



## A practical implication of the Astolfo Effect: bias in AI generated images

Rodrigo Brincalepe Salvador<sup>1,2</sup>, Henrique Magalhães Soares<sup>3</sup> & João Vitor Tomotani<sup>3</sup>

<sup>1</sup>The Arctic University Museum of Norway, UiT – The Arctic University of Norway, Tromsø, Norway.

<sup>2</sup>Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, UiT – The Arctic University of Norway, Tromsø, Norway.

<sup>3</sup>Independent Researchers. São Paulo, SP, Brazil.

Emails: salvador.rodrigo.b@gmail.com; hemagso@gmail.com; tjvitor@gmail.com



**Figure 1.** The lovely paladorks. Fanart showing Roland (left inset), Astolfo, Charlie (a.k.a. Charlemagne), Brad-amante, and Merlin (right inset). Illustration by Rabi ([https://twitter.com/b\\_rabi](https://twitter.com/b_rabi); <https://www.pixiv.net/users/4959742>); used with permission.

The year of 2022 saw a huge advance in AI technology, especially Large Language Models, or LLMs. This culminated in the release of Chat GPT, an AI Chatbot assistant that, as of the time of this writing, is wowing the public with its uncanny performance.

However, chatbots are not the only application of LLMs. One such application is the artificial generation of images. Although such idea is not a novel one (it dates back to the 1970s; Elgammal, 2022), the advance-

ments on large language models allowed a new breakthrough in what these methods are able to achieve.

A non-obvious application of such models is as a “probe” for bias in its learning set. Since these models are trained on public datasets collected from the Internet, they tend to reflect the inherent biases present in human generated content. As such, we see the advent of AI generated image as an opportunity to further test the ‘Astolfo Effect’

hypothesis, as first outlined by Tomotani & Salvador (2021).

## WHAT IS THE ASTOLFO EFFECT?

We really encourage you to check out our original article and its follow-up (Tomotani & Salvador, 2021, 2022) to get the full story, but we will summarize the Astolfo Effect here.

It all starts with the *Fate* franchise, which began with the visual novel (and later anime) *Fate/stay night* in 2004 but became massively popular with the mobile and arcade versions of the game *Fate/Grand Order* (henceforth FGO; mobile version by De-lightworks, Lasengle, 2014–present, and arcade version by Sega AM2, 2018–present). In the *Fate* universe, mages can summon heroic spirits (known as ‘Servants’) to fight alongside them or, more usually, on their behalf. The servants are almost all taken from real-world material and can be historical people, legendary/mythological beings, or literary characters.

*Fate* is so popular and has such an amazing quantity of fanart that, typically, if you’re doing an internet search for a given character, you will get results including both “actual” real-world entries and *Fate*-related entries.

For the Astolfo Effect, we hypothesized that the most obscure characters (e.g., Astolfo and Bradamante; Fig. 1) present in FGO would, on a Google search, have more hits of their *Fate* incarnations than their original real-world ones. We further hypothesized that those FGO-related Google hits would appear sooner rather than later in the search. Conversely, widely popular characters (particularly in cinema and TV, such as Sherlock Holmes) would have fewer hits about their FGO incarnations and those

would appear later in the Google search.

We have shown that the Astolfo Effect is real and provided a list of the characters most affected by this (Tomotani & Salvador, 2021). Besides the ever-lovable Astolfo, that list included servants such as Nitocris, Sitonai, Yu Mei-ren, Li Shuwen, Mandricardo, Osakabehime, Scathach, and Yan Qing.

Given that the Astolfo Effect specifically pertains to internet searches, it is expected to play a role on AI-generated images, which fully depends upon images available online. However, in the original Astolfo Effect, we were interested in how fast internet searches got “flooded” with *Fate* material and the proportion of the first results that were from *Fate*. AI will use all available material, so we must make an adjustment and come up with a corollary: in this case, we expect that the AI-generated images affected by the Astolfo Effect will be those belonging to the characters with most fanart (e.g., Astolfo, Cu Chulainn).

## HOW ARE IMAGES GENERATED BY AI?

There is a lot of debate currently ongoing in the AI and Art community about the impact of this technology on the art scene, as well on the ethical and copyright implications of this technology.<sup>1</sup> As our goal is to seriously study the *Fate* phenomenon, we are not going to weigh in on this debate in this article.<sup>2</sup> Also, we will refrain from explanations that anthropomorphize AI models, such as stating that AI models “learn” a specific artist style by looking at a thousand images. They are not humans, and do not learn like humans.

The current state-of-the-art approach for generating images using AI are called diffusion models. They are a class of models that

<sup>1</sup> Here are some articles and viewpoints if you’re interested: Guadamuz (2017), Gillotte (2020), Cetinic & Che (2022), Elgammal (2022), Plunkett (2022). There are also some fun Last Week Tonight (HBO) episodes about it too.

