TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Advanced Seminar on Computer Graphics and Visualization

Recent Trends in Generative AI

Kick-off Meeting

APB 1004 // 13th April 2023 // 2 DS

Computer Graphics and Visualization

Introduction to Generative AI

- Generative AI is a subset of artificial intelligence that generates new data.
- Examples of data that can be generated, includes images, text, and videos.
- Potential uses of Generative AI in developing new products, services, and ideas.

Image generated by DALL-E based on text
„A lion wearing sunglasses and reading a book in library“

Objective of the Seminar

- Learn advanced concepts of Generative AI.
- Learn how to conduct high-quality scientific writing.
- The essential aspects of literature study for success in scientific writing.
- The importance of sharing research work in science through publications.

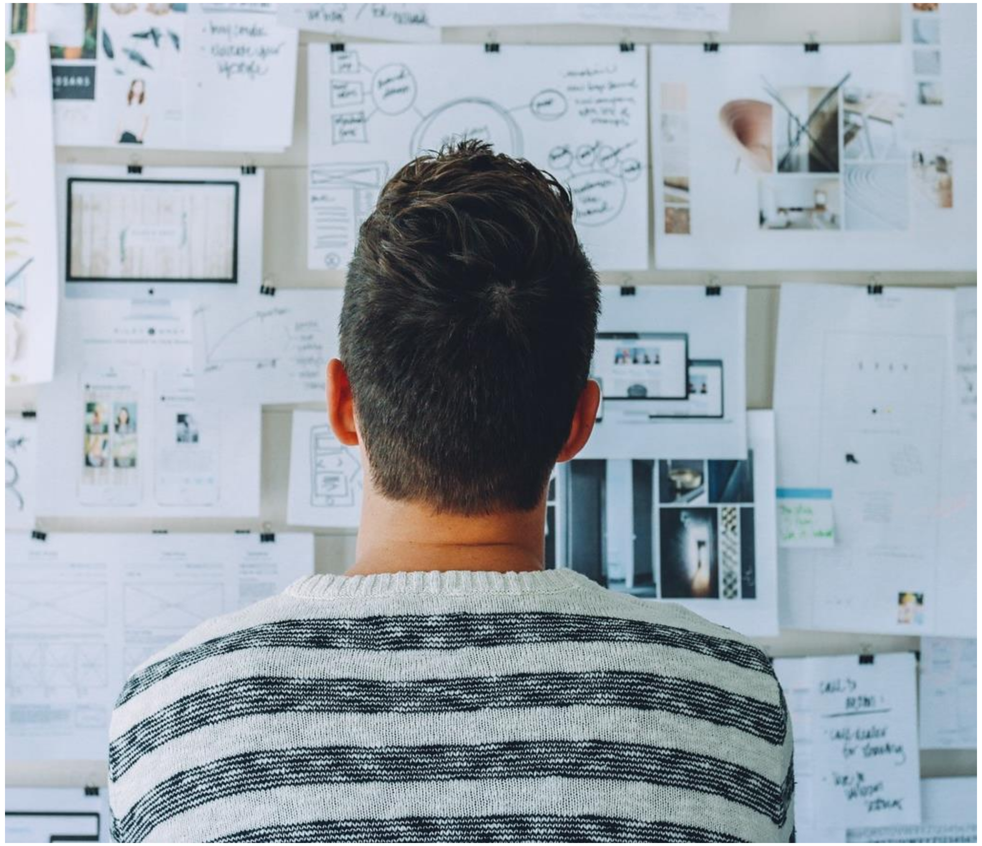


Image source: SlidesAI.io

Teamwork and Grading

- Teams of up to three members can take a topic.
- Grading is done on an individual basis based on the contribution in each milestones.
- Maintain a team diary for marking individual contributions.
- Maintain a database of papers (e.g. with Zotero).
- Study the papers in depth and collaborate with teammates. Advisable to divide the task equally.

Image source: SlidesAI.io

Topic Selection and Contact with Supervisor

- A doodle poll will be live soon for the selection of the topic.
- Check the seminar webpage for updates.
- Deadline to select a topic is 20th April.
- The team should contact their supervisor after the topics has been assigned.
- Have a regular contact with your supervisor for guidance throughout the seminar duration.

Image source: SlidesAI.io

Tips for Scientific Writing

- A special lecture to provide you guidance on conducting quality scientific research and writing.
- Planned for 27th April, 09:20 am, 2 DS.
- Will also be conducted onsite. The location will be revealed on the seminar webpage.

https://handmadewriting.com/blog/guides/scientific-paper/

Ethical Considerations in Scientific Writing

- The ethical considerations in scientific writing including plagiarism, conflicts of interest, or falsification.

- The importance of adhering to ethical standards for the credibility and reproducibility of research.

https://www.mpi-mps.org/about/code-of-ethics/

1st Milestone - 1 Page Review

- The team will submit a 1 page review of publications that are highly important for the selected topic.
- It is advisable to send the 1 page draft to your supervisor before final submission.
- Submission Deadline: 18th May

Image source: SlidesAI.io

2nd Milestone - Detailed Scientific Report

- After the submission of the 1 page review, the team will study relevant publications in detail.
- The team will write a detailed scientific report of 15-20 pages following ethical and scientific standards.
- It is encouraged to collaborate with teammates and find solutions to open problems.
- Have regular meetings with your supervisor to seek guidance.
- Submission Deadline: 16th June

Image source: SlidesAI.io

3rd Milestone – Final Presentation

- The team will discuss their work with a PowerPoint style presentation.
- Importance of clear and effective communication in scientific presentations.
- The final presentation will also be conducted purely onsite.
- Tentative Schedule: 13th July

Image Source: DALL-E

Summary

- The scientific research in the field of Generative AI has enormous potential for advancing knowledge in the field.

- High-quality scientific writing is essential for success in science and requires careful planning, collaboration, and adherence to ethical standards.

