Article

Supplement 1: LLM-Guided Matryoshka Generation 1

Yulia Kumar[1,3,*] 0000−0002−7621−2734, **Srotriyo Sengupta**[2]

[1] Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ; yulia.kumar@rutgers.edu;
[2] Princeton University ss0352@princeton.edu Srotriyo Sengupta;
[3] Department of Computer Science and Technology, Kean University, Union, NJ; yulia.kumar@kean.edu
* Correspondence: yulia.kumar@kean.edu; (Y.K.)

1. LLM-Guided Matryoshka Generation

This study evaluates whether contemporary large language models (LLMs) and their associated image generators can synthesize Matryoshka imagery that is both visually plausible and class-consistent. To establish a controlled generative baseline, a two-stage prompting pipeline is adopted: a multimodal LLM produces structured, class-aware prompts, and a separate image model renders the corresponding images. In the reported configuration, Gemini-PRO generates prompts that are then executed by ChatGPT-4o, with one prompt template instantiated for each of the eight Matryoshka classes. This design improves reproducibility and makes it possible to audit how well the class taxonomy is internalized at the prompting stage. An example prompt used for the *Artistic* class is shown below.

> *"Please generate an image of a Matryoshka doll. Focus on an artistic type, emphasizing intricate, hand-painted details and unique thematic elements that showcase the craftsmanship beyond typical folk art. The doll should reflect a sophisticated design, potentially drawing from specific historical art movements or contemporary artistic interpretations while retaining the traditional nesting doll form."*

The prompt explicitly specifies stylistic constraints (detail level, ornament density, and optional art-historical influences) that act as high-level controls over palette, visual complexity, and mood. The template is further adjusted to produce coherent families of synthetic dolls per class, targeting consistency in pose, facial structure, and decorative grammar while preserving class-specific variability.

Zero-shot generation is also evaluated using the Google Nano-Banana-Pro model (2), issuing single-turn prompts without conversational context. Its outputs differ qualitatively from those produced by ChatGPT-4o and Gemini-PRO: in several cases, prompts intended for *non- Matryoshkas* produce plausible Matryoshka families, and political or religious themes are rendered in unpredictable ways. These failure modes indicate that current generators often preserve global "Matryoshka-ness" while weakening class boundaries. Despite these limitations, the results demonstrate substantial generative capacity and motivate future work on prompt alignment and controllable synthesis for cultural-heritage imagery.

Although synthetic generation is not the primary objective, it provides a controlled mechanism for distribution-shift experiments. A long-term direction is to blend real and synthetic Matryoshkas into a mixed dataset and evaluate how 2D, 3D, and multimodal backbones transfer to, or detect, synthetic counterparts. This enables stress-testing of

Received:
Revised:
Accepted:
Published:

https://doi.org/10.3390/1010000

Figure 1. Synthetic examples for 8 classes: ChatGPT-4o (top), Gemini-2-PRO (mid), and Nano-Banana-PRO (bottom).

authenticity cues and robustness under systematic perturbations in motif, style, and composition.

Lessons Learned: LLM-guided prompting enables reproducible, class-conditioned synthesis and supports rapid qualitative auditing across the full taxonomy. However, current generators frequently blur semantic boundaries between classes and do not reliably reproduce geometric and construction cues required for RU vs. non-RU authentication. Consequently, synthetic Matryoshka generation is treated primarily as a controlled stress-test source rather than a direct augmentation mechanism for authenticity modeling.

After several unsuccessful attempts with alternative generators and prompting strategies, SDXL–Juggernaut is adopted to produce Matryoshka images (1). Figure 2 shows a representative grid for the *Artistic* category. The samples exhibit high photorealism, consistent lighting, and moderate intra-class diversity. Nevertheless, the outputs remain insufficient for correcting the class imbalance of the authentic dataset: pose and composition diversity are limited, subtle generation artifacts persist, and style averaging reduces the presence of distinctive workshop-specific features. Overall, the SDXL–Juggernaut pipeline, combined with grid visualization, enables rapid human-in-the-loop screening and clarifies where additional prompt constraints, guidance tuning, or targeted augmentation is required to better match the statistics of the real Matryoshka dataset.

Lessons Learned: SDXL–Juggernaut produces visually convincing Matryoshka imagery and supports efficient quality control via grid-based inspection, but it does not reliably generate the fine-grained construction and geometry signals needed for authenticity inference. In the current setting, the primary value is diagnostic: the synthetic grids expose which attributes are easily controlled (style, palette, motif density) and which remain underconstrained (pose variation, workshop-specific structure, and class-separating semantics).

References

1. RunDiffusion, Juggernaut XL v9 + RunDiffusion Photo v2 Official,” *Hugging Face*, 2024. Available: https://huggingface.co/RunDiffusion/Juggernaut-XL-v9. Accessed: 6 Dec. 2025.
2. K. Chedraoui, *Nano Banana Pro Is the Best AI Image Tool I've Tested. It's Also Deeply Troubling*. CNET, Dec. 4, 2025. [Online]. Available: https://www.cnet.com/tech/services-and-software/google-nano-banana-pro-ai-image-generator-review/. [Accessed: Dec. 2025].

Figure 2. Matryoshkas generated via SDXL–Juggernaut.