

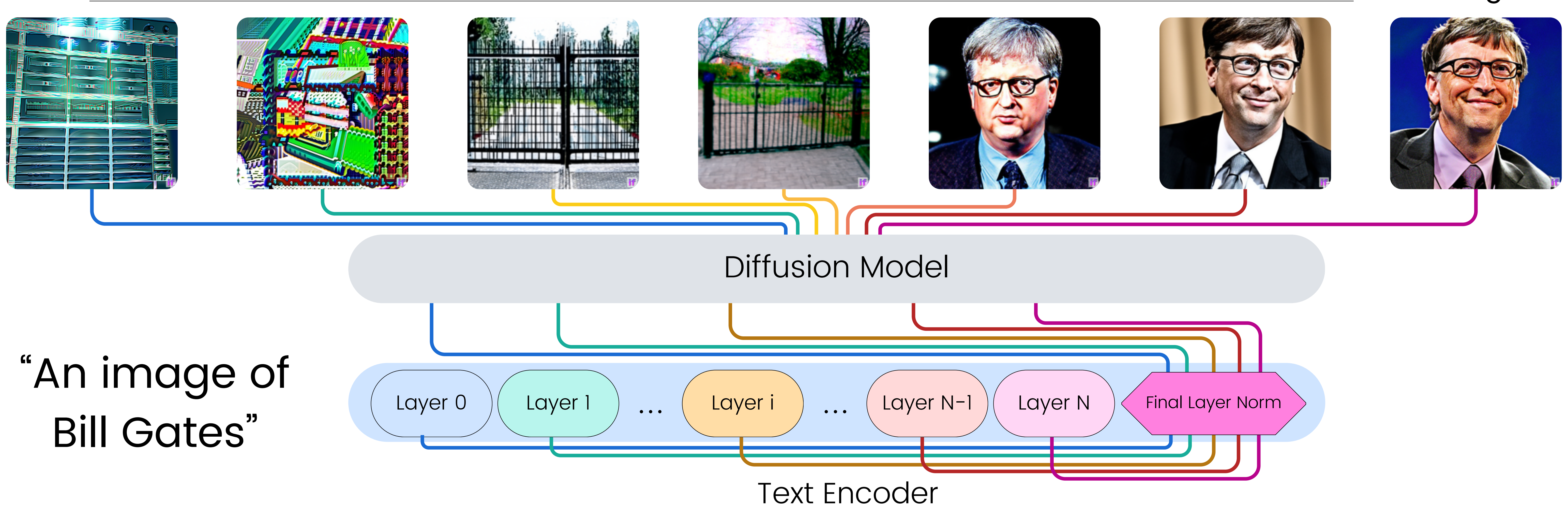


Diffusion Lens: Generating images from hidden states reveals insights into computation process