Image source: SlidesAI.io

Lets check out the topics

Seminar Topics on Generative AI

- Three broad topics focusing on technical concepts behind Generative AI.
- Other sub-topics with emphasis on both technical concepts and application use cases.

Broad topics

- Stable Diffusion Networks
- Neural Radiance Fields (NeRFs)
- Generative Transformers

- Have your own topic in mind?

  Contact me nishant.kumar@tu-dresden.de

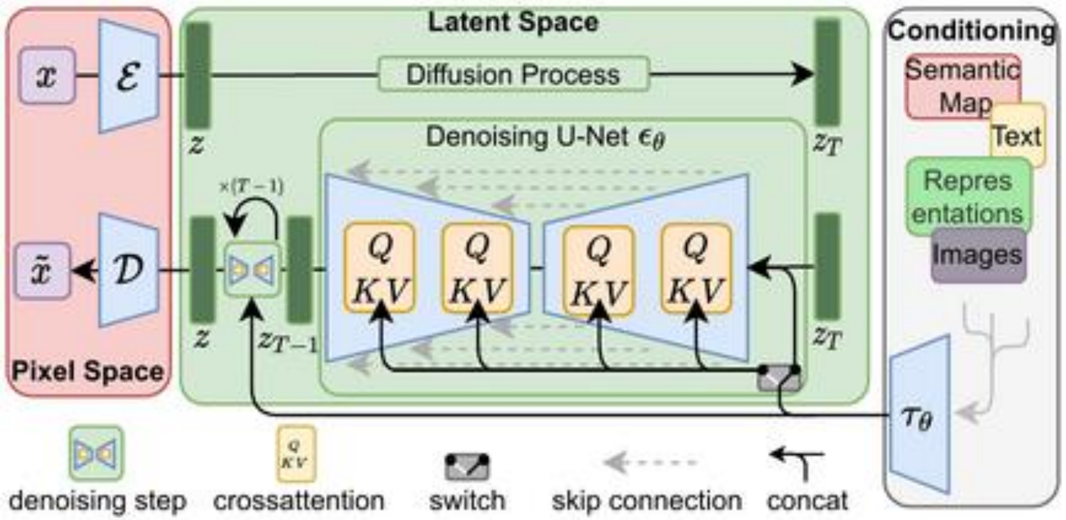
Sub topics

- Text-to-Image Generation
- AI-driven video synthesis
- Ethical aspects of AI-generated content
- Text-to-Human Motion
- Images, Text, and Human Body Shapes
- Text-to-Mesh
- Text-to-point cloud
- Other topics?

Broad Topic: Stable Diffusion Networks

Contact: Nishant Kumar nishant.kumar@tu-dresden.de

- A **forward diffusion** process adds noise to a training image, gradually turning it into an uncharacteristic noise image.
- Eventually, you won't be able to tell the semantics from the original image.
- A **reverse diffusion** process generates a noisy image and propagates this image back to the pixel space.



Source: High-Resolution Image Synthesis with Latent Diffusion Models

Relevant papers:

[1] Diffusion Models: A Comprehensive Survey of Methods and Applications
[2] Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding
[3] High-Resolution Image Synthesis with Latent Diffusion Models

Sub-Topic: Text-to-Image Generation

Contact: Nishant Kumar nishant.kumar@tu-dresden.de

- Give the model a text prompt, and it returns an image.
- Combines a language model with a generative image model.
- The language model transforms the input text into a latent representation.
- The generative image model, produces an image conditioned on the latent representation of text.