<sup>2</sup> If you are interested on the perspective of each author on this topic, see the “About the authors” section at the end of this article.

are created by adding noise to an image, and by training a neural network to predict and remove this noise from an input image. By feeding a random noise input image to the model we can generate a new image by applying this “de-noise” procedure.

In a sense, these models learn a function that approximates the joint probability distribution of pixels in the images of its training set. With that function at hand, we can sample this distribution to generate new images. And by conditioning the distribution on the image prompt we can guide the generating process, creating images that correspond to a textual input given by the user.<sup>3</sup>

Given that these models approximate the distribution of its training dataset, they are, of course, vulnerable to bias present in those datasets – such as the bias introduced by the Astolfo Effect. The following section describes our approach to identify and quantify this bias.

## METHODOLOGY

To evaluate the bias introduced in the joint distributions learned by these kinds of Diffusion Models, we utilized “Stable Diffusion”, a model trained and open-sourced by Stability AI, an AI product and consulting company. This model was trained on the LAION Aesthetics dataset,<sup>4</sup> a subset of the LAION-5B dataset curated to contain images that humans find “aesthetically pleasing”. As the LAION-5B dataset contains over 5 billion images collected from the internet we conjectured that the Astolfo Effect bias would be present in it.

Setting up these models can be quite a hassle, as they have loads of dependencies and pre-requisites both on the hardware

(GPU acceleration) and on the software (OS'es, drivers and libraries) sides. Fortunately, there are containerized docker applications that take care of all that for us. In this article we utilized the “stable-diffusion-docker” image provided by GitHub user ‘fboulnois’.<sup>5</sup> This docker image allow us to prompt the model using a very simple command line script bundled with the docker image. For example, the following prompt produced the following image:

```
./build.sh run “An impressionist painting of a parakeet eating spaghetti in the desert”
```



Using this container, we generated images for a series of historical and mythological figures that are present in FGO, using the following command:

```
./build.sh run “{name}”
```

<sup>3</sup> This is a huge oversimplification. If you’re interested learning more about these models, we recommend the article by Allamar (2022).

<sup>4</sup> <https://laion.ai/blog/laion-aesthetics/>

<sup>5</sup> <https://github.com/fboulnois/stable-diffusion-docker> – run the official Stable Diffusion releases in a Docker container with txt2img, img2img, upscale4x, and inpaint.



where {name} is a placeholder for the actual name of the historical or mythological figure in question. So, for example, when generating images of “Arthur Pendragon” the command used was:

```
./build.sh run “Arthur Pendragon”
```

There were no attempts at “prompt engineering” such as prompting a specific art-style or specifying that we were interested in historical or mythological figures. Only the name of the figure was given to the model.

### Servants

As in Tomotani & Salvador (2021), we considered only the characters present in the mobile version of FGO (not the Fateverse as a whole), considering both the North American and Japanese servers in their state in July 2022 (last announced servants were Kyokutei Bakin and Minamoto no Tametomo). We used the North American spelling of the names (even though some are a bit “off”; see discussion in Tomotani & Salvador, 2021); for those characters that are yet to be released in North America, we used the names as given by the Fate Grand Order Wiki.<sup>6</sup>

We had to exclude some servants from our study, such as: repeated entries of the same character (e.g., Alter/Summer/Prototype/etc. versions); original characters (Mash, Emiya); and servants that were based on real-world material but in a way that makes them exclusive to the Fateverse (Hessian Lobo, Kijyo Koyo, Senji Muramasa).

## RESULTS

In total, we had 199 servants, and the full set of images can be found in the Supple-

mentary File to this article. The majority of the images produced was based on either (1) photographs or still images from films/TV series, or (2) “classic” art (mostly paintings, but also illustrations and sculptures). But many of the AI-generated images were affected by the Fateverse to different extents, from completely (e.g., Astolfo, Cu Chulainn) to tangentially (e.g., Euryale, Gilgamesh).

Here’s the full list of affected characters (see Fig. 2 for the images): Astolfo, Astraea, Bedivere, Bradamante, Cu Chulainn, Enkidu, Ereshkigal (not shown here because it was NSFW), Euryale (the purple hair gives it away), Gilgamesh, Ishtar, Iskandar (looks mostly like a classic artwork, but his anime eyes are priceless), Jing Ke, Kiyohime, Mordred, Nero Claudius, Osakabehime (shown in the typical pose of the official artwork, but what the hell happened here?), Scathach (also influenced by her incarnation in *Shin Megami Tensei* and *Persona* games), Semiramis, Sionai, Tamamo no Mae, and Yan Qing (also not shown due to NSFW reasons).

Most of the lesser-known servants from the results of Tomotani & Salvador (2021) that we mentioned earlier also popped up on the present list. But, as predicted, better-known characters (or characters with several other representations) also make the list if they are popular enough to have tons of fanart: Euryale, Gilgamesh, Ishtar, Iskandar, and Mordred.