Michael Toker, Hadas Orgad, Mor Ventura, Dana Arad, Yonatan Belinkov

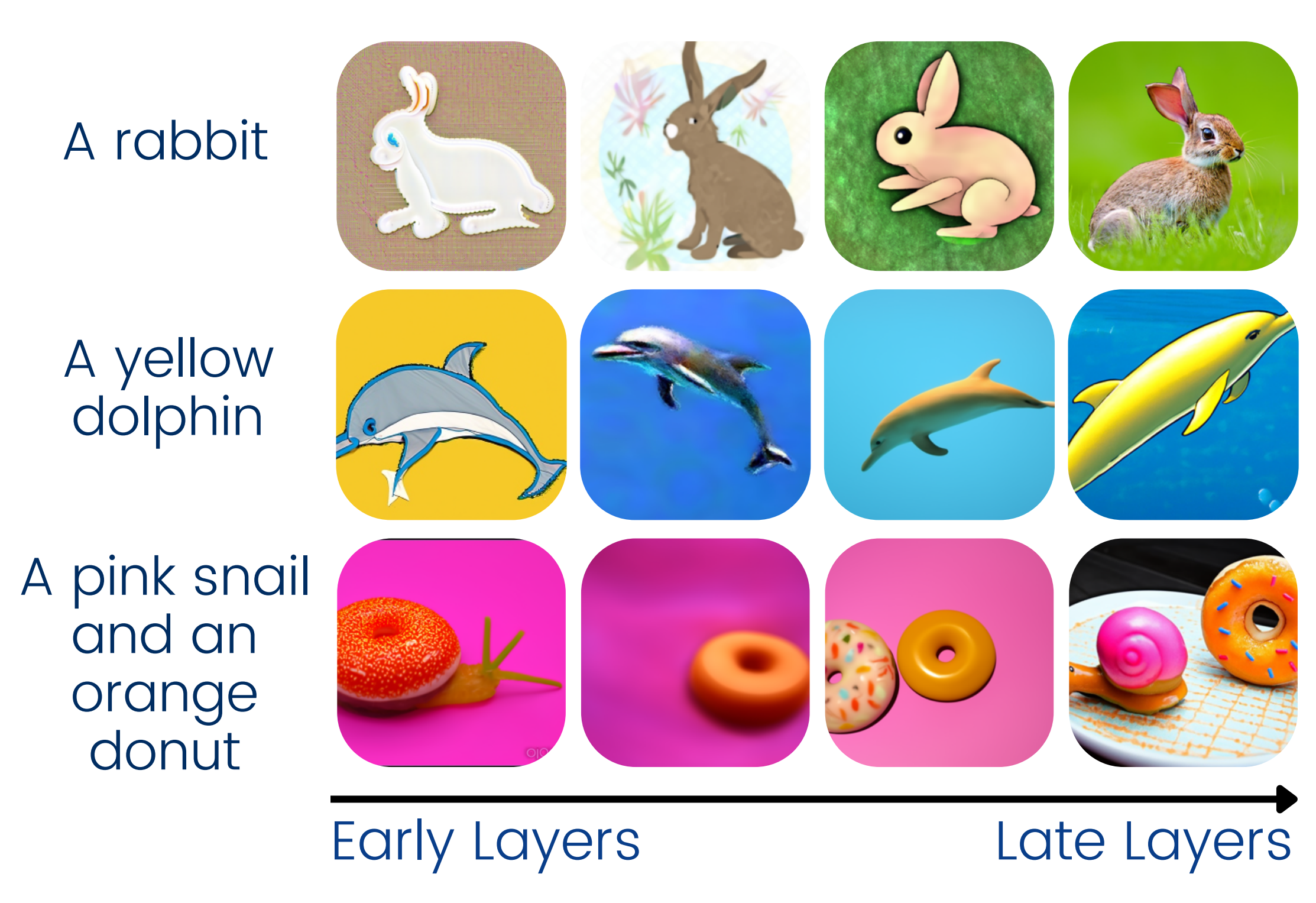
Demo & Paper

Diffusion Lens 🔍