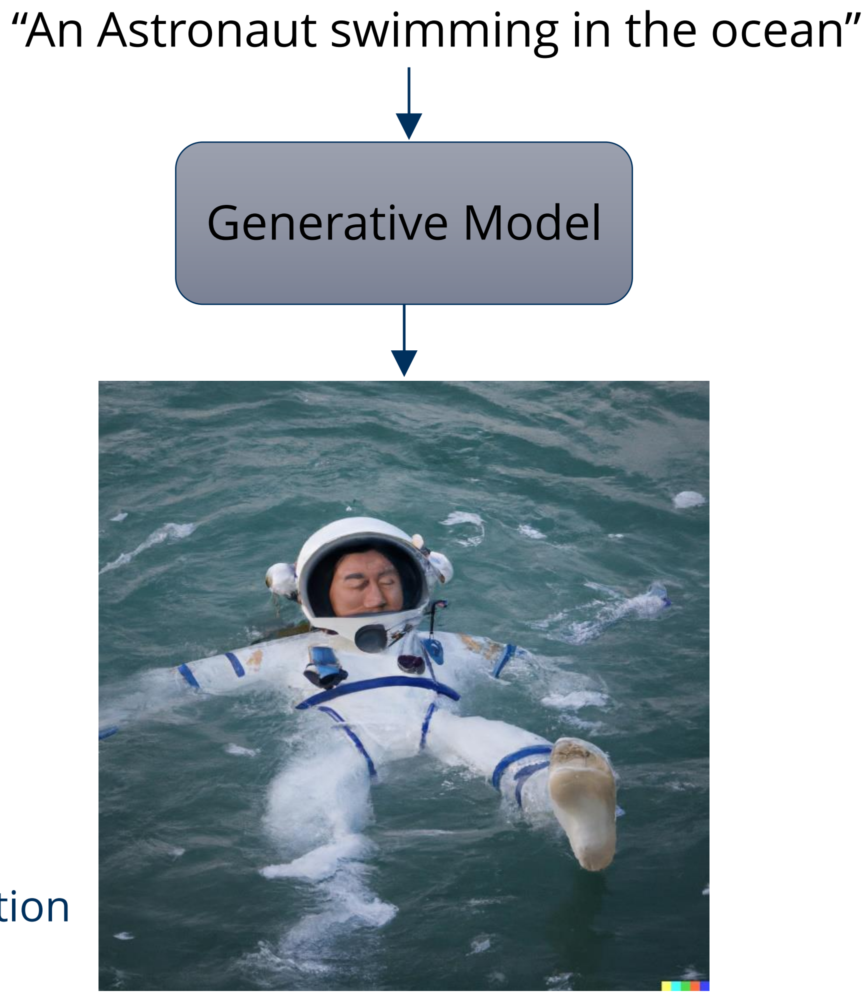


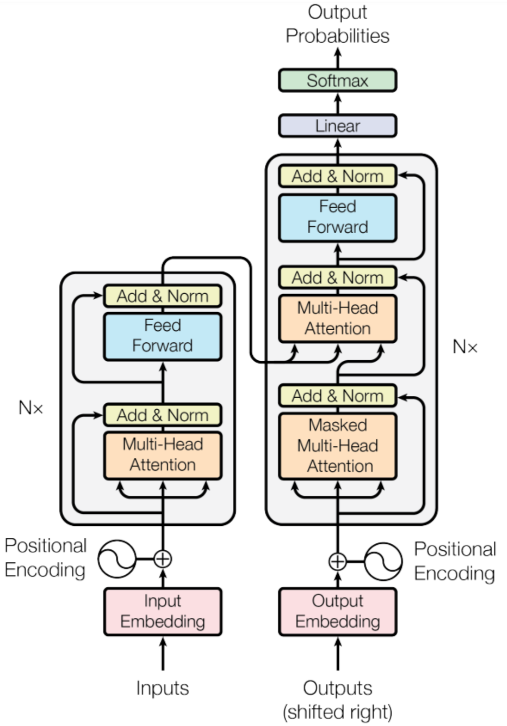
Image source: DALL-E

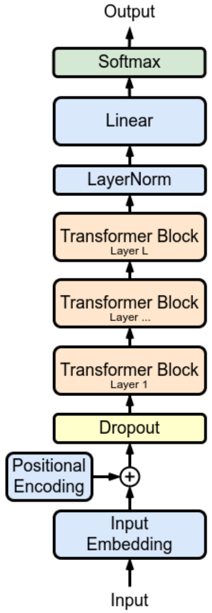
Relevant papers:

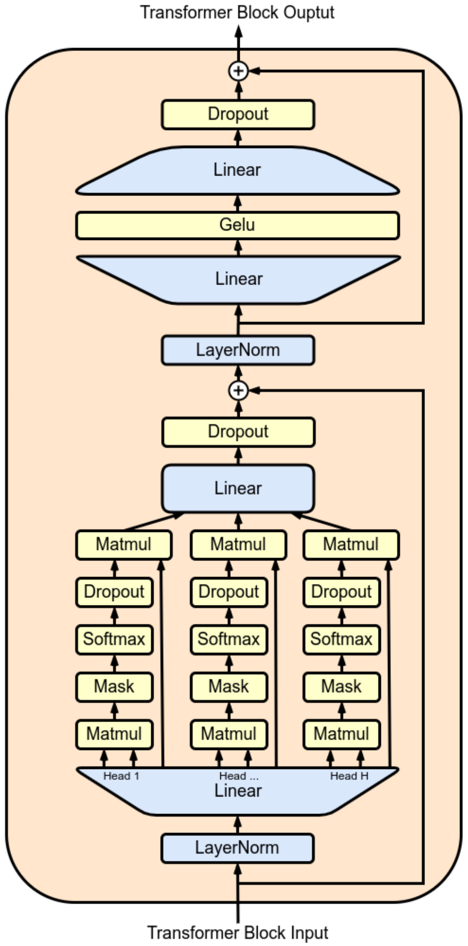
[1] Variational Distribution Learning for Unsupervised Text-to-Image Generation
[2] Cogview: Mastering Text-to-Image Generation via Transformers
[3] Taming Transformers for High-Resolution Image Synthesis
[4] Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models

