Theory of memorisation and generalisation in diffusion models of generative AI: a statistical physics perspective

Giulio Biroli[1]

[1]ENS Paris , France

Generative AI represents a groundbreaking development within the broader "Machine Learning Revolution," significantly influencing technology, science, and society. In this talk, I will focus on remarkable connections between generative AI, in particular diffusion models, and statistical physics. Diffusion models are the state of the art to generate images, videos, and sounds. They are very fascinating algorithms for physicists, as they are very much connected to concepts from stochastic thermodynamics, particularly time-reversed Langevin dynamics. These diffusion models start from a simple white noise input and make it evolve through a Langevin process to generate complex outputs such as images, videos, and sounds. I will show that statistical physics provides principles and methods to characterise the generative process. I will focus a central puzzle of this field: how these models, which are trained to memorize the data, instead end up producing genuinely new outputs. I will explain what is the mechanism underlying the emergence of creativity (generalisation), or its absence (memorisation), and how it can be analyzed using theories of disordered systems, and dynamics in complex energy landscapes.