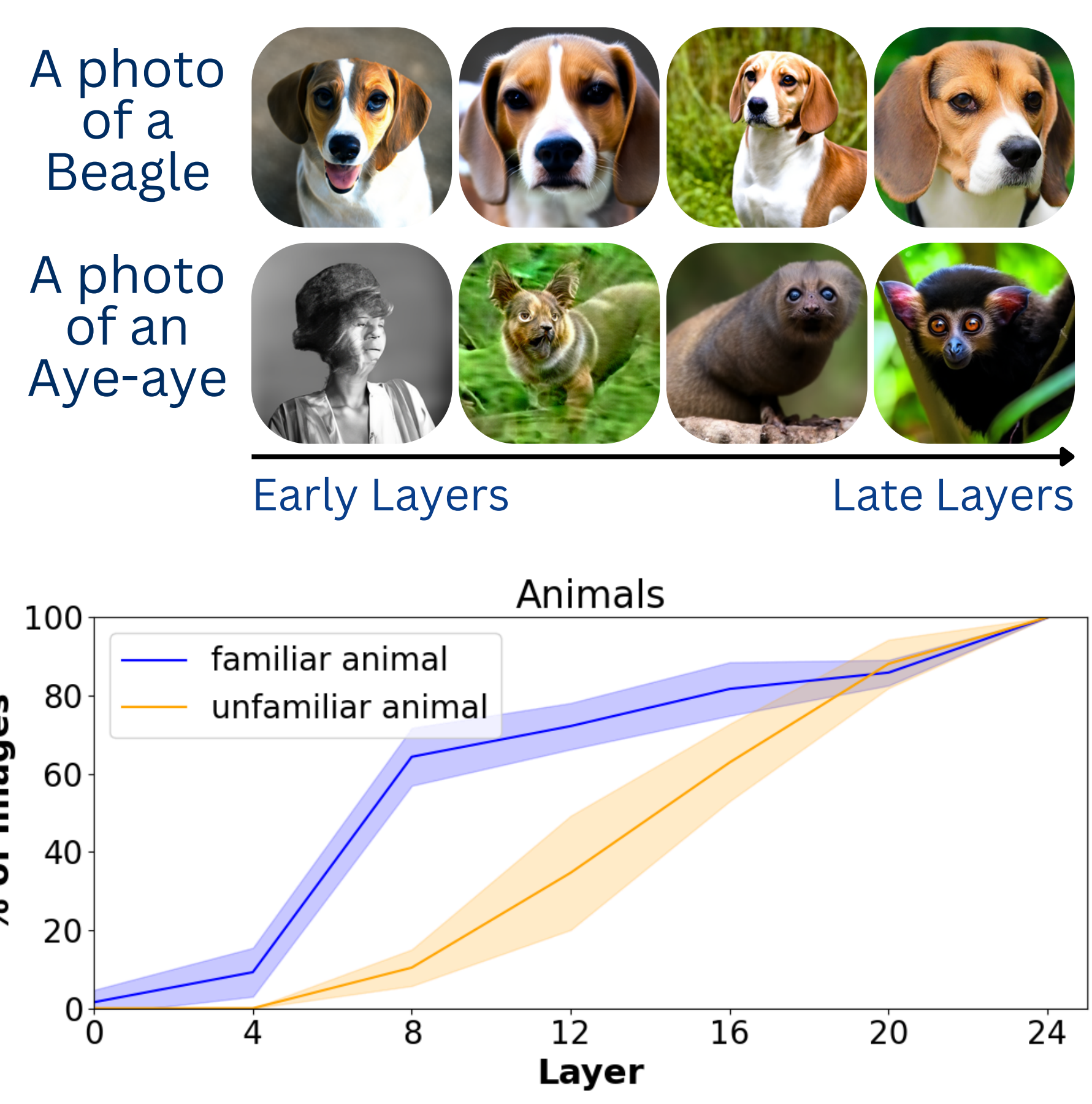


Insights

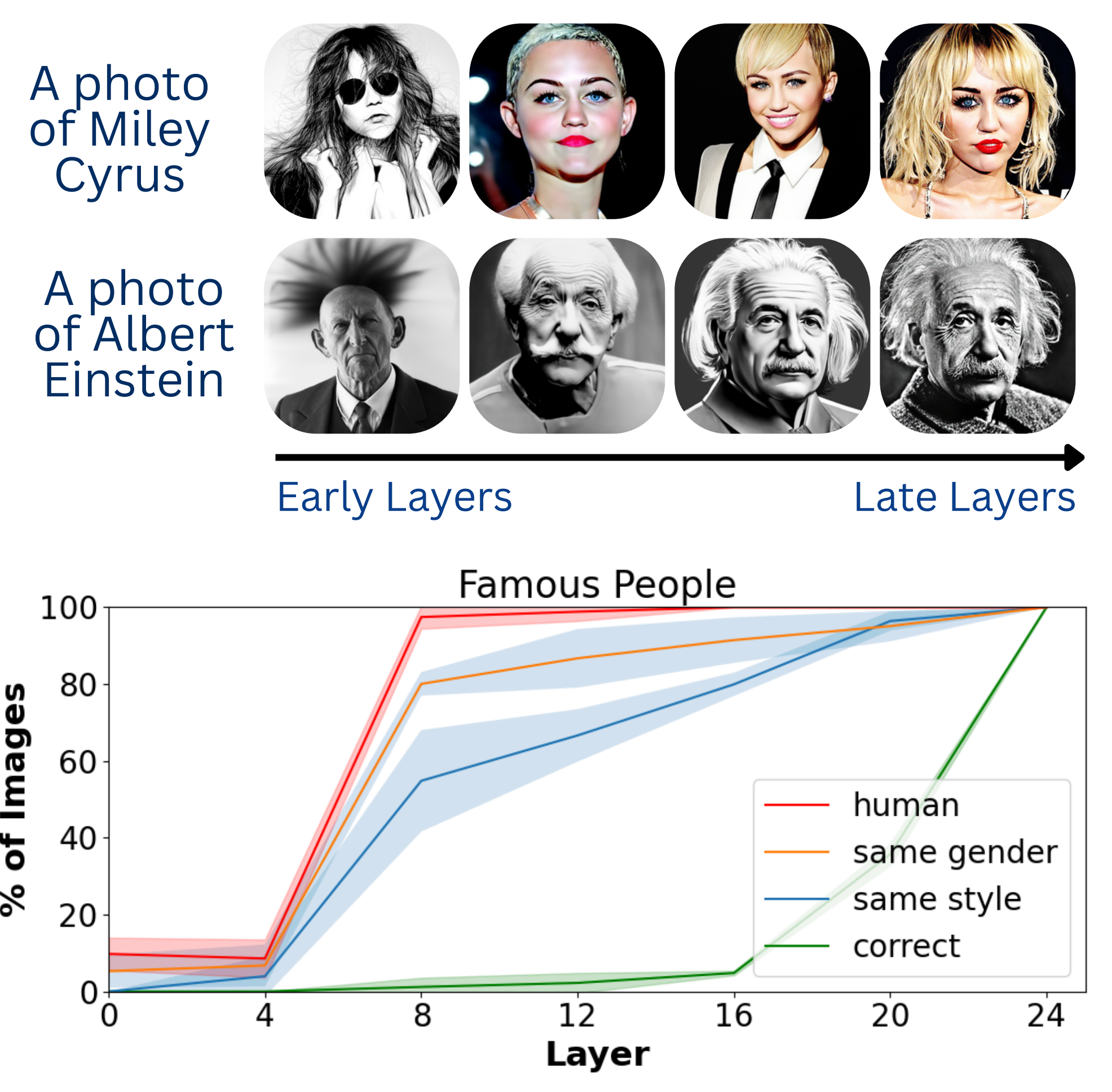
1 Complex prompts emerge later than simpler prompts



