Alien Recombination: Exploring Concept Blends Beyond Human Cognitive Availability in Visual Art

Alejandro Hernandez[1], Levin Brinkmann[1], Ignacio Serna[1], Nasim Rahaman[2], Hassan Abu Alhaija[3], Hiromu Yakura[1], Mar Canet Sola[1,4], Bernhard Schölkopf[2], and Iyad Rahwan[1]

[1]Max Planck Institute for Human Development, Berlin, Germany.
[2]Max Planck Institute for Intelligent Systems, Tübingen, Germany.
[3]NVIDIA.
[4]BFM, Tallinn University, Estonia.

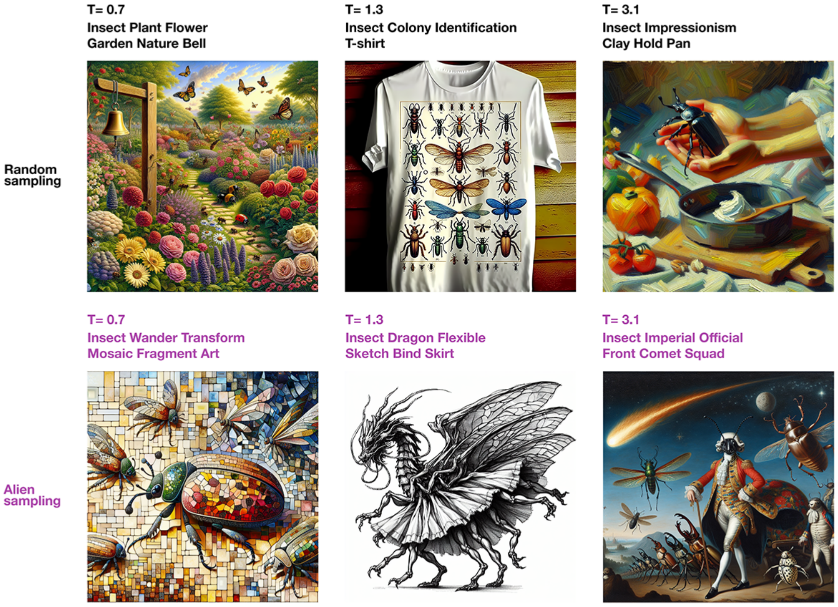


Figure 1: Comparison of images generated using random sampling versus Alien sampling with the Art model for the input sequence “Insect”. The Art model, fine-tuned on WikiArt concepts [1], combines likely concepts to produce artworks, with novelty controlled by temperature. Random sampling selects concept combinations arbitrarily, while Alien sampling ranks combinations based on cognitive unavailability and artistic fit, generating images from top-ranked sequences. For the full Alien Recombination method, see Figure 4.

1 Introduction

There is ongoing debate about whether generative AI models can genuinely be considered creative and capable of producing original cultural artifacts. Some AI systems, however, have already broken new intellectual ground. For example, AlphaGo [2] discovered Go strategies that were previously overlooked or unimaginable and have since been learned, adopted, and expanded by human players [3]. Yet, for open-ended domains like art, the question remains: to what extent can these models unlock truly novel and valuable connections that no human mind has conceived?

We hypothesize that in visual art, a vast, unexplored space of concept combinations exists, not due to inherent incompatibility, but because of the limits of the individual artist cultural horizon, including their geographical, temporal, and social embedding. For instance, the concept of “airplane” did not exist during the Renaissance period, making it cognitively unavailable to artists of that era.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Today, despite familiarity with both concepts, there remains a bias against combining Renaissance style and airplanes. Cognitive science has intensively discussed human availability bias, heavily relying on immediate examples that come to mind when evaluating a specific topic [4], thereby potentially constraining exploration of novel ideas. Although trained across temporal and cultural boundaries, generative AI models inevitably absorb human biases. As a result, these models can expect to predominantly produce cultural artifacts that align with human cognitive availability. James Evans and his team showed that counteracting such a bias could be key to algorithmic augmented scientific discovery [5, 6]. Inspired by this work, we developed a system that generates novel visual art concept combinations by modeling and counteracting cognitive biases. It produces combinations not previously attempted in our dataset and cognitively unavailable to any artist in the domain.

