Ordered List of Projects (From Most to Least Desired)

1. PhotoNode project
2. Deep Learning-Driven 3D Reconstruction, Data Synthesis, and Size Estimation for Road Signs
3. Reinforcement Learning for Road Sign Detection and False Positive Elimination

1. **Proposed PhotoNodes Project: Intelligent Simple Object Detection and Anonymous Privacy Protection Feature for unclassified images/photographs loaded to a project and gallery Swipe or Blur rule**

Why I Want to Work on This Project:
I find this project because of its emphasis on synthetic data generation, a field I have hands-on experience in through my work with Conditional Categorical GANs (CCGAN). The opportunity to apply and expand this knowledge in a privacy-focused application excites me, as it combines technical challenges in computer vision, deep learning, and privacy protection with real-world impact. Additionally, the integration of multi-modal learning techniques offers a unique chance to work on innovative solutions for blending visual and textual data, which aligns closely with my research interests.

Experience:

- **Research**: I worked on a Wafer Project (CNN, GAN), addressing the challenge of imbalanced data in semiconductor wafer fabrication systems (SWFSs). Defective wafer maps were underrepresented, leading to misidentifications. To tackle this, I implemented a Conditional Categorical GAN (CCGAN) to generate high-fidelity wafer map simulations for each defect pattern, augmenting the dataset and improving recognition accuracy. This experience equipped me with a strong foundation in synthetic data generation and domain-specific model fine-tuning.

- I also have experience fine-tuning models for low-resource languages through Parameter-Efficient Fine-Tuning (PEFT), where I worked on adapting multilingual models like mBART to perform effectively on underrepresented languages, further enhancing their domain-specific utility.

- Developed a Reasoning and Act Agent with a Multi-Modal Contextual Retrieval-Augmented Generation (RAG) system to optimize circuit designs. The system efficiently provides designs with specifications, NETLISTs, graphs, and tables while comparing them with previously available designs. Leveraged multi-modal embeddings using image, graph, and table summaries to enhance retrieval performance. Incorporated original images as context for multi-modal models. (In the process of publishing at IEEE International Conference on LLM-Aided Design, 2025).

These have experience in multi-modality, synthetic data generation, and fine-tuning models for domain-specific applications.

Strategy to Approach This Project

1. Synthetic Data Generation:

- Leverage GANs and Diffusion Models:
    - Use GANs like StyleGAN2 or CCGAN to generate high-quality synthetic credentials with customizable layouts, fonts, sizes, and alphanumeric combinations. These models are particularly useful for generating diverse outputs with controlled variations.
    - Diffusion Models: Once trained on a relevant dataset, these models can generate specific types of credential data using text prompts, allowing flexibility in the type and style of credentials produced. For instance, a text prompt like *"Event badge with blue border and white background"* can yield the desired output.
- Domain Randomization for Robustness:
    - Apply random transformations like rotation, scaling, lighting changes, and occlusion to make the synthetic data more robust.
    - Combine synthetic credentials with real-world backgrounds (e.g., event photos) to create hybrid images that simulate realistic conditions. Tools like OpenCV and PIL can be used for such augmentations.

2. Deep Learning Model Development:

- Credential Detection with CNNs:
    - Use pre-trained CNN-based object detection models like YOLOv8 or Faster R-CNN to locate credentials within images. Fine-tune these models on the synthetic and hybrid dataset to improve detection accuracy for diverse credential layouts.
    - Use OCR models for accurate text recognition. These models can handle alphanumeric text variations and work seamlessly across multiple languages and font styles.
    - Fine-tune the text recognition model to enhance its performance on credential-specific text, such as names and event details.
- Privacy Protection:
    - Implement automated privacy masking for sensitive information by applying image processing techniques like Gaussian blurring or pixelation.

    This approach combines cutting-edge synthetic data generation techniques with advanced deep learning and multi-modal learning strategies to develop a robust and adaptable system for credential detection and privacy protection.