INDIAN INSTITUTE OF TECHNOLOGY - JODHPUR

OAN7010 Summer 2022 - System Engineering and Project Management

Assignment: Technology Roadmapping

Technology: AI: Text to Image Generation

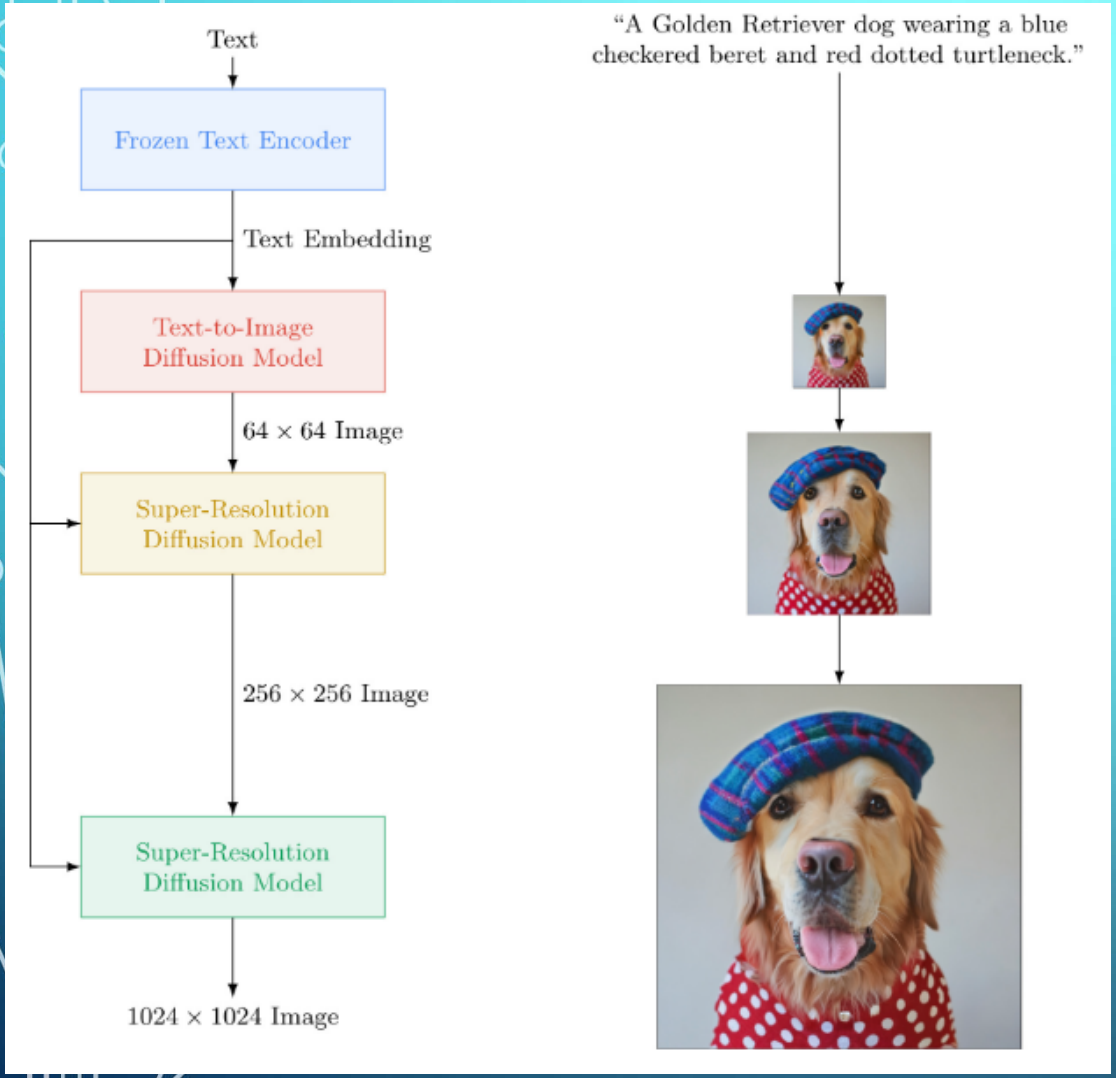
Group 3

Ajeet Kumar Singh (M20AIE208)
Arvind Sriram G (M21AIE218) [Lead]
Bipin Thomas (M20AIE226)
Bishnu Kumar Kashyap (M20AIE227)
Manoj Kumar Maharana (M20AIE252)
Nitish Kumar (M20AIE269)
Rohit Mathur (M21AIE249)

Meeting 1: 25th June 2022: 1900 – 1915 Hrs
Meeting 2: 26th June 2022: 1000 – 1030 Hrs
Meeting 3: 29th June 2022: 1930 – 2030 Hrs