2 Method

The Alien Recombination method employs two large language models (LLMs) to generate and rank novel combinations of artistic concepts. To build our concept space, we first extracted nine semantic-level features for each image in the WikiArt dataset [1] using CLIP [7], selecting the words most similar to the image in CLIP's embedding space. This number was determined through an ablation study showing diminishing returns in concept accuracy beyond nine features. To capture artistic style information, we then added the artwork's style as a tenth concept from metadata, resulting in a comprehensive representation of each artwork.

To facilitate the combinatorial nature of our approach, we constrained concepts to the WordNet Core [8], a curated list of essential English words. This decision reduces the concept space size, preventing CLIP from extracting multiple words referring to the same concept and increasing concept overlap between artworks. Although this may limit the proper representation of some niche concepts, it enables us to focus purely on novel concept combinations rather than vocabulary variations.

Using these concepts, we created two complementary text datasets: the Art dataset, which contains randomly permuted concept lists for each image, and the Cognitive Availability dataset, created by sampling and aggregating concepts associated with individual artists to reflect their total range of ideas. Let $C = \{c_1, c_2, \ldots, c_n\}$ be the set of all possible concepts from WordNet Core. From these datasets, we respectively estimate two key probability distributions: the artwork-level distribution $P_{\text{art}}(c^i|c^0, ..., c^{i-1})$, representing concept co-occurrence within artworks, and the artist-level distribution $P_{\text{cog}}(c^i|c^0, \ldots, c^{i-1})$, capturing cognitive availability of concepts to artists, where $c^i \in C \setminus \{c^0, \ldots, c^{i-1}\}$ for both distributions.

We then fine-tuned two GPT-2 models [9] on their respective namesake datasets: the Art model and the Cognitive Availability model. Novel combinations are generated through a two-step process. First, We generate sequences with the Art model, keeping only those with all concepts in WordNet Core. This constraint not only enables later reliable evaluation of combinations within the source data but also ensures the problem remains fully recombinational. Second, we rank these sequences based on perplexity scores from both models. Given that both models accurately represent their target distributions, the perplexity scores provide an approximation of human cognitive biases by capturing which concepts are not usually seen together in artworks or in an artist's full concept usage. Specifically, high perplexity in the Art model indicates concept combinations that rarely co-occur in existing artworks, while high perplexity in the Cognitive Availability model suggests combinations that do not typically appear within an artist's entire body of work, that is, when artists use some concepts in the sequence, they rarely use the others in any of their artworks. In turn, a lower rank in the Art model suggests higher probability of that combination appearing in existing artworks, while a lower rank in the Cognitive Availability model indicates higher cognitive accessibility.

By inverting the Cognitive Availability rank and using Weighted Rank Aggregation with parameter β, we obtain concepts that could form coherent artworks while being cognitively unexpected. Higher β values emphasize cognitive unavailability, which we term "alienness" as in [5]. The top-k ranked sequences from the combined ranking are then returned. We refer to this stage of the method, which involves ranking and selection of sequences, as Alien Sampling. Finally, because the artistic novelty of plain text combinations of concepts can be difficult to evaluate directly, we assess how well these combinations fit in an artwork by visualizing the returned sequences. This is done using a text-to-image model (DALL-E [10]), with prompts structured as: `A painting that contains the concepts: <input sequence + generated sequence>`. If a particular style is specified in the sequence, we modify the prompt to reflect that style.

3 Experiments and results

We compared the Alien Recombination method, using various values of β for Alien sampling, to a baseline method to generate novel images. The Baseline method generates images using the Art model but replaces Alien sampling with random sampling.

In our experiment, we created 50 unique input sequences, each containing either 1 or 2 concepts. For each input sequence, we generated 150 output sequences at each temperature level, ranging from 0.1 to 3.1 in increments of 0.3, for both the Alien Recombination and Baseline method. From the generated sequences, we then selected the top-ranked sequence for each method based on its respective sampling strategy.

Our experimental design addresses the complex task of assessing artistic novelty through both text-based and image-based evaluation approaches.

3.1 Text-based novelty

Our methodology evaluates the novelty of concept combinations using two complementary measures.

