, doi: 10.1109/ICCV.2019.00467.
23. S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "MoCoGAN: Decomposing motion and content for video generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, 2018, pp. 1526–1535. [Online]. Available: https://ieeexplore.ieee.org/document/8578263
24. X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis.-ECCV Workshops*, Munich, 2018, pp. 63–79, doi: 10.1007/978-3-030-11021-5_5.
25. T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," Feb. 26, 2018. Accessed: Jul. 15, 2022. [Online]. Available: https://arxiv.org/abs/1710.10196
26. Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, "Generative adversarial networks: A survey toward private and secure applications," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–38, 2022, doi: 10.1145/3459992.
27. A. Vahdat and K. Kreis. "Improving diffusion models as an alternative to GANs." nvidia. https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/ (Accessed: Jul. 15, 2022).
28. A. Davies *et al.*, "Advancing mathematics by guiding human intuition with AI," *Nature*, vol. 600, no. 7887, pp. 70–74, 2021, doi: 10.1038/s41586-021-04086-x.
29. P. Schramowski, C. Turan, N. Andersen, C. A. Rothkopf, and K. Kersting, "Large pre-trained language models contain human-like biases of what is right and wrong to do," *Nature Mach. Intell.*, vol. 4, no. 3, pp. 258–268, 2022, doi: 10.1038/s42256-022-00458-8.

MLAĐAN JOVANOVIĆ is an assistant professor at Singidunum University, Belgrade, 11000, Serbia. Contact him at mjovanovic@singidunum.ac.rs.

MARK CAMPBELL is the chief innovation officer for EVOTEK, San Diego, California, 92121, USA. Contact him at mark@evotek.com.

Digital Object Identifier 10.1109/MC.2022.3202583