Gender Dimensions in Generative AI for Image and Video Synthesis

Paraskevi (Vivi) Valergaki,
Anastasios (Tassos) Roussos
https://users.ics.forth.gr/~troussos

Institute of Computer Science (ICS),
Foundation for Research and Technology - Hellas (FORTH)

Heraklion, Crete, 27 February 2025

Introduction

A breakthrough moment in Generative AI

- On February 2024, OpenAI releases results of its text-to-video generator **SORA**:

Prompt: A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a bla

https://openai.com/sora

How did we get here?

Anastasios Roussos

Roots: Image-to-text & text-to-image generation

- Based on the big successes of Deep Learning in **both image analysis and text analysis,**
  - scientists soon started experimenting in **bridging them**, trying to combine image & text representations

- 2014: first AI models on **automatic image captioning** [1]
  - <u>Input</u>: image, <u>output</u>: text description of the image
  - impressive results with **captions that most of the times made sense**

- 2015: several scientists started working on the **inverse** problem: instead of image-to-text, **text-to-image generation** [2]
  - **even more challenging** problem
  - 2021: OpenAI released **DALL-E** text-to-image generation
  - 2022: **Midjourney**, which is widely used by **digital artists, Stable Diffusion,** the **first open-source image generator**
  - the current SOTA is based on **denoising diffusion models**

[1] Vinyals, O., Toshev, A., Bengio, S., Erhan, D. **Show and tell: A neural image caption generator**. CVPR 2015
[2] Mansimov, E., Parisotto, E., Ba, J.L., Salakhutdinov, R. **Generating images from captions with attention**. ArXiv 2015

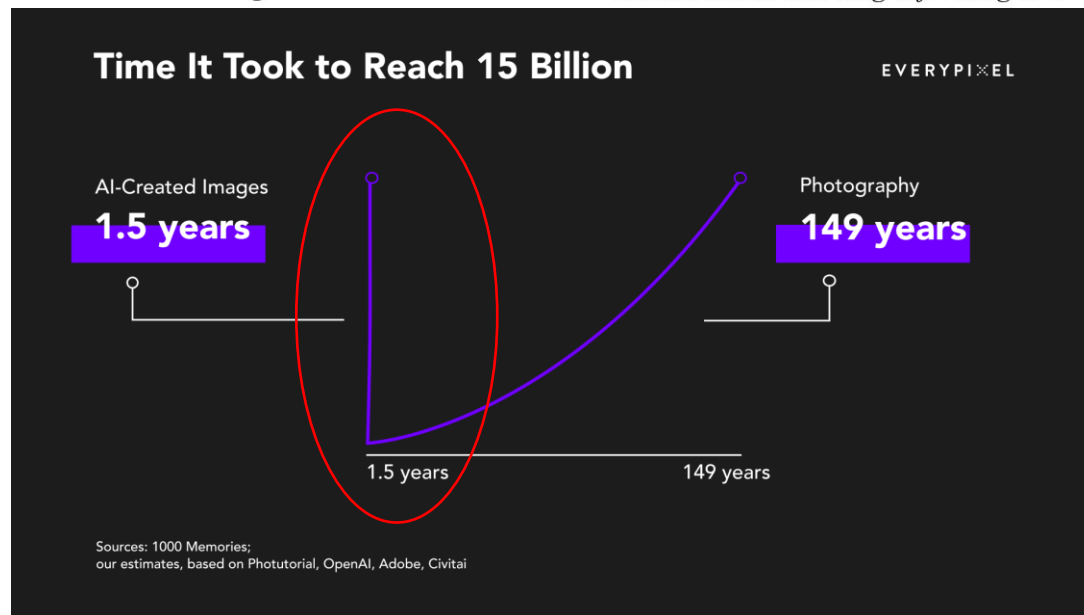# Results of text-to-image generation

- the results attracted a **wide interest and public discussions**

- during the last years, such systems have become **available to the general public** to use online and **Internet is now flooded with synthetic AI images**

