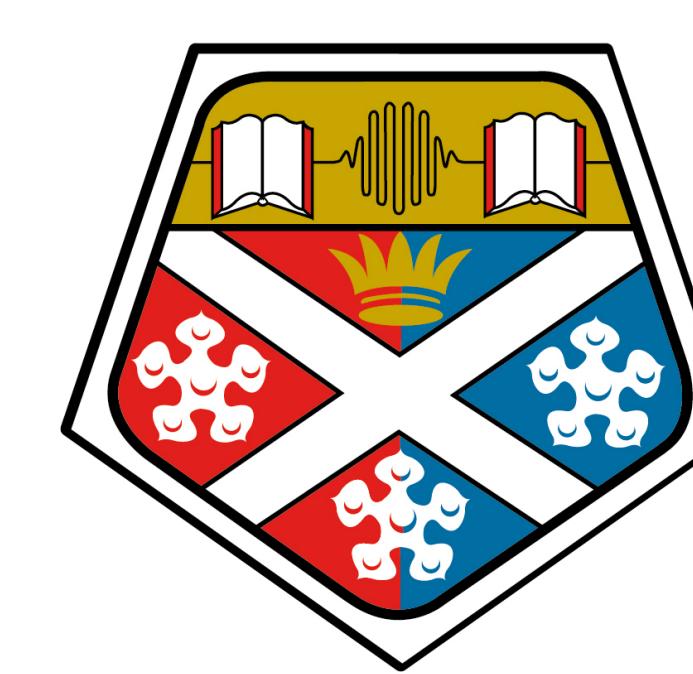


Advancing Assessment Practices in CS Education through AI-Generated Visual Test Cases

Mohamed Elawady

Department of Computer and Information Sciences, University of Strathclyde, Glasgow, UK

mohamed.elawady@strath.ac.uk



University of
Strathclyde
Science

Introduction

Context. Recent advances in generative AI enable high-fidelity image synthesis from text or reference images. This opens up opportunities to modernise assessment in PG courses in AI.

Aim. Integrate AI-generated visual test cases into computer vision coursework to:

- Explore robustness beyond public saturated benchmarks.
- Discourage overfitting to pre-defined test subsets.

Idea. This work proposes a three-stage framework: (1) Input specification, (2) Diffusion-based generation, (3) Automated validation & human-in-the-loop verification.

Literature Review

- Diffusion models (e.g., ImageGen, DALL-E, LDM) produce high-quality images from prompts or references.
- SOTA vision models achieve near-ceiling performance on public datasets, potentially leading to redundant learning experiences when coursework replicates them.
- Promising HE practices in using generative AI in educational assessments.

Contributions

- **Assessment-based synthesis pipeline** for generating curriculum-aligned images.
- **Validation loop** combining baseline models (classification: ResNet/ViT, detection: YOLO/RetinaNet) with tutor review.
- **Case studies** for classification & detection to guide adoption in HE teaching.