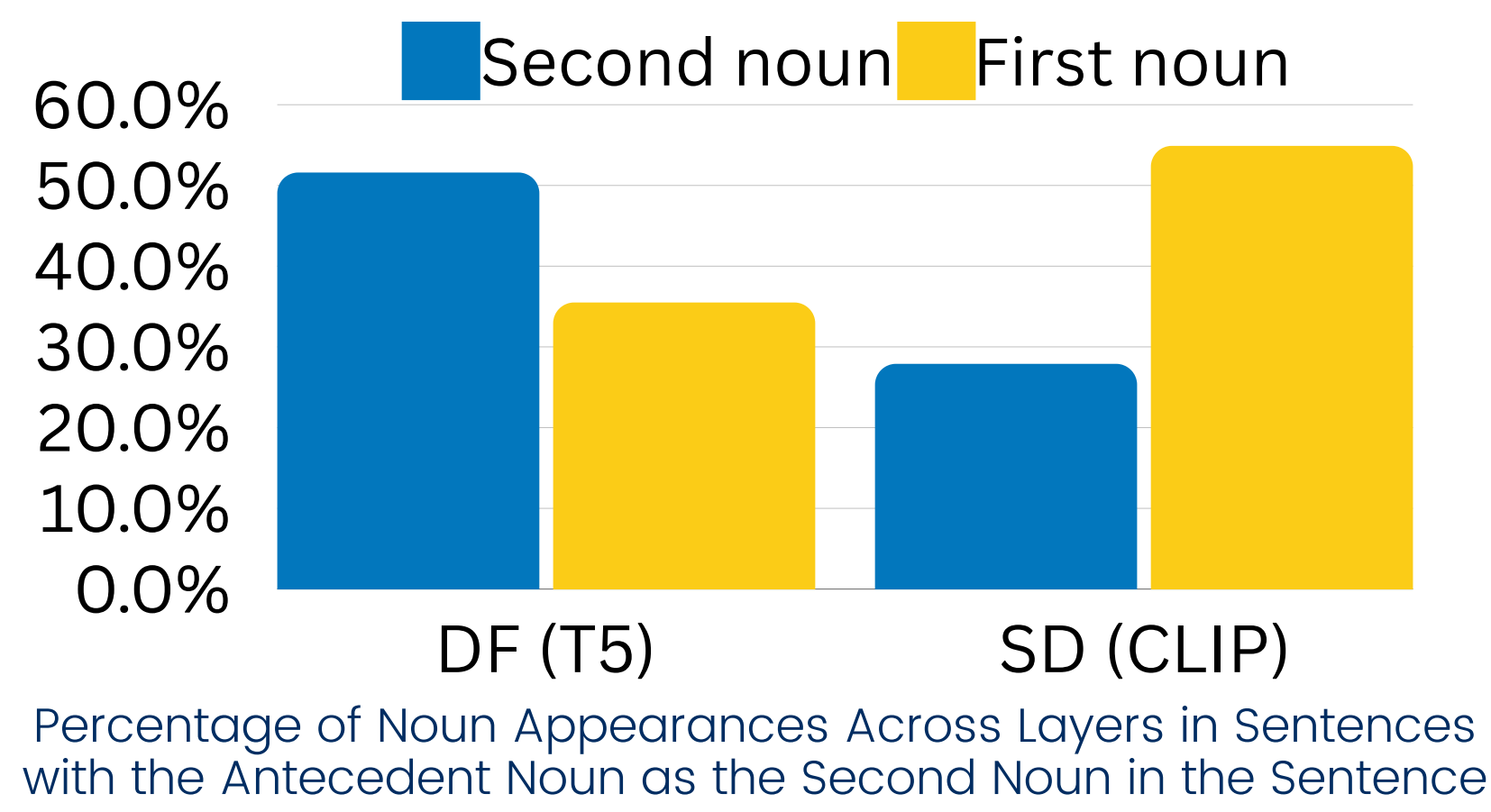
2 Common concepts emerge earlier



3 Gradual refinement



4 Text encoders differ in computation process



A few people are in the ocean splashing around



A photo of a Jerboa



Check out the paper for additional insights