



home
papers
study
life
About me

Chameleon

- Chameleon: Mixed-Modal Early-Fusion Foundation Models
- <https://arxiv.org/abs/2405.09818>
-

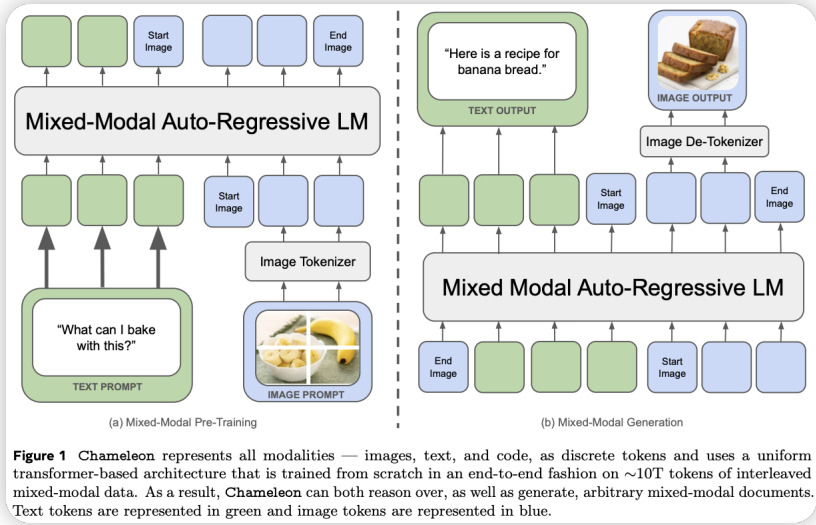


Figure 1 Chameleon represents all modalities — images, text, and code, as discrete tokens and uses a uniform transformer-based architecture that is trained from scratch in an end-to-end fashion on ~10T tokens of interleaved mixed-modal data. As a result, Chameleon can both reason over, as well as generate, arbitrary mixed-modal documents. Text tokens are represented in green and image tokens are represented in blue.

Image Tokenizer token

Newer

Older

2025-05-08 Flamingo	2025-05-07 Emu
------------------------	-------------------