### Other odd stuff

Curiously, it was also possible to see the influence of other pop culture icons on some results (Fig. 3). For starters, Baobhan Sith, because of the term “Sith”, was rendered as an evil-looking *Star Wars* character. Lu Bu and Sima Yi’s image were affected by their more famous incarnations from the *Dynasty Warriors* video game series. Astraea (Fig. 2), Habetrot, Hektor and the Valkyries were illustrated as cards from *Magic: The Gathering*. Eric Bloodaxe, the always-forgotten

<sup>6</sup> <https://gamepress.gg/grandorder/servant-availability>



Figure 2. AI generated images that have been affected to varying extents by the Astolfo Effect.





Figure 3. AI-generated images that have been affected by other things.

character in FGO, also appears as a card, potentially based on the game *Anachronism*. Chen Gong was part of an image that looks like a character/monster entry from a *Dungeons & Dragons* book, complete with depictions of the stats and text blocks. Nemo, of course, was shown as a clown fish. Siegfried appears as a bad Manowar-style heavy metal album cover art (perhaps based on the magician duo Siegfried & Roy).

Finally, Himiko, Izumo no Okuni, Minamoto no Tametomo, Ranmaru, Saito Hajime, Sakata Kintoki, Suzuka Gozen, and Watanabe no Tsuna were clearly based on manga/anime aesthetics, though not related to *Fate*. Saito Hajime's image, in particular, was influenced by his many other manga/anime incarnations, which are generally more popular than FGO's one.

## CONCLUSION

Salvador (2020) remarked that in all likelihood, many historians, archaeologists, and literary scholars must have at some point cursed *Fate* when their Google searches brought up a flood of anime results (sometimes NSFW!). Our results show that AI are also affected by this and, thus, we extended the known sphere of influence of the Astolfo Effect.

## REFERENCES

Allamar, J. (2022) The Illustrated Stable Dif-

fusion. Available from: <https://jalamar.github.io/illustrated-stable-diffusion/> (Date of access: 28/Feb/2023).

Cetinic, E. & Che, J. (2022) Understanding and creating art with AI: review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications* 18(2): 66.

Elgammal, A. (2022) AI is blurring the definition of artist. *American Scientist*. Available from: <https://www.americanscientist.org/article/ai-is-blurring-the-definition-of-artist> (Date of access: 01/Nov/2022).

Gillotte, J. (2020) Copyright infringement in AI-generated artworks. *UC Davis Law Review* 53(5): 2655-2691.

Guadamuz, A. (2017) Do androids dream of electric copyright? Comparative analysis of originality in Artificial Intelligence generated works. *Intellectual Property Quarterly* 2: 169-186.

Plunkett, L. (2022) AI creating 'art' is an ethical and copyright nightmare. *Kotaku*. Available from: <https://kotaku.com/ai-art-dall-e-midjourney-stable-diffusion-copyright-1849388060> (Date of access: 01/Nov/2022).

Salvador, R.B. (2020) Ancient Egyptian royalty in *Fate/Grand Order*. *Journal of Geek Studies* 7(2): 131-148.

Tomotani, J.V. & Salvador, R.B. (2021) The Astolfo Effect: the popularity of *Fate/Grand Order* characters in comparison to their real counterparts. *Journal of Geek Studies* 8(2): 59-69.

Tomotani, J.V. & Salvador, R.B. (2022) Testing the Astolfo Effect on newly-released servants in *Fate/Grand Order*. *Journal of Geek Studies* 9(2): 125-129.

### SUPPLEMENTARY FILE

Full set of AI-generated images created during the present study.

### ACKNOWLEDGEMENTS

We are extremely grateful to Rabi for letting us reproduce here the amazing artwork of chibi Charlie and his paladins.

### ABOUT THE AUTHORS

Dr **Rodrigo B. Salvador** was until recently a curator at the Museum of New Zealand Te Papa Tongarewa and now is a researcher in the Arctic University Museum of Norway. He believes the world would be a better place without so many tech bros. He is notably unlucky in summoning SSR servants, so he is seriously considering

visiting Aachen next year to use the Throne of Charlemagne as a catalyst.

**Henrique M. Soares** is an independent researcher from Brazil and fortunately was never hooked by the “gacha” demons. He believes AI generated images are not artwork, but a tool to further aid artists, and that safeguards must be put in place to protect the rights of artists to their creations (although he has no idea what such safeguards might be).

**João Tomotani**, MSc, is an engineer and Rin simp since the 2006 FSN anime, who seems to be turning into an Astolfo Effect specialist. Despite never having researched or delved into AI topics, the sheer amount of AI-generated FGO fanart being recommended to him has caught his attention. He successfully pulled Space Ishtar this year completing his Rin collection, and is now in the gigantic camp waiting for a new version of Ereshkigal (no, a costume is not enough).