- as of August 2023, **AI had already created as many images as photographers have taken in 150 years!** [1]
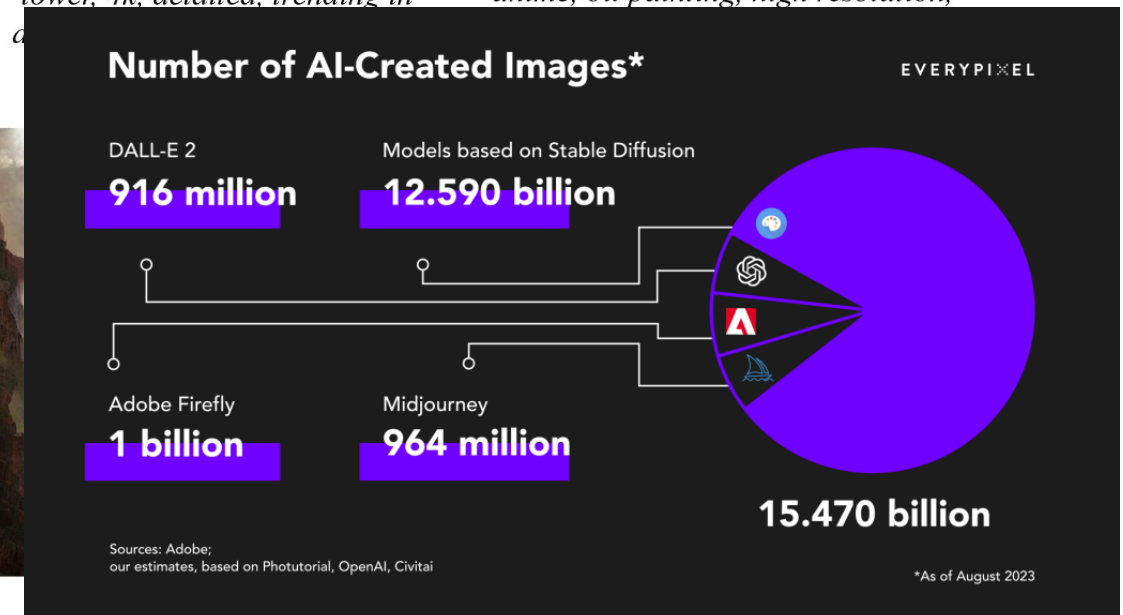
"A cute corgi lives in a house made

"A strawberry mug filled with white sesame seeds. The mug is floating in a

"A digital Illustration of the Babel tower, 4k, detailed, trending in

"Cluttered house in the woods, anime, oil painting, high resolution,

Time It Took to Reach 15 Billion   EVERYPIXEL

AI-Created Images
1.5 years

Photography
149 years

1.5 years   149 years

Sources: 1000 Memories;
our estimates, based on Phututorial, OpenAI, Adobe, Civitai

Number of AI-Created Images*   EVERYPIXEL

DALL-E 2
916 million

Models based on Stable Diffusion
12.590 billion

Adobe Firefly
1 billion

Midjourney
964 million

15.470 billion

Sources: Adobe;
our estimates, based on Phututorial, OpenAI, Civitai

*As of August 2023

results by *Imagen* (Saharia et al. 2022)

results by *Stable Diffusion* (Rombach et al. 2022)

[1] A. Valyaeva. AI Has Already Created As Many Images As Photographers Have Taken in 150 Years. Statistics for 2023. Everypixel, Aug 2023

Impact of AI image & video generators

Anastasios Roussos

Impact of AI image & video generators

- **Positive impact**:

  - **Visual arts** (painting, design, filmmaking, etc): **democratization**, novel means of **artistic experimentation**

  - **Human-like virtual assistants** for learning, healthcare, accessibility, etc

Refik Anadol, **Unsupervised**,
Museum of Modern Art, New York (2022)