PRESENT STATE

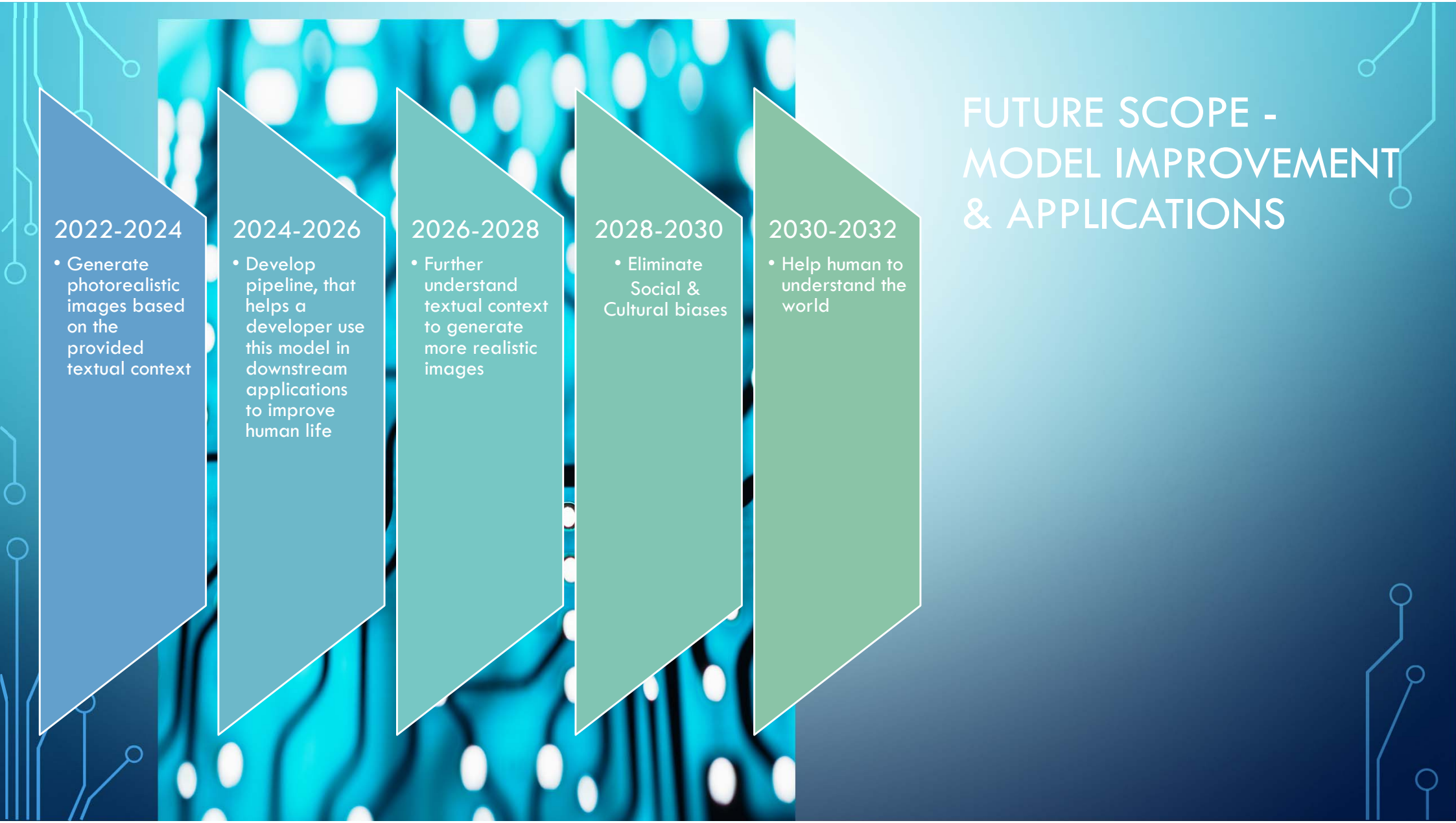
- **Scope of Technology: Generation of realistic image from input text**
  - **Example: (Image courtesy: Google Research's Imagen)**



PRESENT STATE

- Current Significant Works:
  - Google Research's Imagen (https://imagen.research.google/)
    - Text-to-image diffusion model with an unprecedented degree of photorealism and a deep level of language understanding
    - Employs large pre-trained frozen text encoders, its size being more important than diffusion model size
    - Uses novel thresholding diffusion sampler that allows use of very large classifier-free guidance weights
    - Incorporates a compute/memory efficient architecture (U-Net)
    - Performs well on the COCO dataset
  - OpenAI's DALL-E (https://openai.com/blog/dall-e/)
    - 12-billion parameter version of GPT-3 trained to generate images from text descriptions, using a dataset of text–image pairs
  - VQGAN+CLIP (https://creator.nightcafe.studio/text-to-image-art)
    - Generate artworks from short text descriptions
- Current Underlying Technologies:
  - NLP: Natural Language Processing
  - Neural Networks
  - Artificial Intelligence
- Current challenges
  - Responsible AI:
    - May promote disinformation/deep fakes with created images
    - Bias and misinformation/miscommunication through images

FUTURE SCOPE - MODEL IMPROVEMENT & APPLICATIONS



FUTURE SCOPE - MODEL IMPROVEMENT & APPLICATIONS :

Image editing tool - Used it to clean up a photo taken and generate a higher quality image, and the results look more impressive than Photoshop.

Human life improvement - Help people express themselves visually in ways they may not have been able to before.

Replace encoding module with bigger/deeper model architecture, so, it can generate more meaningful information from text.

Current models suffer from different type of biases. It can generate Images which can have social and cultural biases, so, we can filter/remove the biases and AI model will be more ethical.

Robotics – Help humans understand how AI systems see and understand our world.