Check for updates

editorial

Into the latent space

Generative deep learning can produce artificial, natural-looking images and other data, which has many promising applications in research — and in art. But the wide availability of generative models poses a challenge for society, which needs tools and best practices to distinguish between real and synthetic data.

The first wave of interest in deep learning as a transformative tool for scientific research began over a decade ago. Now there is fresh excitement over a new set of capabilities: as deep neural networks can learn underlying patterns in complex data, this can be taken a step further by designing algorithms that can explore the latent possibilities of the learned distributions, to produce original examples of data, without using any design rules.

Among the pioneers of generative tools are artists. The results of AI art made with GANs — generative adversarial networks — have made headlines, for instance when the first AI-generated fictional painting was auctioned at the art auction house Christie's in 2018 and sold for US$432,500. The artwork, a portrait called *Edmond de Belamy*, could be mistaken for a work from a seventeenth century master. Instead, it is the product of the modern technique of training a generative model on 15,000 portraits from the thirteenth to the nineteenth centuries, which was implemented by a group of French artists collectively known as Obvious.

Many questions arise, as indeed they did over the Belamy portrait sale. If an algorithm produces the art, who is the artist? Can the product be called 'art'? And who is the legal owner of the product? Obvious used a model that is openly available. The training data are publicly available. But they still had to choose a specific style of portraits and identify a suitable image in the model's multidimensional latent space. Many more interesting images could be 'found' by the artists and in fact they produced a whole family tree of portraits for the fictitious Belamys. A Perspective in this issue by Jason Eshraghian analyses several legal issues around the use of generative models in art and design, and gives recommendations regarding the need to document the full process of training and the selection of data, models and parameters.

Documenting data sources and training procedures will also be best practice for scientists using generative approaches. An area that, like art, can gain much from combining human creativity with AI generative tools is drug discovery. The chemical space of possible molecules is enormous: for drug-like molecules, it's somewhere near the order of 10^{60}. For more practical small molecules up to 30 atoms, the estimate is between 10^{20}–10^{24} (ref. [1]). Running preliminary trials to synthesize and assess the bioactivity of a potential drug is expensive and time intense, so the candidate molecules have to be selected carefully, with human experts involved in both constraining the generative models into desired properties and final selection of candidate drugs. An Article in this issue by Michael Moret et al. develops a generative machine learning framework to design new molecular entities for specific target applications, where only small sets of training examples are available. This is an example of an approach now gaining popularity: training a generative model with a large selection of generic data and subsequently fine-tuning the model towards a specific kind of data of which few training examples might exist. Intuitively this decouples the task of learning the typical structure of molecules and the specific features desired.

Picture produced with generative AI art tool, artbreeder, by Jacob Huth.

Generative models can also be employed to address challenges in the wide adoption of machine learning applications, such as image recognition in medical diagnosis and by self-driving cars. A thorny issue is that deep learning algorithms can be fooled easily with adversarial attacks, which alter an input image in a way that radically affects the classification. GANs can be employed to filter out such alterations, before feeding an image into a classifier[2]. Another challenge that generative AI could tackle is the need to protect patient privacy in sharing large medical datasets. Generative models can produce realistic but synthetic data where personal information is removed[3].

But flooding the world with artificial data, images, text, videos and more comes at a price, as the world is aware. Faked documents and deceit is nothing new, but powerful deep learning tools are now within everyone's reach. Open-source toolboxes make it easy for anyone to produce photographs of people who do not exist, write fake news articles using OpenAI's GPT-2 model, and of course, create a new Van Gogh. How will we know what's real and fake? The spread of made-up news and other disinformation disrupts democratic processes. There is an urgent need for regulations, as well as for tools to identify fakes such as Assembler from Jisgsaw and Google research, which spots a specific type of deepfake that is made with an approach called styleGAN[4].

Generative AI is still a process in which humans are actively involved at various stages. It seems more important than ever for developers, users and researchers to document data sources, training procedures and outputs. This will be essential for establishing ownership, but also to offer transparency regarding the origins, motivation and intended use of a generated dataset. It may also be a good time to promote again the proposal from ref. [5], which encourages AI developers to provide data sheets for datasets, to add detailed information about how and why a publicly available dataset has been created. Future potential users will be grateful for the additional effort. ❐

Published online: 16 March 2020
https://doi.org/10.1038/s42256-020-0164-7

References

1. Reymond, J.-L. *Acc. Chem. Res.* **48**, 722–730 (2015).
2. Samangouei, P., Kabkab, M. & Chellappa, R. In *Proc. 6th Int. Conf. Learning Representations* (ICLR, 2018).
3. Beaulieu-Jones, B. K. et al. *Circ. Cardiovasc. Qual. Outcomes* **12**, e005122 (2019).
4. Beridze, I. & Butcher, J. *Nat. Mach. Intell.* **1**, 332–334 (2019).
5. Gebru, T. et al. Preprint at https://arxiv.org/abs/1803.09010 (2018).