

# Meta's Image Generator That Plans, Drafts, Critiques, and Refines Like a Human Artist

Kabui, Charles

2026-04-22

---

[Read at ToKnow.ai](#)

---

**Meta's Image Generator That Plans, Drafts, and Refines Like an Artist**

Process-driven generation with interleaved reasoning

- 4 Stages**  
Plan, draft, critique, refine per iteration cycle
- Dense**  
Step-level supervision on every intermediate state
- Self-Fix**  
No external evaluator needed  
Self-sampled critique traces

April 22, 2026

ToKnow.ai

Researchers at [Meta AI \(FAIR\)](#) have built an image generation system that works the way a human painter does: plan what to draw, sketch a draft, look at what's wrong, then fix it. Instead of producing an image in a single pass, the model follows an interleaved reasoning trajectory across four stages per iteration: textual planning (deciding what to draw next), visual drafting (generating the current state), textual reflection (critiquing the draft against the

prompt), and visual refinement (correcting what the critique identified). Each stage receives dense step-level supervision, meaning the model is trained not just on final outputs but on every intermediate state. Self-sampled critique traces teach the model to spot and fix its own errors without needing a separate evaluator. On standard text-to-image benchmarks, this approach significantly outperforms single-shot generation in compositional accuracy, the ability to correctly arrange multiple objects with specific attributes and spatial relationships.

Complex prompts routinely break single-shot generators. Ask for “a red teapot on a wooden table next to a blue vase with sunflowers” and something will be wrong: the teapot turns blue, the vase disappears, or the flowers end up in the teapot. Process-driven generation gives the model a chance to notice and fix those failures before finalizing. The approach is architecture-agnostic, so it can layer on top of existing generators.

This mirrors what happened in language models when chain-of-thought prompting replaced single-pass answers. The same principle, let the model reason step by step, now applies to pixels. For an earlier example of search-augmented image generation, see [Gen-Searcher](#).

Sources:

- [Process-Driven Image Generation \(arXiv\)](#)
- [Meta AI Research \(FAIR\)](#)
- [Chain-of-Thought Prompting \(Wei et al., 2022\)](#)

---

*Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)*