Conclusions and Future Work

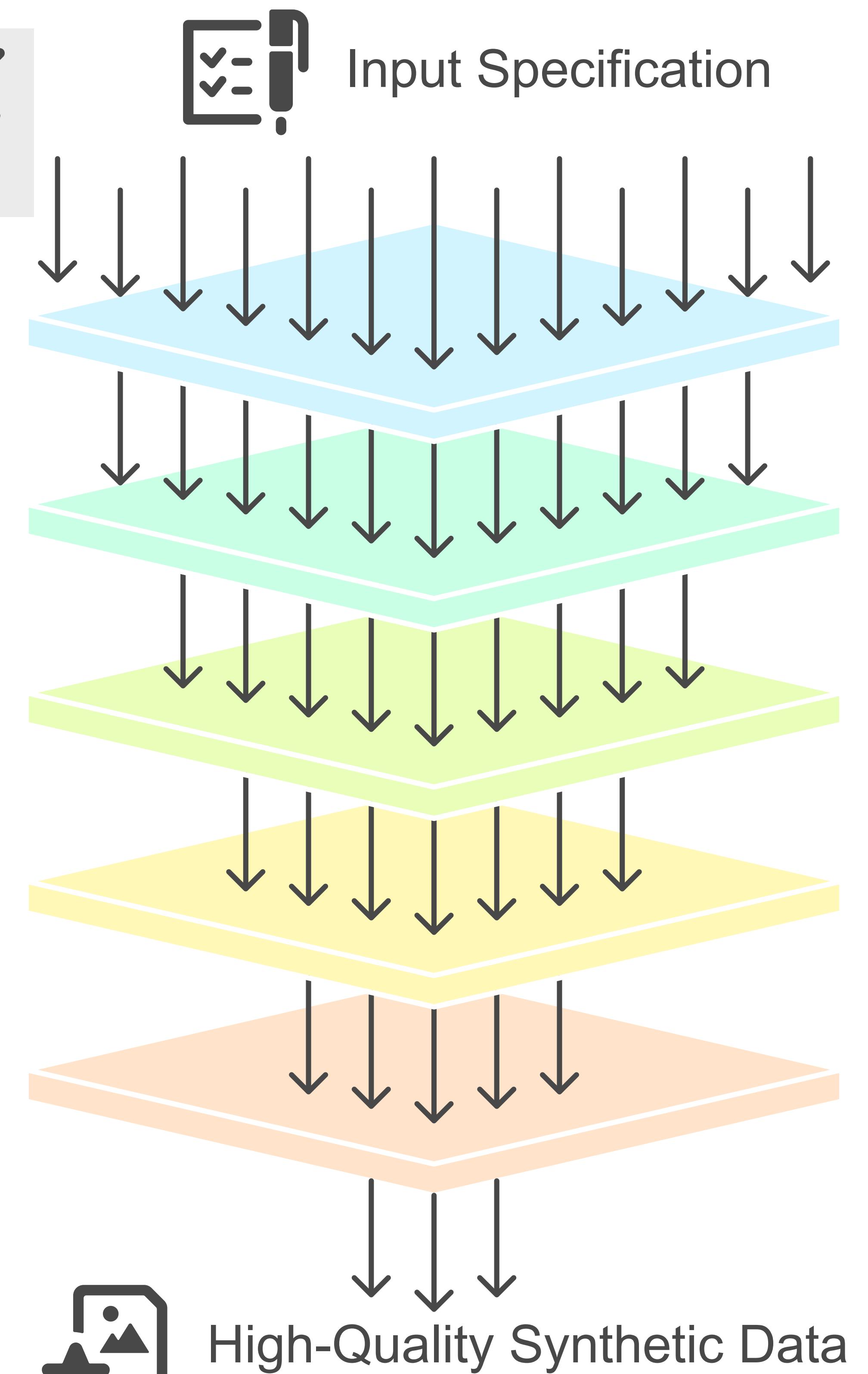
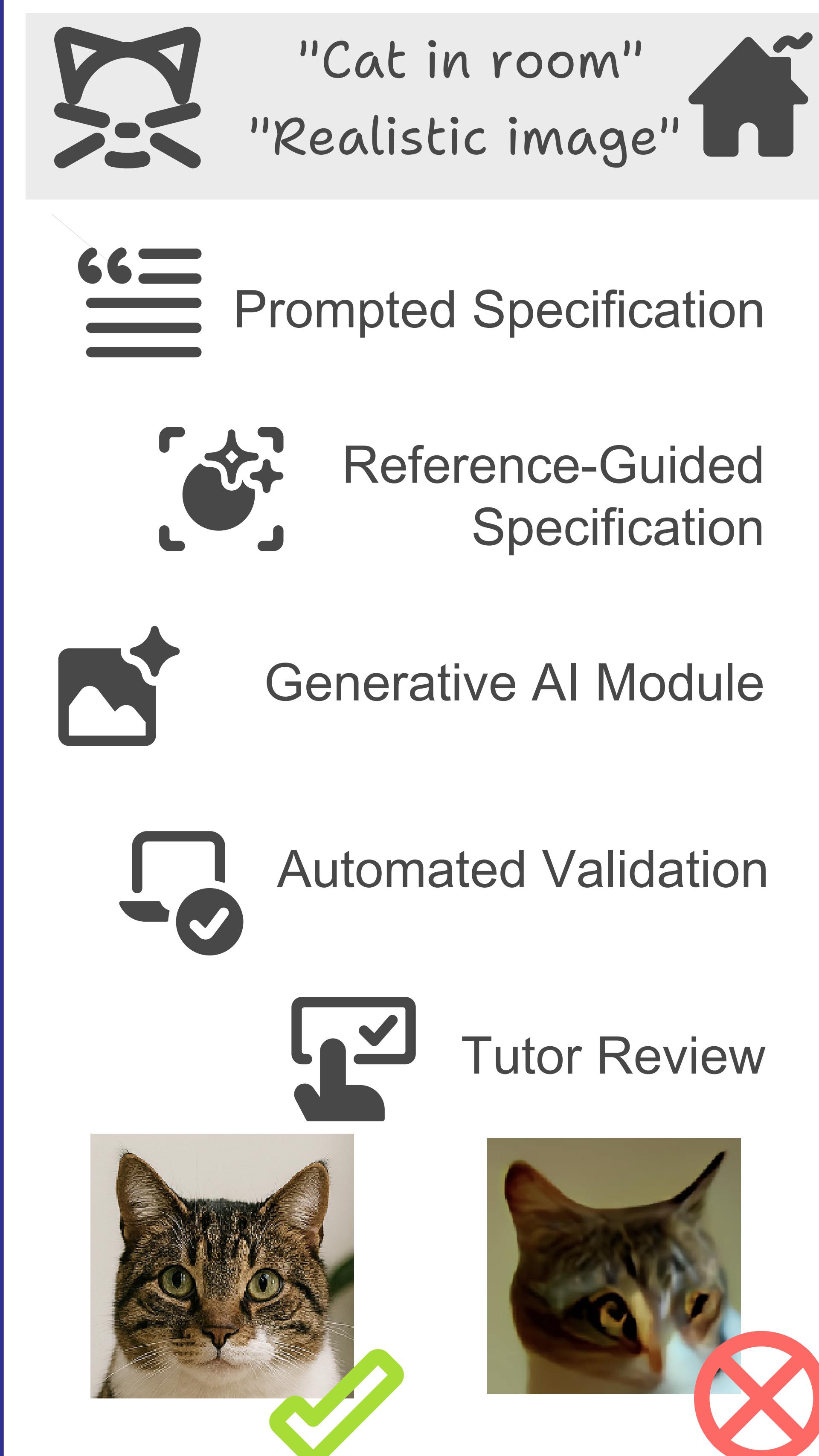
Summary. Generative visual technologies can provide challenging, semantically controlled test cases that align with course outcomes and reduce overfitting in coursework evaluation.