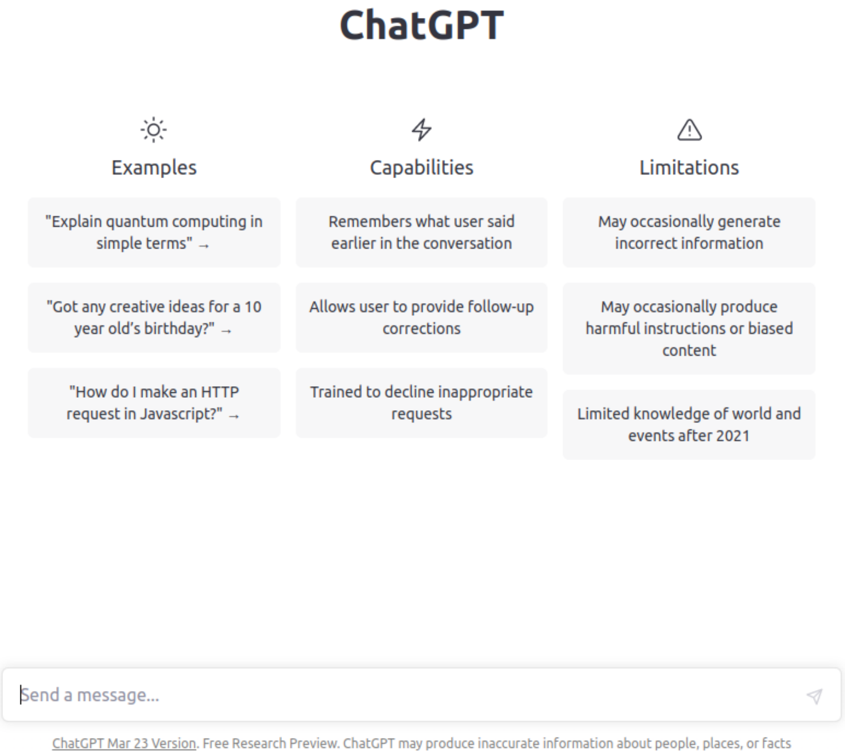
Broad Topic: Generative Transformers

Contact: Kristijan Bartol kristijan.bartol@tu-dresden.de







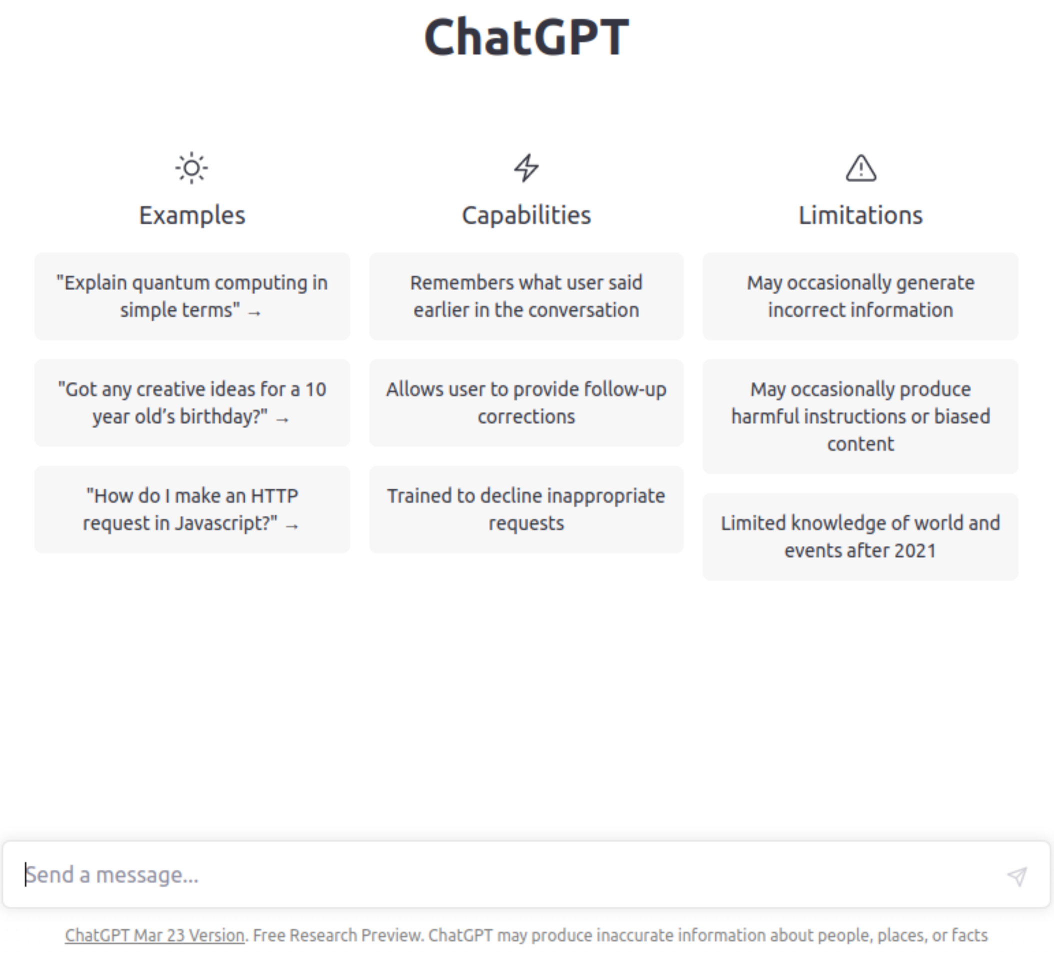


Broad Topic: Generative Transformers

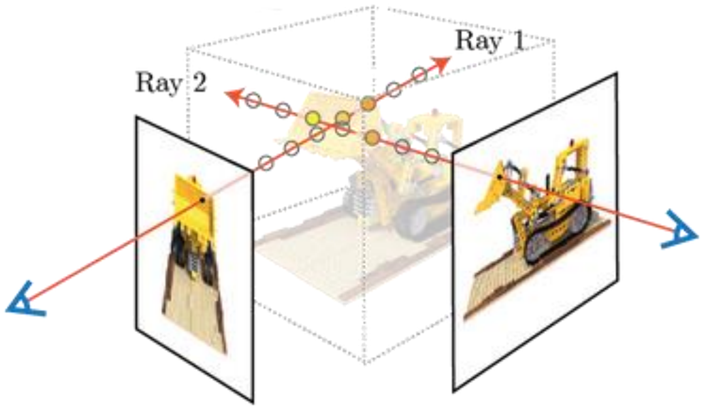
- Describe how to get from lower-level architectures (e.g. Transformers) to systems like ChatGPT

- Keywords: natural language processing (NLP), next-token-prediction, LSTMs, transformers, …

- References:
  - Attention is All You Need
  - Language Models are Few-Shot Learners
  - …



Broad Topic: Neural Radiance Fields (NeRFs)



