# AI Judges Photos

By Don Sutherland

Generative Artificial Intelligence (AI) has made headlines with the arrival of ChatGPT and Gemini (formerly Bard). AI is already performing a growing range of functions, although its performance remains uneven with select tasks. For example, ChatGPT was asked to describe Connecticut's worst tornado on record and listed a fictional tornado that never occurred. Nevertheless, AI-related performance ranging from summarizing information to generating images has improved markedly over a very short period of time making AI an increasingly attractive tool for a growing range of uses.

One AI tool, **RateMyPhoto** attempts to evaluate photos on the basis of their aesthetic appeal. It could hold the promise of complementing teachers' scoring in art classes, assisting students in developing their work, and aiding the professional assessment of photo submissions.

The site provides a score of 1-10 based on how it assesses humans would rate a photo and also ranks the photos in terms of percentile ($1^{st}$ to $99^{th}$) with $99^{th}$ being in the top 1% of photos overall.

At present, even as RateMyPhoto shows promise, it likely is not yet ready to take the place of professional judges or the broad population. Nevertheless, its performance lies somewhere between good and impressive (based on how it fared against actual human judging).

Human judging was based on the following:

- World Photography.Org's top photo from 2023
- *Time's* top photo from 2023
- The front page gallery in a very large photo site and the front gallery in a small photo site
- Selective photos from professional or iconic photographers (Eliot Elisofon, David Benjamin Sherry, Cath Simar, Denise Ippolito, Ansel Adams) were also scored by the system. All of their photos have been sold on the basis of their creative and aesthetic appeal or featured in galleries worldwide
- Solid black and white squares

Time's photo of the year scored 7.53 ($85^{th}$ percentile). World Photography.Org's top photo scored 7.11 ($75^{th}$ percentile). The top 10 photos in the large photo site's top 10 photos had an average score of 8.27 ($92^{nd}$ percentile) and while those in the top 10 on the small site averaged 7.28 ($74^{th}$ percentile).

As the small site's front page photo gallery is comprised of 18 photos, I also looked at the lowest score for each site's top 18 photos. Both sites had a single photo among the top 18 that had a score of 6.27 ($48^{th}$ percentile).

**Randomized Ranking from Front Page Galleries:**

**Large Site: 112 million members/60 million active users**

Top 10 Photos:

| Score | Percentile |
|-------|------------|
| 7.11 | 75% |
| 7.53 | 85% |
| 9.32 | 99% |
| 7.95 | 92% |
| 8.64 | 98% |
| 9.15 | 99% |
| 7.39 | 82% |
| 8.56 | 98% |
| 8.98 | 99% |
| 8.09 | 93% |

| Summary: | Score | Percentile |
|----------|-------|------------|
| Mean | 8.27 | 92% |
| Median | 8.33 | 96% |
| Std. Dev. | 0.77 | |
| Minimum | 7.11 | 75% |
| Maximum | 9.32 | 99% |

**Small Site: 7,500 members**

Top 10 Photos:

| Score | Percentile |
|-------|------------|
| 7.39 | 82% |
| 8.39 | 97% |
| 6.27 | 48% |
| 6.55 | 58% |
| 6.69 | 62% |
| 6.41 | 53% |
| 6.83 | 67% |
| 7.67 | 87% |
| 8.48 | 97% |
| 8.09 | 93% |

| Summary: | Score | Percentile |
|----------|-------|------------|
| Mean | 7.28 | 74% |
| Median | 7.11 | 75% |
| Std. Dev. | 0.84 | |
| Minimum | 6.27 | 48% |
| Maximum | 8.48 | 97% |

Ansel Adams' "Clearing Winter Storm, Yosemite National Park" scored 9.74 (99th percentile). In contrast, Eliot Elisofon's "Man fishing in waterfalls" featured at the Smithsonian Institute scored 7.53 (85th percentile).

For a further idea, I used a few higher quality photographs I had taken and scored them in the system. Two photographs are below:



SCORE: 9.57

**Compare to others:** Your picture has a higher rating than 99%% of all analyzed photos.

99%

SCORE: 7.25

**Compare to others:** Your picture has a higher rating than 79%% of all analyzed photos.

79%

I also ran some AI-generated images by the system and provided it with both a solid black and solid white square. Both squares scored 2.9 (7$^{th}$ percentile). I ran the same image at different times and the system consistently provided the same scores.

Based on my quick evaluation of the system, several points are in order:

- Bright and vivid colors and striking contrasts score high. Aesthetic soft or pastel images don't fare well in the system
- Images with cloudy, gray, or white skies don't fare well
- Out-of-season snowy scenes e.g., snow-covered blossoms, don't fare well.
- Underwater photos don't fare well
- Sharp and/or high-resolution photos do better than depth-of-field photos
- The system seems to be generous in providing high ratings to AI-generated images

In sum, at present, it appears that human artistic interpretation is more complex than what can currently be captured by AI. The large standard deviations among the top 10 photos in the two analyzed galleries illustrate the complexity of human artistic interpretation. Finally, it is worth noting that not every important photo e.g., documentary photos that cover critical subjects or events, are aesthetic. Those photos remain beyond the reach of the a system such as RateMyPhoto.