- **Concerns** and **ethical considerations**:

  - **artists** (painters, graphic designers, actors, filmmakers, …) fear that the developments in AI **will make them lose their jobs**

  - **copyright** concerns: training of AI models on copyrighted media

  - **environmental impact** due to computational resources needed

  - **misuse of deepfakes**: scams, sexual cyberbullying and fake news

  - **bias** in AI image & video synthesis

Strike in Hollywood, demanding **regulations on AI usage**, July 2023

Impact of AI image & video generators

- **Positive impact**:

  - **Visual arts** (painting, design, filmmaking, etc): **democratization**, novel means of **artistic experimentation**

  - **Human-like virtual assistants** for learning, healthcare, accessibility, etc

- **Concerns** and **ethical considerations**:

  - **artists** (painters, graphic designers, actors, filmmakers, …) fear that the developments in AI **will make them lose their jobs**

  - **copyright** concerns: training of AI models on copyrighted media

  - **environmental impact** due to computational resources needed

  - **misuse of deepfakes**: scams, sexual cyberbullying and fake news

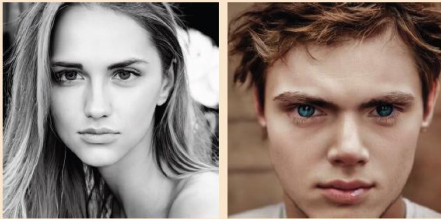 - **bias** in AI image & video synthesis

Synthesized

C.O. Tze, P.P. Filntisis, A.L. Dimou, A. Roussos, P. Maragos.
Deep Photorealistic **Sign Language Retargeting**. AI4CC 2023

Strike in Hollywood, demanding
regulations on AI usage, July 2023

Bias in AI Image Synthesis

- **text-to-image generators** are very prone to reinforcing many forms of stereotypes [1,2]:

  - gender, race, nationality, class &other identities:

TRAITS
"an attractive person"
"a poor person"
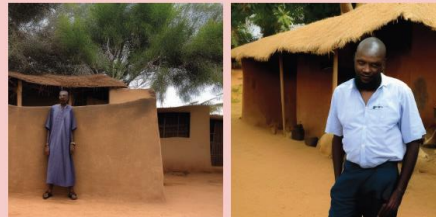
NATIONAL IDENTITIES
"a man from the USA"
"an Iraqi man"

ETHNIC IDENTITIES WITH COUNTER-STEREOTYPES
"a wealthy African man and his house"
"a poor white person"

**ETHNIC IDENTITIES WITH OBJECTS**
"Turkish clothing"
"an African house"

[1] Bianchi, F., ..., Caliskan, A. **Easily accessible text-to-image generation amplifies demographic stereotypes at large scale**. ACM FAT 2023.
[2] L. Nicoletti and D. Bass. **Humans Are Biased. Generative AI Is Even Worse**. Bloomberg, July 2023.

- **gender bias** regarding professions [1,2]:

AI-generated images of **scientists**:

0% women, despite the fact that ~43% of all scientists are women [3]
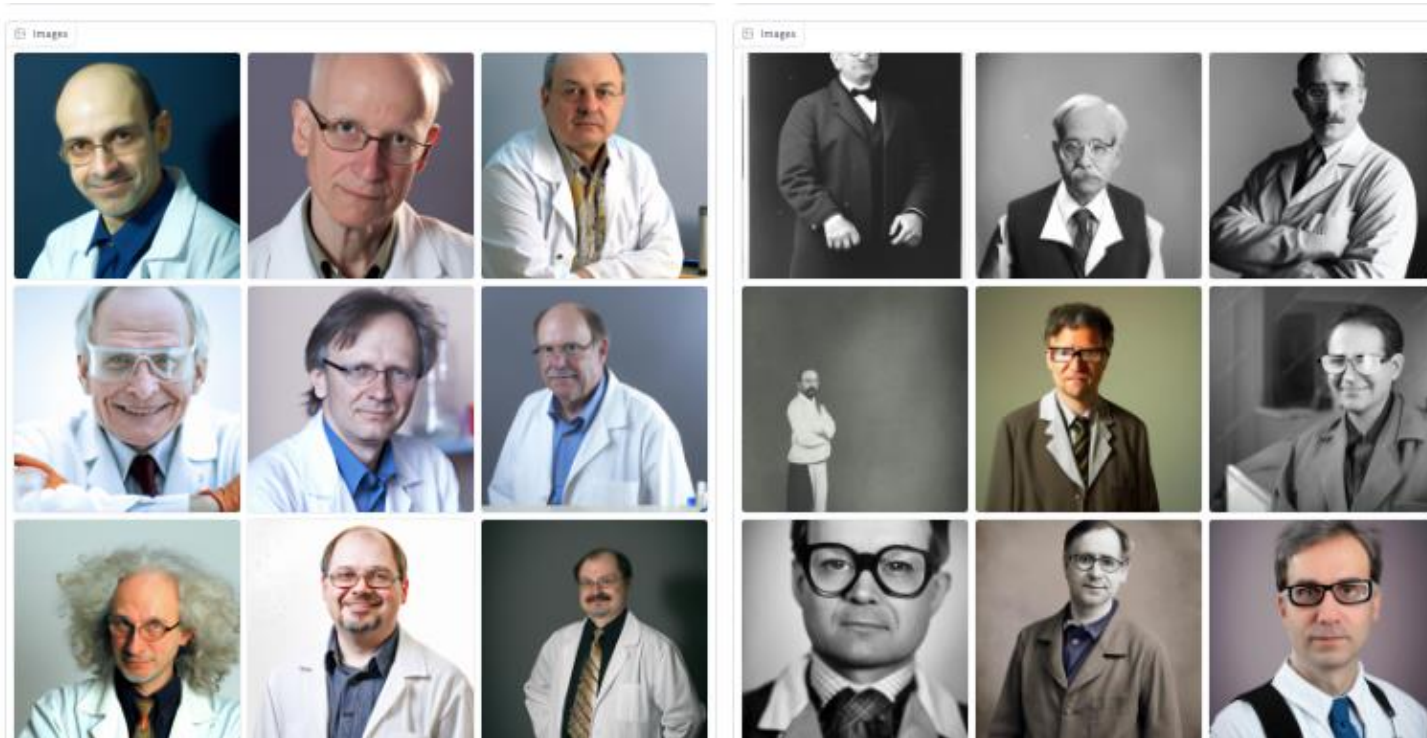
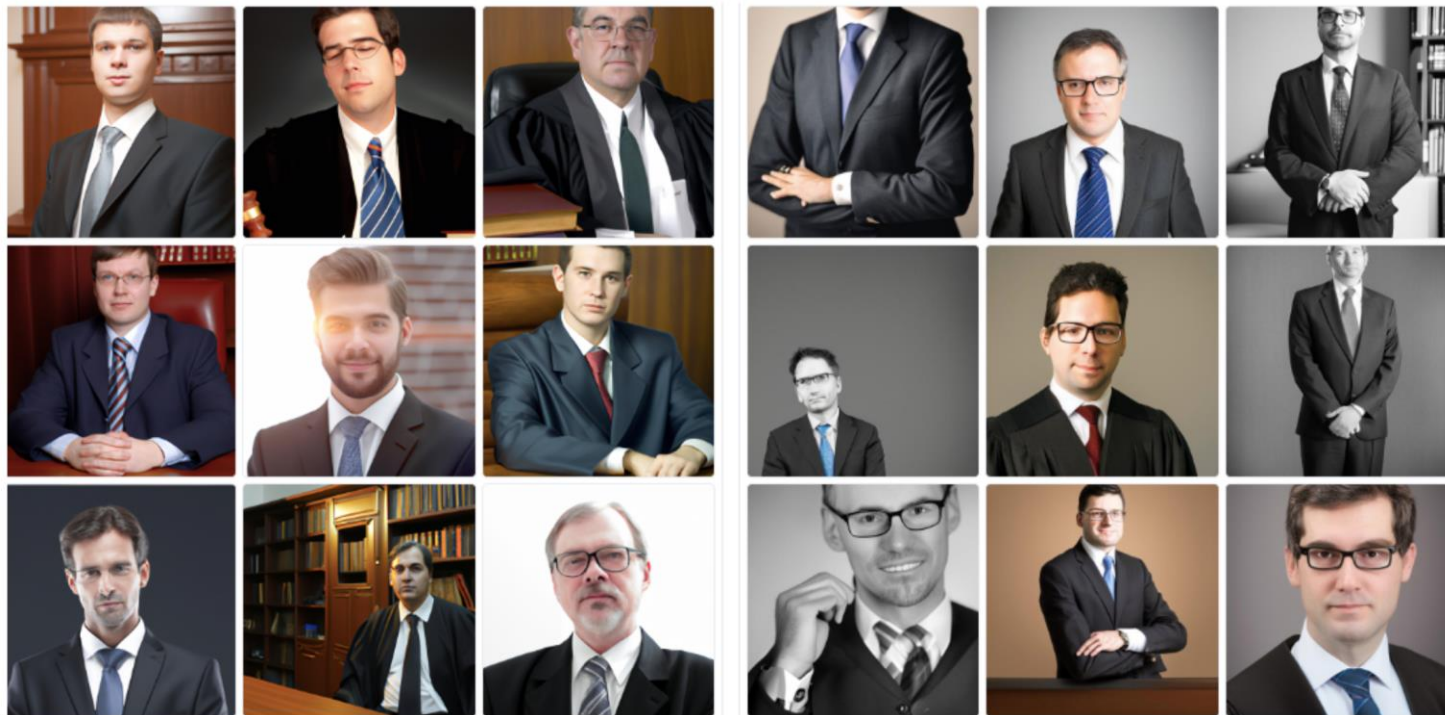[1] Diffusion Bias Explorer by Hugging Face , https://writer.com/blog/ai-bias/
[2] L. Nicoletti and D. Bass. **Humans Are Biased. Generative AI Is Even Worse**. Bloomberg, July 2023
[3] https://www.zippia.com/scientist-jobs/demographics/