Goal

Represent complex scene with relatively simple neural network

How

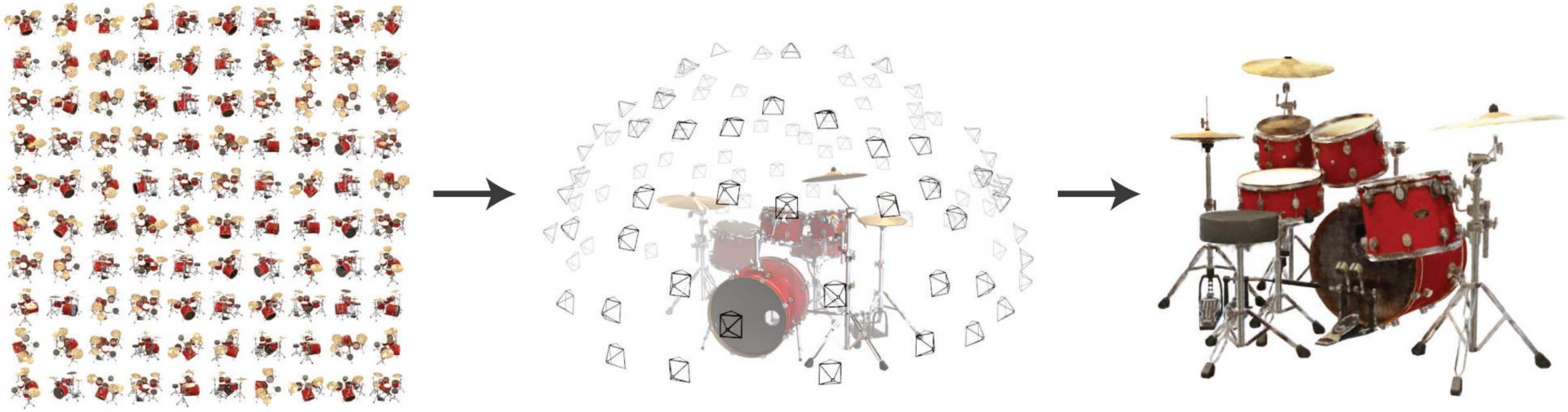
Sample positions along view rays

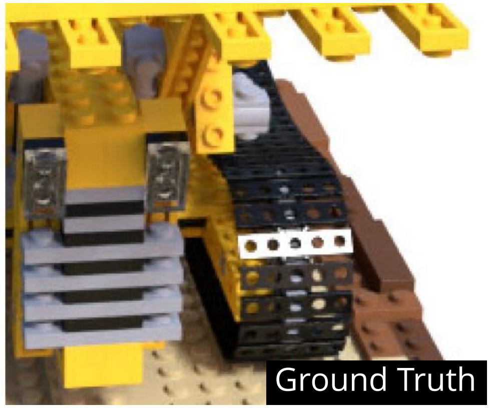
Infer scene information per sample

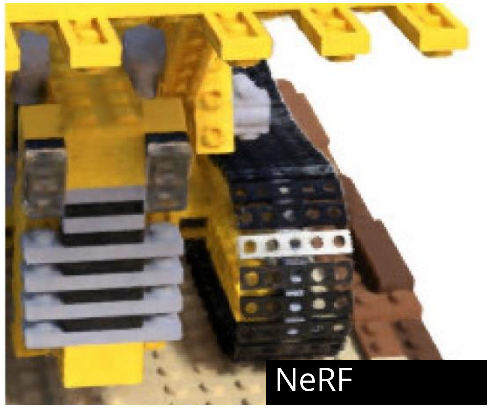
Combine using volume rendering

$$(x, y, z, \theta, \phi) \rightarrow \mathrm{DNN} \rightarrow (\mathrm{R}, \mathrm{G}, \mathrm{B}, \sigma)$$

Input position and direction.
Get Color and opacity.







Broad Topic: Neural Radiance Fields (NeRFs)

Applications

- Generate novel views from a limited set of input images
- Reduce memory requirements to represent complex scenes
- Capture real-world objects and convert them to meshes

Questions

- Which scene representation approaches exist?
- How do the methods compare with respect to training time, rendering performance and image quality?

References

- ***Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi and Ren Ng:*** *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.* In Computer Vision – ECCV 2020, 405 – 421. DOI: https://doi.org/10.1007/978-3-030-58452-8_24
- ***Thomas Müller, Alex Evans, Christoph Schied and Alexander Keller:*** *Instant Neural Graphics Primitives with a Multiresolution Hash Encoding*. In ACM Trans. Graph., New York, NY, USA (July 2022). DOI: https://doi.org/10.1145/3528223.3530127
- ***S. Weiss, P. Hermüller and R. Westermann:*** *Fast Neural Representations for Direct Volume Rendering*. Computer Graphics Forum, 41: 196-211 (September 2022). DOI: https://doi.org/10.1111/cgf.14578

Contact

David Groß david.gross1@tu-dresden.de

Sub-topic: AI-driven video synthesis

with stable diffusion or similar techniques (Runway Gen-1/2)

Contact: Ludwig Schmutzler ludwig.schmutzler@tu-dresden.de

Intro (generated via ChatGPT)

Stable diffusion is an emerging approach in the field of AI-generated text-to-video that shows great potential for creating high-quality and diverse video content from textual input.

One of the significant challenges in this approach is ensuring that the generated video sequence is visually consistent and coherent with smooth transitions between frames to produce visually compelling video sequences.

This approach involves using natural language processing (NLP) to analyze the input text and generate a visual representation of the text, which is then transformed into a video sequence using an iterative refinement process. Despite challenges such as the need for large amounts of training data and computational resources, stable diffusion models are a promising area of research and development in the field of AI-generated text-to-video, and they are expected to have a profound impact on the future of video content creation and consumption.

Sub-topic: AI-driven video synthesis
with stable diffusion or similar techniques (Runway Gen-1/2)

Tasks
— do research on AI video synthesis (e.g. stable diffusion, Gen-1, Gen-2) and related techniques (e.g. morphing in latent space)
— describe basic techniques e.g. text-to-video, image-to-video, image+text-to-video, video stylization, video inpainting
— what challenges arise regarding "moving pictures" (in comparison to still pictures)?
— what influences have refined SD models in the process?

Starting points:
— VideoFusion: Decomposed Diffusion Models for High-Quality Video Generation (March 2023), arxiv.org/pdf/2303.08320v2.pdf
— Structure and Content-Guided Video Synthesis with Diffusion Models (Feb 2023), arxiv.org/pdf/2302.03011.pdf
— Make-A-Video: Text-to-Video Generation without Text-Video Data, arxiv.org/pdf/2209.14792.pdf
— Video Diffusion Models, arxiv.org/abs/2204.03458
— InstructPix2Pix: Learning to Follow Image Editing Instructions, arxiv.org/pdf/2211.09800.pdf
— High-Resolution Image Synthesis with Latent Diffusion Models, arxiv.org/pdf/2112.10752.pdf

Wanna play around? Get in touch and follow the Google Colab links on the pages below:
https://github.com/nateraw/stable-diffusion-videos
https://github.com/camenduru/pix2pix-video-colab
https://huggingface.co/spaces/damo-vilab/modelscope-text-to-video-synthesis
https://huggingface.co/spaces/nateraw/stable-diffusion-music-videos

Sub-topic: Ethical aspects of AI generated content

on how to deal with models that have internet-scale bias

Contact: Ludwig Schmutzler ludwig.schmutzler@tu-dresden.de

Intro (generated via ChatGPT)

As we delve deeper into the capabilities of generative AI models such as DALL-E 2, Imagen, and Stable Diffusion, we must also consider the potential impact of these models on different aspects of our society.

