EMERGING TECHNOLOGIES
IN PROPAGANDA AND DISINFORMATION

GLOBAL ENGAGEMENT CENTER (GEC)
TECHNOLOGY ENGAGEMENT DIVISION (TE)

March 27, 2023

Executive Summary

Text to Image Generative Artificial Intelligence (AI) Program Midjourney Released Its Version 5 Software, Dramatically Improving on Previous Versions: Previous GEC reporting highlighted text-to-image generative AI technology, this week we look at Midjourney version 5, showcase its astronomical improvements and possible impact to the global digital information environment.

Five images, each generated by a different version of Midjourney with the same prompt:
"Red haired woman wearing sunglasses standing with the Statue of Liberty in the background, photograph, 35mm film"

Midjourney v1 (March 2022)   **Midjourney v2 (April 2022)**

Midjourney v3 (July 2022)

Midjourney v4 (November 2022)

Midjourney v5 (March 2023)

Text-to-Image Generative AI: A Generation of Improvement In 18 Months

Text generating AI like ChatGPT is not the only type of AI improving seemingly at exponential scale. Midjourney's latest release represents a massive improvement in image quality, while reducing the level of expertise needed to make highly believable synthetic imagery. Midjourney version 5 has achieved photorealism, where its predecessor has a more cartoonish, overly-stylized output. Midjourney may have defeated the persistent "[Uncanny Valley](#)" effect, producing images that no longer look fake to the human eye.

I. **Why It Matters:** Midjourney allows anyone with a single line of text to create synthetic, but

highly realistic, photographs. Until recently, a convincing fake required time and artistic talent, now such images could become ubiquitous and difficult for the average user to detect.

Several images, generated in seconds, with simple prompts: "photoshoot of a delicious hamburger, food photography, 8k, realistic"

The above images received accusations of being photoshopped instead of generated, so the creator added otters:

II. **What is the Upside?** Generative AI has the potential to dramatically increase productivity of artists and creative workers. Adobe announced "Firefly", an AI-powered tool meant to augment

the ability of digital artists to utilize generative AI in their work. Computer graphics hardware maker NVIDIA AI [announced](#) "Foundations," a generative AI for enterprise customers. NVIDIA CEO Jensen Huang CEO [has predicted](#): "every single pixel will be generated soon. Not rendered: generated". Abode's platform could give rise to a new industry: "AI as a Service" (AIaaS), allowing third parties to access foundation models via an API. Simply put, companies can offer remote access to custom AI models customized to the end user's proprietary data and demands. OpenAI [has launched a similar platform](#) for ChatGPT in the mold of an App Store, allowing access to its generative AI technology via third-party plugin.

Early images from Adobe's "Firefly", showcasing an artistic style without the artist

III. **What is the Problem?** Midjourney allows creation of synthetic imagery with unprecedented ease and scale, with no preset limits on generating potentially misleading or harmful content. Previous text-to-image programs [had strict limits](#) on creating content that could be used in disinformation and propaganda campaigns. Midjourney is [attempting to ban misuse](#) of its platform, but with a full-time staff of only 11 people, addressing all misuse after a full public release may be too great of a challenge.
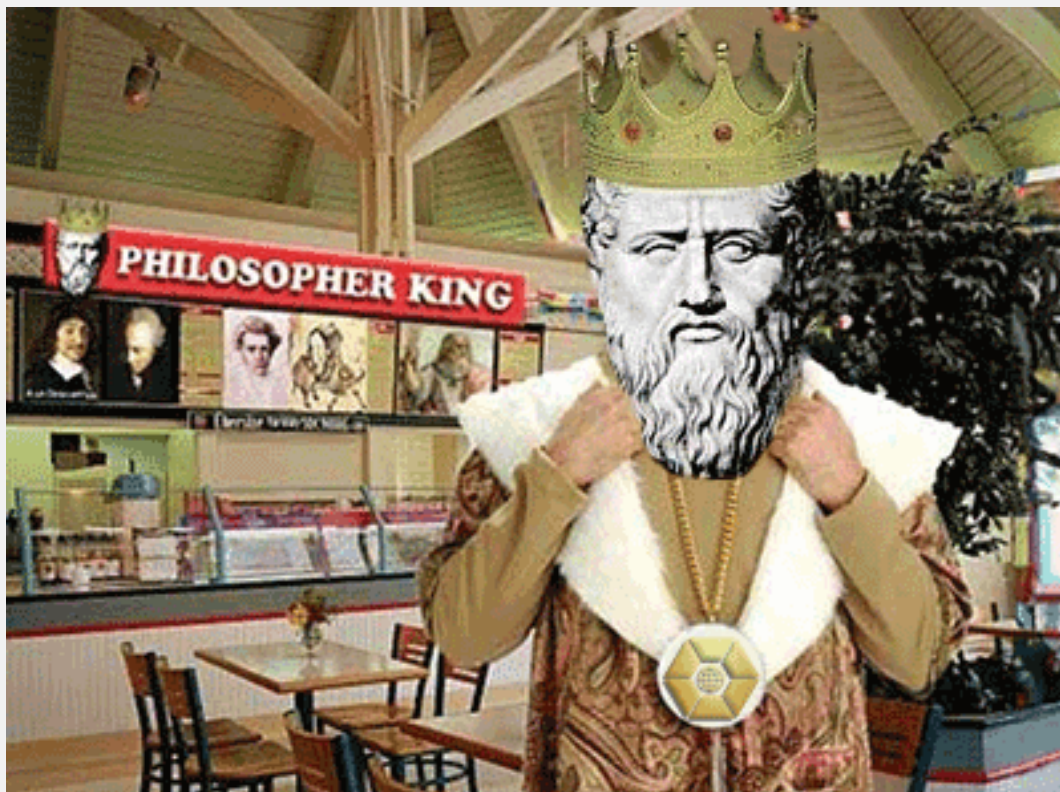
IV. What Can Be Done: [Empower solutions that can be built into platforms](#) for reverse image search and authenticity verification. Twitter's [current policy](#) toward generative AI images reads: "You may not share synthetic, manipulated, or out-of-context media that may deceive or confuse people and lead to harm ('misleading media')." Current synthetic image detection techniques for novel images created by generative AI [have proven to be easily circumvented,](#) and former "tells" of generative AI (such as "[it can't do hands](#)") [may no longer help](#).
   - Eliot Higgins, founder of [Bellingcat](#), with some tips [how to spot](#) a new generation of