- **Novelty relative to artworks:** Let $A = \{A_1, A_2, \ldots, A_m\}$ be the set of sets, where each $A_i \subset C$ represents the set of concepts in a single artwork within the Art dataset. Let $S \subset C$ be the set of concepts in the generated sequence. We define the novelty measure N_{art} as:

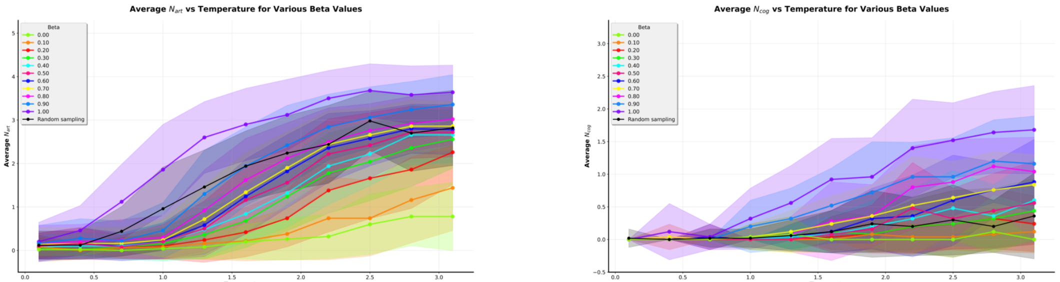
$$N_{\text{art}} = \min_{i \in \{1,\ldots,m\}} |S \setminus A_i| \tag{1}$$

  This measure represents the minimum number of concepts in S that do not appear together in any single artwork, when compared individually to each artwork in the dataset. A more intuitive way to interpret this measure is that it quantifies the number of distinct concepts in the generated artwork compared to the most similar artwork in the dataset.

- **Novelty relative to cognitive availability:** Let $B = \{B_1, B_2, \ldots, B_k\}$ be the set of sets, where each $B_j \subset C$ represents the set of all unique concepts used by artist j in the Art dataset. We define the novelty measure N_{cog} as:

$$N_{\text{cog}} = \min_{j \in \{1,\ldots,k\}} |S \setminus B_j| \tag{2}$$

  This measure represents the minimum number of concepts in S that do not appear together in any single artist's cognitive framework, when compared individually to each artist's set of concepts. Thus, N_{cog} quantifies the number of distinct concepts in the generated artwork compared to the most similar artist set.



(a) Average N_{art} vs temperature for multiple values of β

(b) Average N_{cog} vs temperature for multiple values of β

Figure 2: Comparison on the novelty of the generated sequences with respect to the artworks and the cognitive availability for multiple values of β

Our findings reveal that while increasing the Art model's temperature can generate novel combinations absent from the dataset at artwork-level (Figure 2a), it does not reliably produce cognitively unavailable combinations. The Alien Recombination method, through its explicit search for cognitive unavailability, demonstrates a consistently higher likelihood of generating such combinations (Figure 2b). Empirically, we show that generating unseen combinations in artworks when increasing the temperature is relatively straightforward, with 85% of the combinations containing at least one new concept when surpassing temperature 1. However, this phenomenon does not extend to cognitively unavailable combinations: at temperature 3, more than half of the combinations have $N_{\text{cog}} = 0$, indicating no cognitively unavailable concepts. This reveals that finding completely cognitively unavailable combinations is a fundamentally much harder task, requiring explicit search strategies within the concept space. Further analysis and details are provided in the Appendix.

3.2 Image-based Novelty

For both the Baseline method and the Alien Recombination method with $\beta = 0.85$, we generated sequences and converted the top-ranked sequences into images for evaluation. Image novelty was then evaluated using two independent approaches:

- **Evaluation using GPT-4:** We used GPT-4 [11] to perform pairwise comparisons between images produced by the Alien Recombination and Baseline methods. Based on GPT-4's alignment with human evaluators in vision-language tasks [12], we asked it to evaluate concept combination novelty. For each pair of images created with the same input and temperature, GPT-4 was prompted with the following request: `As an art expert, please write a sentence indicating which image is more novel, focusing on concept combination novelty.`
- **Embedding-based Analysis:** Following the approach of [13], we computed ResNet152 [14] embeddings for both the generated images and those in the WikiArt dataset. Then, we assessed image novelty in two ways: (1) For each generated image, we computed its maximum cosine similarity with any WikiArt image, where lower maximum similarity indicates higher novelty. We then compared the average of these maximum similarities between methods. (2) As before, we performed pairwise comparisons between images generated by both methods, considering the image with lower similarity to be more novel.