One such aspect is the issue of equality, particularly when it comes to pornographic imagery, racism, hate speech, harmful social stereotypes or religous bias. While generative AI models have the potential to produce diverse and inclusive content, they can also perpetuate biases and stereotypes* that exist in our society. For instance, a generative AI model trained on a (un-curated) dataset that is predominantly male or white may struggle to generate content that is representative of diverse perspectives.

Moreover, generative AI models have the potential to create content that could be deemed offensive or harmful to certain cultures. This raises important ethical questions about the responsibility of developers and users of generative AI models to ensure that the content generated by these models is respectful and considerate of all individuals.

Your task is to explore these challenges and examine potential solutions to ensure that generative AI models are not perpetuating biases and inequalities in our society.

*Those models typically tend to present stereotypes from a western point of view.

Sub-topic: Ethical aspects of AI generated content

on how to deal with models that have internet-scale bias

Tasks

- do research and present the problems regarding different ethical aspects (see intro)
- starting from related literature, describe different approach to cope with the problems
- how can one achieve accurate representations of cultures (as well as minorities, women, LGBTQIA+, …)
- examine the challenges of curated training data
- compare generated data for different models (curated vs. non-curated)
- conclude

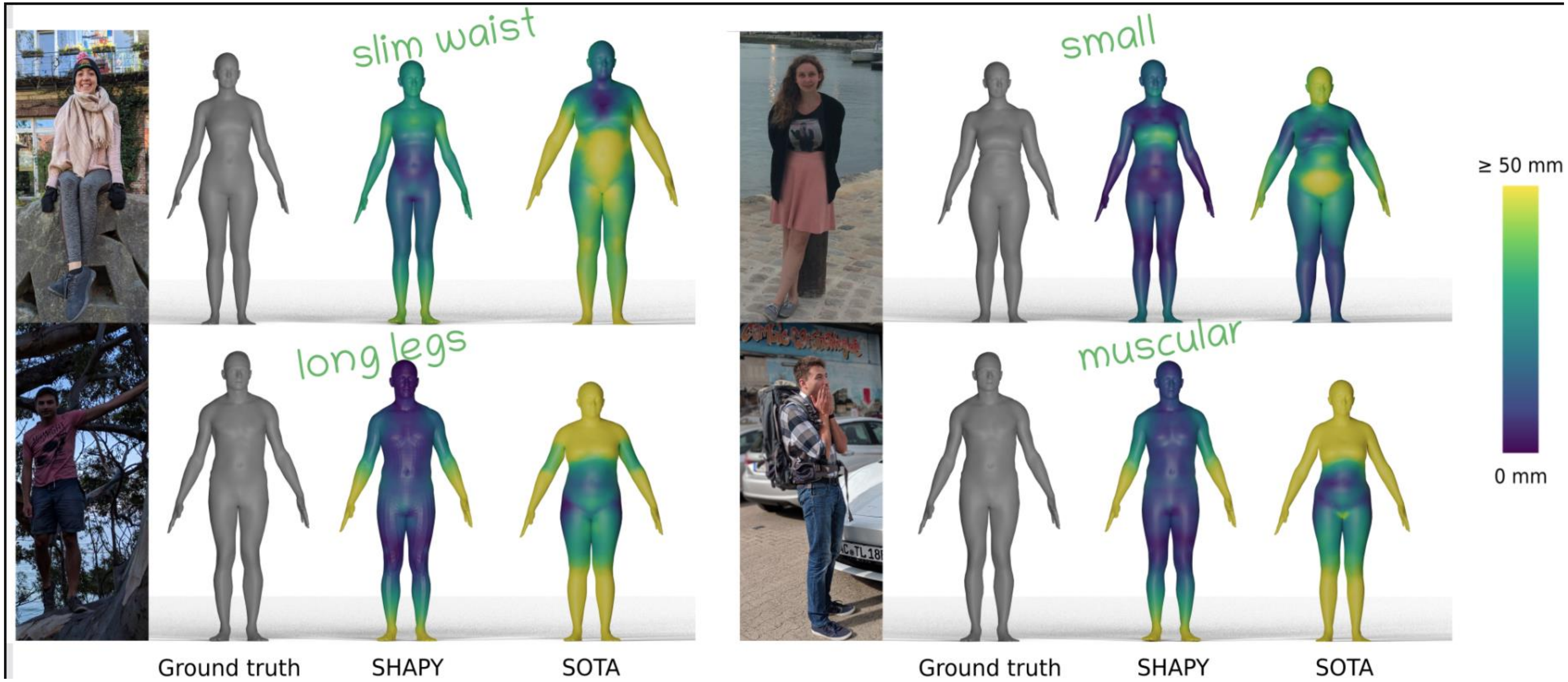
Starting points:

- https://www.technologyreview.com/2022/05/25/1052695/dark-secret-cute-ai-animal-images-dalle-openai-imagen-google/
- https://matthewpburruss.com/post/the-unethical-story-of-gpt-3-openais-million-dollar-model/
- https://huggingface.co/blog/japanese-stable-diffusion
- https://www.herzindagi.com/society-culture/artificial-intelligence-ai-artwork-based-on-stereotypes-article-218643
- https://openreview.net/pdf?id=uGmiCvop2zT
- https://www.cultural-ai.nl/projects/aicult-culturally-aware-ai