generative AI deepfake images:

1. "It's apparent that it [Midjourney] often focuses on the first object described...with everything around it often having more flaws. Look outside of the image's focal point. Does the rest of the image appear to be an afterthought?"
2. Despite incredible progress, mistakes in the smaller details remain a common sign of fake images. As AI art grows in popularity, many artists point out that the algorithms still struggle to replicate the human body in a consistent, natural manner. It is possible that the algorithm might be able to avoid peculiar-looking body parts with more training and refinement.
3. "Look for odd writing on the walls, clothing, or other visible items. Messy rendering of text can differentiate fake images from real photos. If you look on text written on signs and walls, the words are often nonsensical."
4. Finally, a staple of AI- generated imagery seems to be over-the-top facial expressions. Higgins adds "I've also noticed that if you ask for expressions, Midjourney tends to render them in an exaggerated way, with skin creases from things like smiling being very pronounced".

A Brief Interlude from the GEC Philosopher King:

It's interesting how nobody thinks image generating AIs are conscious...that's some serious left-brain bias.

This week the GEC Philosopher King asks: The internet is already full of junk, we may find a point where we have [Kessler Syndrome](#) but for online content?

Is there a point where existing efforts to verify what is "true" or "false" break down when the internet is overwhelmed with synthetically generated content?

Recent technological advancements make it trivial to flood social media with convincing deep fakes and spam the comments section with realistic but fake opinions. What's left to help humans know and prove what is true?

Citations may no longer help because they rely on the existing knowledge graph (which may be overwhelmed by automated data pollution). What could remain is bespoke sources, verifiably associated with real individuals, holding up externally trusted, authentic information, surrounded by a sea of junk and pollution.

Additional Insights Provided by GEC Data Analytics Tool [NMA](#)

- A quick roundup of AI news:
  - Stanford University's [Alpaca 7B claims](#) "GPT-3.5 performance at an operation cost less than $600, running locally on a phone".
  - Google [released generative AI features](#) into its workspace products, including generating images, presentations, emails, docs and more.
  - Microsoft [announced 365 Copilot](#), embedding AI capabilities into Microsoft 365.

The United States Copyright Office declared: "[AI technologies can never be considered authors](...)

- [MIT discussed how](...) we can "fortify the truth", showing reality in context of deepfakes from generative AI.

- [Google announced](...) AI startups like Midjourney and AI21 Labs have chosen its cloud platform and announced a bunch of AI features, including for notetaking and email drafting.

- [Independent developers have launched](...) ChatLLaMA, a ChatGPT style chatbot based on Meta's LLaMA technology.

- [Research from Stanford shows](...) we can only accurately identify AI writers about 50% of the time.

- Chinese company Baidu [has revealed](...) a chatbot rival of ChatGPT. It can write poems in the style of the Tang Dynasty. Dismisses questions about Chinese President Xi Jinping, saying he has not "learned to answer them yet."

- This viral photo of the pope [was generated by Midjourney](...):

In this case, the old trick of zooming in on the hand worked, showing a generative mistake.

More About the GEC

GEC's Technology Engagement Division (GEC/TE) weekly reports inform the United States

government and international partners of future trends in counter propaganda and disinformation technologies, and potential countermeasures.

The mission of GEC/TE is to defend against foreign disinformation and propaganda by transitioning counter disinformation technologies from concept to application in support of smart policies and operations. As part of the mission, GEC/TE convenes technology experts and programmatic authorities from the public and private sectors. If you would like to be a part of these discussions, contact GECTech@state.gov.