Contrat doctoral – ED Galilée

Titre du sujet :

- Unité de recherche : Laboratoire de Traitement et Transport de l'information (L2TI, UR 3043)
- Discipline : Traitement de l'information
- Direction de thèse : Aladine CHETOUANI
- Co-encadrant : Zuheng MING
- Contact : aladine.chetouani@univ-paris13.fr, zuheng.ming@univ-paris13.fr
- Domaine de recherche : IA et Santé
- Mots clés : Deep Learning, Perception, Gaze, Multimodal, Vision, Language, Medical imaging, Medial Reports, LLM, reinforcement learning, Eye-tracker

---

TITLE: Expert-Gaze-Guided Vision-Language Models for Reliable Automated Radiology Reports

Context:

Recent vision-language models (VLMs) such as GPT-4V [1] and LLaVA [2] have sparked interest in automated radiology reporting, yet three barriers still block routine clinical use: 1/ **Misplaced visual focus –** models attend to artefacts or normal anatomy while missing subtle, clinically relevant findings; 2/ **Hallucinated statements –** reports may invent or misclassify lesions because image evidence is weakly grounded and 3/ **Low clinician trust –** training on static image–text pairs hide how radiologists actually read images, offering no transparent confidence signal. Radiologists' eye-gaze naturally reveals where experts look and how they reason. Capturing this signal during routine reading lets us turn gaze into a live feedback channel for reinforcement learning [3], guiding VLMs to "see" like experts and to speak with greater clinical fidelity.

Objectives

1. **Gaze and VLM** – fuse real-time fixation information with image deep representation so the model's attention aligns with radiologists' visual search patterns.
2. **Gaze and reinforcement learning** – train with two rewards: *Visual-grounding reward* (encourages overlap between model attention and expert gaze) and *Diagnostic-preference reward* (thumbs-up/-down on generated sentences)
3. **Uncertainty-aware deployment** [4]– introduce a *Gaze–Report Consistency Score* that combines spatial alignment and language plausibility to flag reports that deserve extra review.
4. **Continual human-in-the-loop learning** – stream radiologists' on-the-fly corrections (edited text or gaze re-annotations) back into the model, preventing drift and catastrophic forgetting.

Key Innovations:

1. **Gaze as a grounding signal** – converts passive eye-tracking into active supervision that links pixels with diagnostic language.
2. **Dual-channel RL** – unifies gaze alignment and binary expert feedback to curb hallucinations and enforce evidence-based wording.
3. **Explainability by design** – delivers an intuitive trust cue (consistency score) without interrupting clinical workflow.
4. **Self-improving clinical AI** – closes the loop between expert knowledge and model updates, keeping the system current with evolving standards.

By embedding expert gaze and feedback directly into the learning process, this project seeks to produce VLMs that generate radiology reports clinicians can understand, trust, and adopt in everyday practice.

It is worth noting that we have already formalized a collaboration with Northwestern University's Department of Radiology (Bagci Lab) [5] to run an eye-tracking study with practicing radiologists, building on our recent ETRA-2025 [6] and CVPR-W-2025 [7] papers that showed medical gaze can significantly improve AI grounding and accuracy.

[1] Achiam, J., et al. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.
[2] Liu, Haotian, et al. "Visual instruction tuning." Advances in neural information processing systems 36 (2023): 34892-34916.
[3] Hu, Mingzhe, et al. "Reinforcement learning in medical image analysis: Concepts, applications, challenges, and future directions." Journal of Applied Clinical Medical Physics 24.2 (2023): e13898.
[4] Zou, K., et al. (2023). A review of uncertainty estimation and its application in medical imaging. Meta-Radiology, 1(1), 100003.
[5] https://bagcilab.com/
[6] David C Wong, Bin Wang, Gorkem Durak, Marouane Tliba, Mohamed Amine Kerkouri, **Aladine Chetouani**, et al. "Shifts in Doctors' Eye Movements Between Real and AI-Generated Medical Images", ACM ETRA 2025 (to appear)
[7] David Wong, Bin Wang, Gorkem Durak, Marouane Tliba, Akshay Chaudhari, **Aladine Chetouani**, et al., "Eyes Tell the Truth: GazeVal Highlights Shortcomings of Generative AI in Medical Imaging", CVPR-W, 2025 (to appear)