

Thanks for your interest in the lab, and **sorry for the delay**—we couldn't get back to you sooner.  
**Deadline: Saturday, 11:59 PM (local time).**

You **must use Google Colab** and publish to a **public GitHub repo containing only:**

- q1.ipynb
- q2.ipynb
- README.md

Also **report your best results** via the provided Google Form.

You may use any resources, but you're **expected to know every line** of your code. If selected, the interview will test your understanding of the paper, core deep-learning concepts, Python skills, and proficiency with packages like NumPy/PyTorch—plus a **live-coding** component.

---

### Q1 — Vision Transformer on CIFAR-10 (PyTorch)

**Goal.** Implement a ViT and train on **CIFAR-10** (10 classes). Your objective is to achieve the **highest possible test accuracy**. You are free to try various improvements and tricks to improve performance.

**Note:** You must only use Colab

Paper: **An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (Dosovitskiy et al., ICLR 2021).**

#### Requirements

- Patchify images, add learnable positional embeddings, prepend a **CLS** token, stack Transformer encoder blocks (MHSA + MLP with residual + norm), classify from CLS.
- In README.md, include: how to run in Colab, the config for your **best** model, and a tiny results table (Overall classification test accuracy).

#### (Bonus — optional analysis)

Any concise analysis earns bonus marks (e.g., patch size choices, depth/width trade-offs, augmentation effects, optimizer/schedule variants, **overlapping vs. non-overlapping patches**, etc.). Keep it short and crisp. The analysis should also be the part of the README.md

---

### Q2 — Text-Driven Image Segmentation with SAM 2

**Goal.** For a **single image**, perform **text-prompted segmentation** of a chosen object using **SAM 2**.

#### What to show in q2.ipynb

- Load image → accept a **text prompt** → convert text to region seeds (e.g., via GroundingDINO/GLIP/CLIPSeg or similar) → feed seeds to **SAM 2** → display the final mask overlay.
- Keep it runnable end-to-end on Colab (install cells at top).
- In README.md, briefly describe your pipeline and note limitations.

**(Bonus — video extension)**

If you also demonstrate **text-driven video object segmentation** (10–30s clip) by propagating masks across frames with SAM 2, you'll receive additional bonus marks.

---

**Submission Rules (all required)**

- **Colab only.** Both notebooks must execute from top to bottom on Colab (GPU).
  - **GitHub repo contains only** q1.ipynb, q2.ipynb, README.md.
  - **Google Form:** submit your best CIFAR-10 Average Accuracy (%) (**in percentage**) from Q1 and your repo link.
  - **Note: The Submission must be made via the google form only. Don't send personal emails**
-