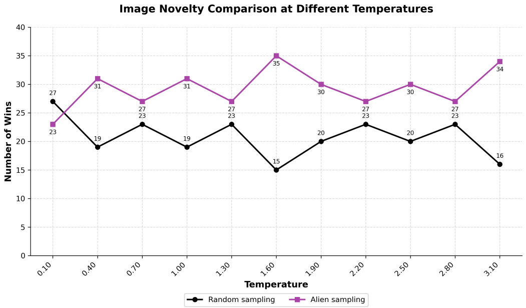


Figure 3: GPT-4 pairwise evaluation

GPT-4 evaluations consistently rated Alien Recombination images as more novel on average, suggesting that cognitively unavailable combinations are perceived as more innovative (Figure 3). The embedding analysis corroborated these findings, showing lower average similarity scores and more favorable pairwise comparisons for the Alien Recombination method (Figure 11).

However, it is important to note that ResNet embeddings also account for visual factors such as lighting and contrast, making them sensitive to elements beyond just conceptual differences. In contrast, GPT-4 can be explicitly prompted to evaluate novelty based on conceptual combinations. Additionally, since DALL-E generates in-distribution images, its outputs may be constrained by familiar patterns and styles, potentially limiting its ability to fully capture the novelty of the prompt. Further details can be found in the Appendix.

4 Conclusion

We present the Alien Recombination method, designed to generate novel artistic combinations within the space of visual art. This system not only produces combinations that have never been attempted before within our dataset but also identifies and generates combinations that are cognitively unavailable to all artists in the domain. Additionally, our results suggest that cognitive unavailability is a promising metric for optimizing artistic novelty, outperforming mere temperature scaling. This approach uses generative models to connect previously unconnected ideas, providing new insight into the potential of framing AI-driven creativity as a combinatorial problem [15, 16].

References

[1] Babak Saleh and Ahmed Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. *arXiv preprint*, 1505.00855:1–21, 2015.

[2] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[3] Minkyu Shin, Jin Kim, Bas van Opheusden, and Thomas L. Griffiths. Superhuman artificial intelligence can improve human decision-making by increasing novelty. *Proceedings of the National Academy of Sciences*, 120(12):e2214840120, March 2023.

[4] Amos Tversky and Daniel Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.

[5] Jamshid Sourati and James A Evans. Accelerating science with human-aware artificial intelligence. *Nature human behaviour*, 7(10):1682–1696, 2023.

[6] Feng Shi and James Evans. Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, 14(1):1641, 2023.

[7] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

[8] Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson, and Robert Schapire. Adding dense, weighted connections to wordnet. In *Proceedings of the International WordNet Conference*, pages 29–36. KAIST, 2006.

[9] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners, 2019. OpenAI blog: `https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf`.

[10] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint*, 2204.06125:1–27, 2022.

[11] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. GPT-4 technical report. *arXiv preprint*, 2303.08774:1–100, 2023.

[12] Xinlu Zhang, Yujie Lu, Weizhi Wang, An Yan, Jun Yan, Lianke Qin, Heng Wang, Xifeng Yan, William Yang Wang, and Linda Ruth Petzold. GPT-4V(ision) as a generalist evaluator for vision-language tasks. *arXiv preprint*, 2311.01361:1–31, 2023.

[13] Ahmed Elgammal, Bingchen Liu, Diana Kim, Mohamed Elhoseiny, and Marian Mazzone. The shape of art history in the eyes of the machine. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[15] Paul Thagard and Terrence C Stewart. The aha! experience: Creativity through emergent binding in neural networks. *Cognitive science*, 35(1):1–33, 2011.

[16] John S Gero. Creativity, emergence and evolution in design. *Knowledge-Based Systems*, 9(7):435–448, 1996.