Sub-topic: Images, Text, and Human Body Shapes

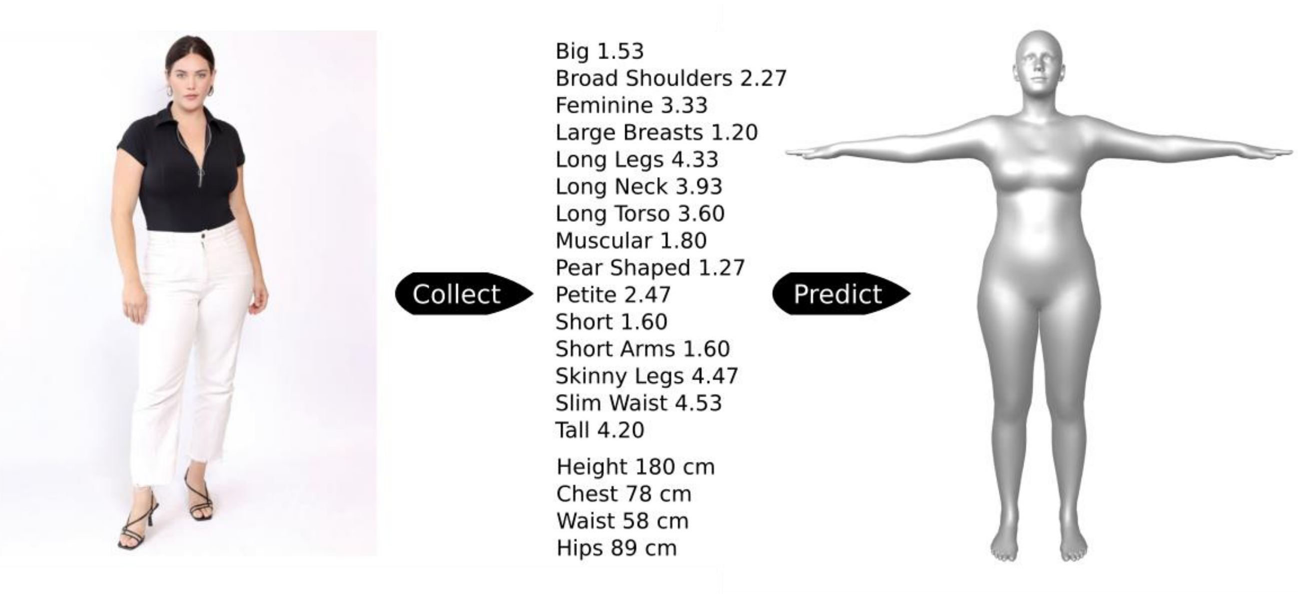
Contact: Kristijan Bartol kristijan.bartol@tu-dresden.de

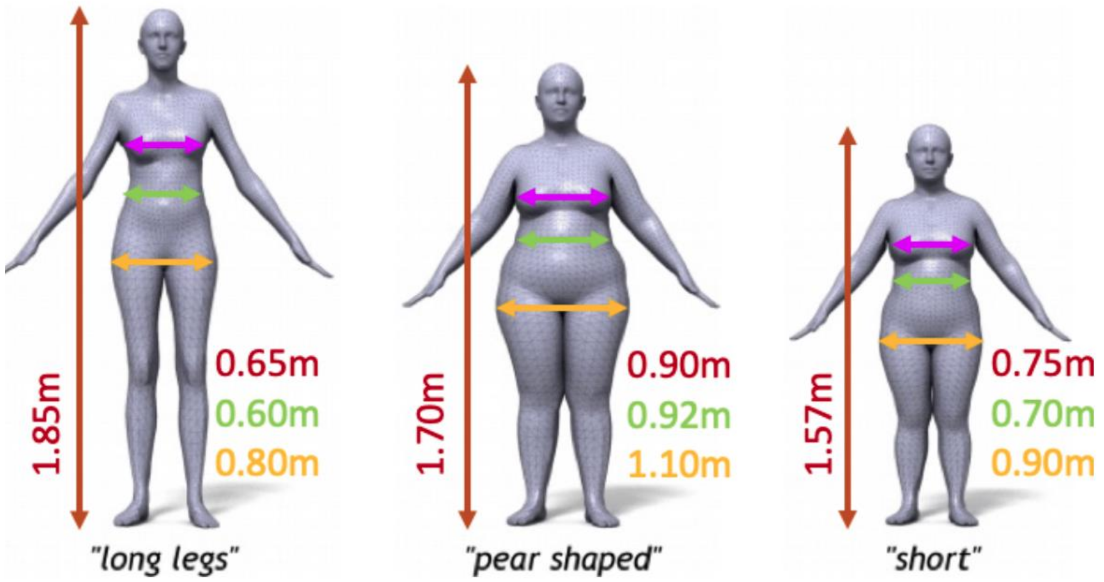
- Human body shapes extracted from images and described in simple words



Sub-topic: Images, Text, and Human Body Shapes

- The crucial component is extracting and interpreting the body measurements





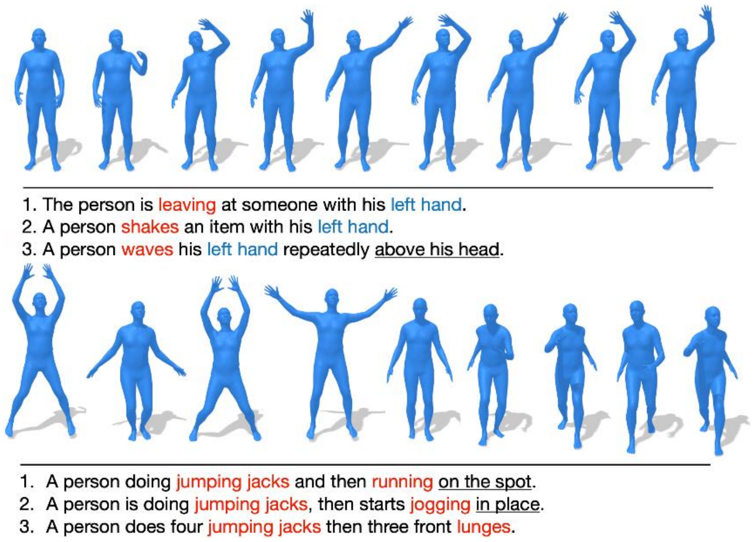
[Reference paper]: Accurate 3D Body Shape Regression using Metric and Semantic Attributes, CVPR, 2022