Next Steps.

- Extend to *video tagging* and *3D object scenarios*.
- Explore *NLP* tasks (sentiment, topic classification) for broader assessment coverage.
- Share reproducible prompt banks across AI-based modules and cohorts.

Acknowledgement. ChatGPT and Copilot (grammar and style suggestions), Napkin (diagram generation), Claude and ChatGPT (vibe coding), Gradio and HF spaces (demo).

Proposed Framework



Case Study I - Classification



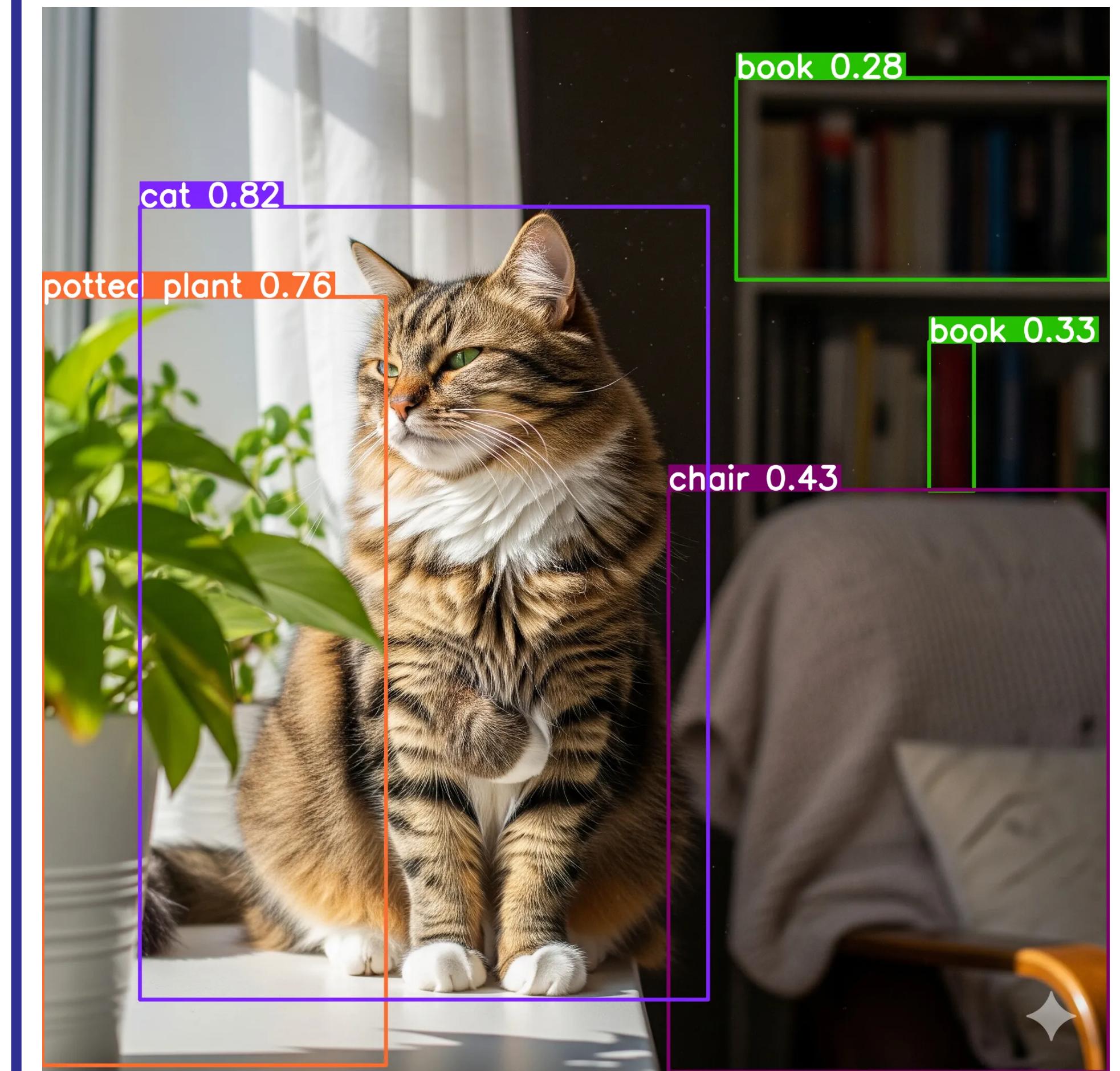
ResNet-18

Egyptian cat (40.09%)
tabby (31.38%)
tiger cat (24.99%)

ConvNeXt-Tiny

Egyptian cat (56.75%)
tiger cat (8.77%)
tabby (5.12%)

Case Study II - Detection



YOLOv8n

cat (83.0%)
vase (58.1%)
potted plant (45.0%)

YOLOv12n

cat (81.7%)
potted plant (75.9%)
chair (42.8%)

References

- Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33 (2020): 6840-6851.
- Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- Ramesh, Aditya, et al. "Zero-shot text-to-image generation." International conference on machine learning. Pmlr, 2021.
- Kaladaras, Leonora, Hope O. Akaeze, and Mark D. Reckase. "Developing valid assessments in the era of generative artificial intelligence." Frontiers in education. Vol. 9. Frontiers Media SA, 2024.
- Charles, Kellie A., et al. "AI in action: Changes to student perceptions when using generative artificial intelligence for the creation of a multimedia project-based assessment." European Journal of Pharmacology 998 (2025): 177508.

QR Code