- **gender bias** regarding professions [1,2]:

AI-generated images of **lawyers**:

[1] Diffusion Bias Explorer by Hugging Face , https://writer.com/blog/ai-bias/
[2] L. Nicoletti and D. Bass. **Humans Are Biased. Generative AI Is Even Worse**. Bloomberg, July 2023

- **gender bias** regarding professions [1,2]:
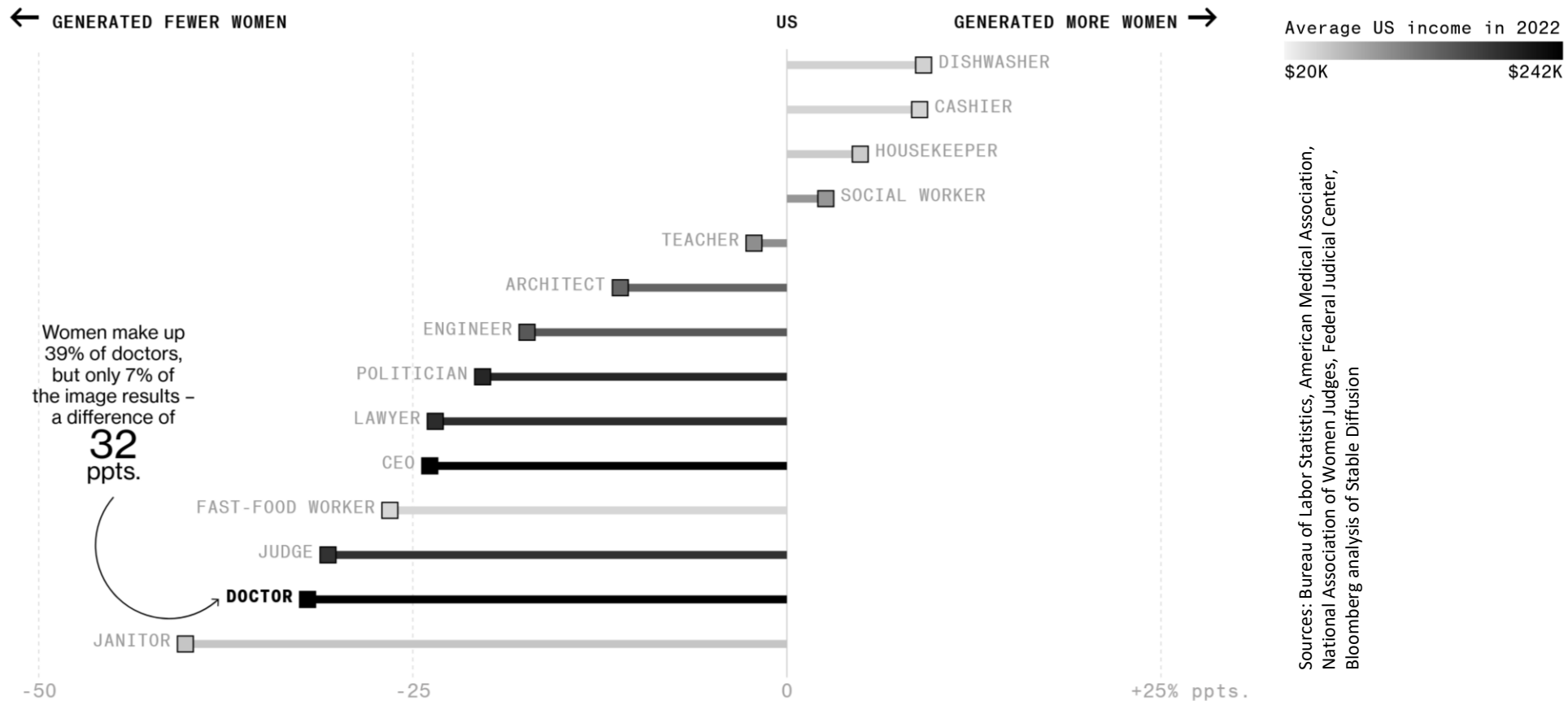
AI-generated images of **nurses**:

[1] Diffusion Bias Explorer by Hugging Face , https://writer.com/blog/ai-bias/
[2] L. Nicoletti and D. Bass. **Humans Are Biased. Generative AI Is Even Worse**. Bloomberg, July 2023

Gender Bias in AI Image Synthesis: Examples

- gender bias in **Stable Diffusion results** versus the **real labor market in USA**:

L. Nicoletti and D. Bass. **Humans Are Biased. Generative AI Is Even Worse**. Bloomberg, July 2023.

14

Bias in AI Image Synthesis: Why is this a problem?

- Many viewers may interpret AI-generated images/videos as **entirely objective or neutral**

- Repeated biased depictions **risk becoming normalized over time**, **shaping perceptions and reinforcing stereotypes**.

- Communities portrayed negatively can **internalize these portrayals**, creating **self-fulfilling prophecies**, while more represented groups **continue to be privileged**
    - e.g. if the prompt "**corporate leadership**" always generates **tall, masculine, formally dressed men**, this implies that authoritative roles are inherently male. This perpetuates a biased perception of leadership and marginalizes women, people of color, and other identities in professional contexts.

