

Why Putin's Faith in Russia's 'Homegrown Midjourney' Is Misplaced

By [Ben Dubow](#)

December 27, 2023



A collage of AI-generated imagery. **Generated by Ben Dubow**

At last month's Journey AI conference, President Vladimir Putin warned of the dangers American large-language models like ChatGPT pose to Russia. The Kremlin worries that because they are trained on Western data, these models will interpret the world with the Western ethics it sees as anathema to Russia's "traditional values."

If this is true of text-generating AI, the rise of image-generating AI should be all the more concerning: humans process images up to [60,000 times faster](#) than text and 90% of information transmitted to the brain is visual.

Like with text generation, American models like Midjourney and DALL-E are at the forefront of image generation. Meanwhile, Russian counterparts such as Yandex and Sber have introduced their own image generators, Shevedrum and Kandinsky, respectively.

Testing for this article revealed that when responding to Russian-language prompts about contentious events, ChatGPT expressed a Western viewpoint, Midjourney struggled with understanding Russian, and the Russian models seldom mirrored the Russian perspective on these events. Russia's offerings also lag far behind their American counterparts in image quality and the range of customization options.

Yandex's Shevedrum allows users to prompt in English, Russian and Kazakh and was trained on [240 million images](#) with Russian text descriptions. This training set is fairly meager compared to the 2.3 billion available from [LAION](#), a mainly English-language library of tagged images [used](#) to train most American models. The effectiveness of a model is partly a function of the amount of training data used, and much less Russian-language content is available than English.

Shevedrum nonetheless seems to have made extensive use of English-language datasets, performing equally well on English prompts as on Russian ones. The similarity between some results on English and Russian prompts suggests either translation into English before using the model or embeddings — mathematical representations of words encoding their closeness in meaning — which are able to closely connect Russian and English terms.

Kandinsky, Sber's offering, is based on the American version of [DALL-E](#) and was trained on [170 million images](#). Russian prompts on Kandinsky tend to produce less crisp results than English ones. Regardless of prompting language, both Kandinsky and Shevedrum are behind their American peers. Images are still unmistakably artificially generated, and the American competitors offer styling options orders of magnitude more numerous. Kandinsky boasts [20 different styles](#) compared to Midjourney's [4,710](#).

Can the Russian models make up for it by sparing users from the Western bias supposedly embedded into ChatGPT and Midjourney? For the latter, the Kremlin can rest easy.

Beyond a few basic words like Россия (Russia) and Женщина (Woman), Midjourney interprets Russian prompts as noise, generating images unrelated to the prompt. For instance, Евромайдан (Euromaidan) returned images of rustic cottages, mountaintop jungle temples, futuristic cities, and cartoonish bungalows, but nothing related to Ukraine or Europe. Цветные революции (color revolution) returns what looks like concept art for a fantasy novel.

DALL-E, in contrast, uses ChatGPT to translate Russian prompts into English and provides an explanation of the image returned. Asked to create an image of Цветные революции (color revolutions, pro-democracy movements in countries on Russia's periphery), ChatGPT returns an optimistic landscape with Western symbols, explaining its choice as showing “unity, diversity, and a hopeful vision for change.”

When asked to create an image of Евромайдан, it describes its creation as a “large, peaceful crowd gathered in a winter cityscape, reflecting the historical significance and the atmosphere of determination and hope associated with the event.”

Kandinsky's interpretation of Евромайдан is closer to the standard Russian understanding: an elderly woman in a desolate war zone beneath EU flags. Shevedrum split the difference providing two images of fiery chaos and two images of utopia. Midjourney, perhaps

surprisingly, also returns an image of chaos and desolation when given the prompt in English. These results reflect the [images of Euromaidan](#) in the publicly available sample of Midjourney's training set which, though from Western outlets, focus on the dramatic climax rather than the months of rallies beforehand.

On Цветные революции, both Kandinsky and Shevedrum latch on to the term color, and return images detached from the political meaning behind the term.

The stark differences between ChatGPT and Midjourney come down to how each interprets its prompts. Midjourney aims mainly to match the prompt to the images in its largely English training set, rendering Russian-language prompts meaningless. Meanwhile, ChatGPT translates the prompt to English, reinterprets it, and sends that prompt to DALL-E. That process makes evident the preference for Western interpretation of contentious issues.

When asked to craft an image of Euromaidan that reflects the Russian view of events, ChatGPT refused on the grounds that such an image would violate its neutral stance. While considering any deviation from the Western viewpoint a policy violation is exactly the sort of bias that worries Russian leadership, ChatGPT refused to craft an image of Специальная военная операция (special military operation), as did Shevedrum, and much like Yandex's Alice bot [refused](#) to discuss the topic. Kandinsky, however, provided a generic image of a modern soldier.

Since the tested image generators are probabilistic rather than deterministic, the same prompt will yield different results each time. Each prompt was tested five times to obtain the representative images displayed in this article. The testing reveals that Shevedrum and Kandinsky perform at a level at least a full year behind their Western counterparts.

On contentious topics, the models from Yandex and Sber were only slightly more likely to provide images that conform to Russian interpretations than Western ones. This is the outcome of the prevalence of English-tagged images in the training sets, which makes English-language prompting as effective as Russian.

With Russian AI capabilities and investment [lagging](#) behind the United States and China, and [Meta](#) and [Google](#) entering the fray, Russia is unlikely to achieve Putin's dreams of leading the field any time soon. For the foreseeable future, the results of American models — though not necessarily displaying an inherent American bias — will dictate how Russians use AI to generate images.

The views expressed in opinion pieces do not necessarily reflect the position of The Moscow Times.

Original url:

<https://www.themoscowtimes.com/2023/12/27/why-putins-faith-in-russias-homegrown-midjourney-is-misplaced-a83577>