From Art to Science: Bridging gaps in diffusion–based generative modeling

Generative models represent a groundbreaking leap in artificial intelligence, captivating researchers, developers, and the public with their unprecedented ability to create content that mirrors the intricacies of human-generated data. This is especially visible in numerous creative usages such as text-to-image and, recently, text-to-video generations, where generative models create astonishing visualizations from short textual descriptions. Those impressive results are possible because of the objective of generative modeling, which is to learn how to mimic training data distribution. In the context of image generation, the distribution is defined as a set of training images (usually colossal in size – over 5 billion examples). Hence, the model learns how to create realistic images by arranging the image pixels so that the final output is similar to the training data, and not just a combination of unrelated random colors.

Apart from the exciting possibilities of generating new funny images of cats, the same generative methods can be applied to scientific programming. The possibility of generating new instances similar to the already gathered data brings unique opportunities in areas such as drug and new materials discovery or high energy physics. In those use cases, generative models can be used as simulation techniques that can validate theoretical hypotheses or speed up the new experiment design process.

Nevertheless, despite using similar methods, those two setups are driven by different objectives. In the majority of recent use cases in creative usages, the end goal of the generative model is to create a visually pleasant new example that is similar to the images created by humans. On the contrary, in scientific applications and simulations, the main focus is to precisely model the true data distribution. The difference between those two goals is evident. For example, the under-representation of red cats in a generator of funny images is a less significant drawback than missing physical phenomena or subtle biochemical structures in scientific applications. At the same time, peculiar generations outside of the training data domain can result in new interesting style of generated illustrations, but lead to physically impossible conclusions when used to validate high energy physics experiment.

In this project, we propose to study the missing elements in generative modeling task in order to use them in scientific computation. Our focus will be on assessing the reliability of diffusion-based generative models (DDGM) in data probability modeling. We plan to first analyze recent models' limitations, such as memorization of training examples and limited coverage of training data distribution. On top of this analysis, we will develop new methods that can add this missing knowledge to the existing model. Moreover, we plan to take a closer look into wrongly generated examples – so-called *hallucinations*. We will develop a method that identifies untrue generations and unlearns how to generate them. inally, we will explore strategies for updating the model with additional data without compromising its performance on already known data.