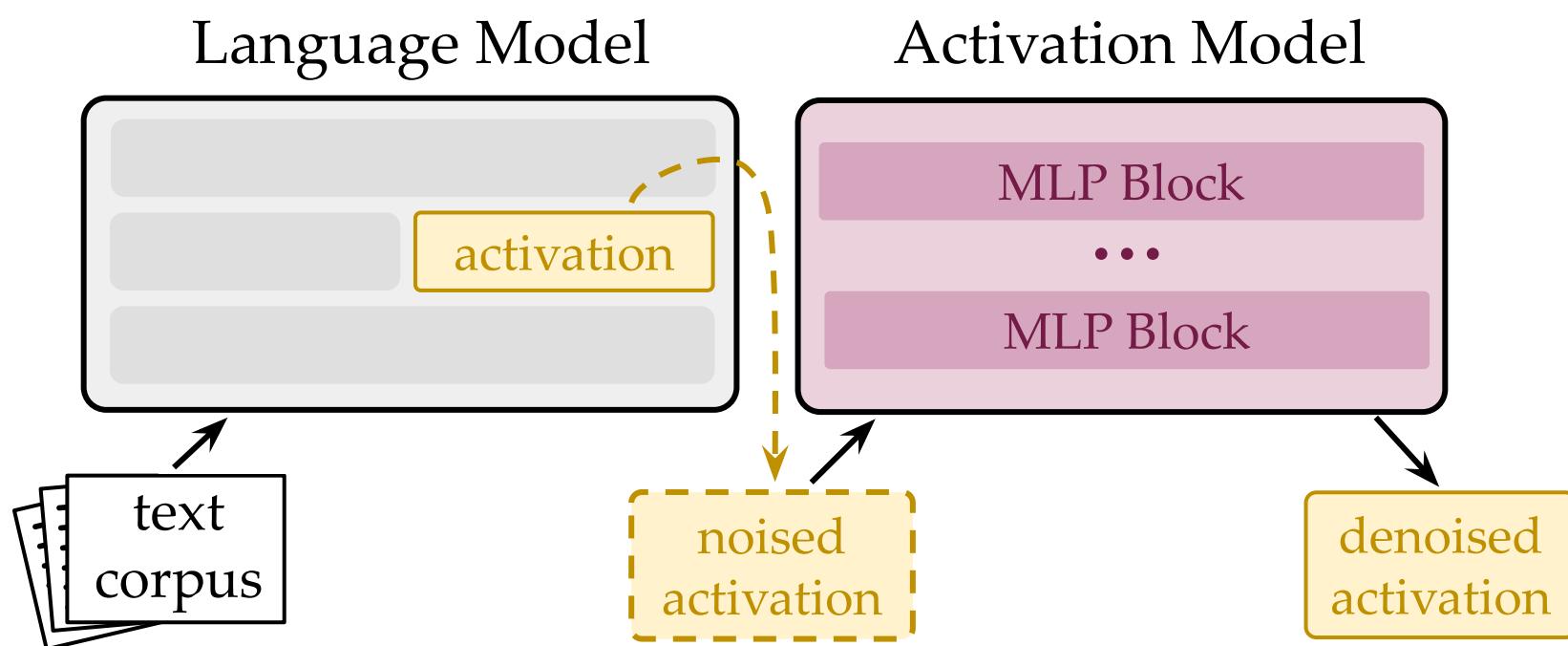
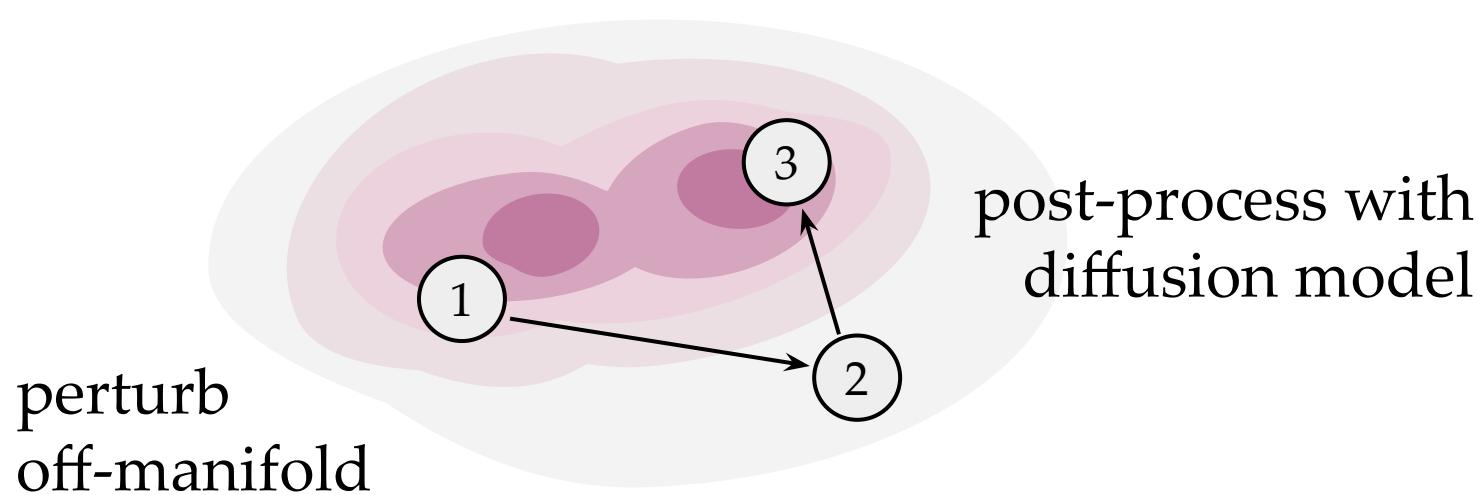


(a) Training an activation diffusion model...



(b) ...learns the structure of the activation manifold...



(c) ...enabling applications such as on-manifold steering

Steering Task:
terms related
to scientific
testing
methods and
protocols

SAE	+ Ours
How to measure the determination of the method for the determination of the method for the determination of the method...	The answer is simple and easy, by following... the National Association of Official Testing Methods...