Sub-topic: Text-to-Human Motion

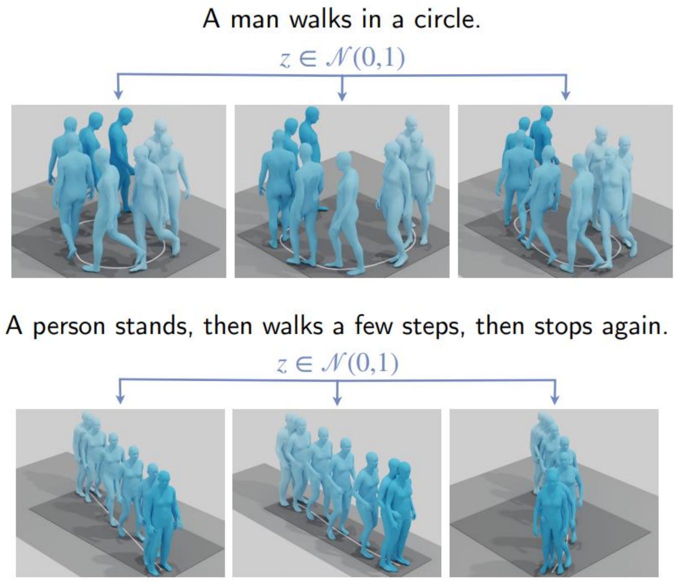
Creation of plausible human motion data is laborious and costly
- Motion capture needs equipment and actors
- Manual creation is time consuming and needs expertise for convincing results

When reading a motion description e.g. in a book, the brain is able to visualize this description with high plausibility, based on our own experience

Text-to-Human Motion tries to mimic this behavior through machine learning



Guo et al. 2022



Petrovich et al. 2022

Sub-topic: Text-to-Human Motion

Research questions:

- How can textual input be used to create plausible human motions?
- Where can this be applied to?
- How good are the results?
- What are the limitations?

Starting points:

- C. Guo, S. Zou, X. Zuo, S. Wang, W. Ji, X. Li, L. Cheng, “Generating Diverse and Natural 3D Human Motions from Texts,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2022
- M. Zhang, Z. Cai, L. Pan, F. Hong, X. Guo, L. Yang, Z. Liu, “MotionDiffuse: Text-Driven Human Motion Generation with Diffusion Model,” *arXiv preprint.* 2022
- M. Petrovich, M.J. Black, G. Varol, “TEMOS: Generating diverse human motions from textual descriptions,” *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXII.* Cham: Springer Nature Switzerland, 2022.
- M. Plappert, C. Mandery, T. Asfour, “The KIT Motion-Language Dataset,” *Big data* 4.4 (2016): 236-252

Contact

Julien Fischer julien.fischer@tu-dresden.de

Sub-topic: Text-to-Mesh

Goal

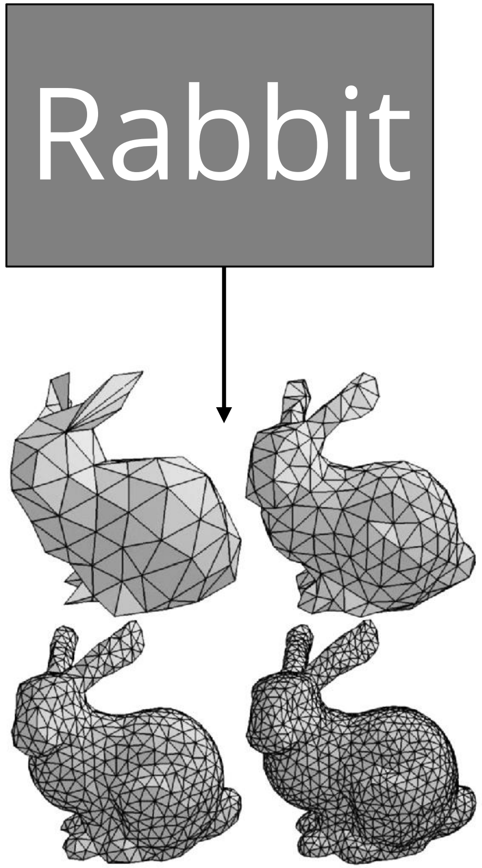
Simplification of Text to mesh Generation

How

Using 3D modeling/specialized software

Programing languages

Source: Khalid et.al (September 2022)



Source: Novakovic et.al (December 2017)

Sub-topic: Text-to-Mesh

Applications

- Detailed 3d meshes of organs and tissues for medical visualization
- Generation of 3D meshes of complex machinery or equipment to support manufacturing process
- Support gaming industry by using text to mesh to streamline their workflow for more realistic game objects

Questions

- How to improve the quality of generated meshes?
- How scale to large scenes by maintaining the level of details?

References

- ***Predgrag Novakovic, Milan Hornak, Mgr. Jan Zachar, Nenad Joncic*** *3D Digital Recording of Archaeological, Architectural and Artistic Heritage.* In Books (December 2017). DOI: 10.4312/9789612378981
- ***Nasir Mohammad Khalid, Tianhao Xie, Eugene Belilovsky, Tiberiu Popa*** *CLIP-Mesh: Generating textured meshes from text using pretrained image-text models.* arXiv:2203.13333v2 [cs.CV] (September 2022). DOI: https://doi.org/10.1145/3550469.3555392

Contact

Marzan Tasnim Oyshi marzan_tasnim.oyshi@tu-dresden.de

Sub-Topic: Text-to-Point cloud

Contact

Tianfang Lin tianfang.lin@tu-dresden.de

Other suitable topics in mind?

Contact

Nishant Kumar nishant.kumar@tu-dresden.de

Thank you. Please feel free to ask any questions. 😄