- The homogenization of visual content **erodes diversity in creative fields**, overshadowing richer and more inclusive narratives.
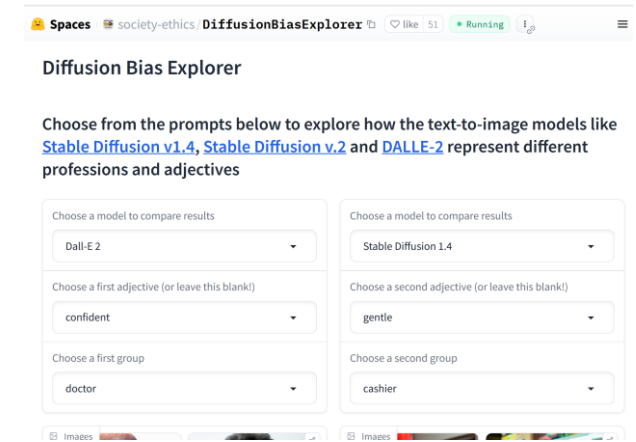
- AI systems **reflect the bias in the datasets that they were trained on:**

  - AI models are often trained on **human-made datasets** that overrepresent certain demographics, aesthetics, or cultural contexts, leading to skewed or limited portrayals in generated content
  - **Subjective labels** created again by humans -like those for beauty or professionalism- may carry implicit cultural assumptions, embedding personal biases of annotators into AI outputs.

- **Feedback Loops:** Biased AI-generated images can become part of new training sets, perpetuating and intensifying inaccuracies or stereotypes over time.

- **Bias Amplification:** When a bias is present in the data, the **model's tendency to generalize patterns in a simplistic manner** can amplify that bias, producing outputs that reinforce existing stereotypes at scale.

Bias in AI Image Synthesis: What can we do?

Countermeasures:

- **growing public discussion**: more and more people (scientists & general public) discuss about this bias and explore ways to reduce it

- **tools** that **help reveal & analyze the bias** in AI image generation

- **technical solutions** (diversifying training data, bias detection & mitigation algorithms, user feedback incorporation, …)

- including **diverse teams** in the design and decision-making regarding AI technologies
  - e.g. it was estimated that in 2019 more than 80% of AI instructors were male [1]

[1] S. Myers West. **Discriminating Systems: Gender, Race, and Power in AI**. Report, AI Now Institute, April 2019

Positive role of Generative AI in Gender Issues

- **Interactive installations**, like the Mirror of Diversity:
  - users approach issues related to **inclusion and diversity** in an **experiential way**

Using Generative AI to raise awareness on Gender Issues

- **Beyond the Gender Binary**:

Zizi Show (2020), by Jake Elwes, an **interactive deep fake drag cabaret** that criticizes the **discriminatory behavior** of the datasets used in AI

  - commenting on the fact that computer vision systems have **difficulty recognizing trans, queer and other marginalized identities**.
  - Zizi show attempts to **expose and subvert this bias**

currently at Victoria and Albert Museum, London

Conclusions

Conclusions

- 13 years after the beginning of the "Deep Learning revolution", AI for image and video synthesis **has achieved many important milestones**
  - the current AI models are able to generate synthetic images & videos of **unprecedented quality and realism**
  - there are models that are **accessible by the general public** and this is **attracting more interest than ever**
  - the technological developments **continue being impressive**

Conclusions

- many examples of bias in AI image and video synthesis, with gender bias being one of the most important types
  - **persisting problem**, not easy to address
  - several strategies to **mitigate** these effects have been explored, however much more need to be done
  - urgent need for **scientists & engineers in our fields to be more engaged and proactive** regarding the relevant **gender issues**

- **AI can merge with art** to **promote gender equality**, while also **challenging social structures** that reinforce **gender inequality**

Thank you for your attention!

Paraskevi (Vivi) Valergaki,
Anastasios (Tassos) Roussos

https://users.ics.forth.gr/~troussos

Institute of Computer Science (ICS),
Foundation for Research and Technology - Hellas (FORTH)