

# Sedigheh (Sarah) Eslami

PhD Candidate

Computer Science and Engineering

Hasso Plattner Institute, Germany

✉ sedigheh.eslami@hpi.de

🔗 Google Scholar

🐙 GitHub Profile

🌐 Homepage

## EDUCATION

---

### • Hasso Plattner Institute / University of Potsdam

Aug. 2022 – Present

PhD in Computer Science

- **Primary Focus:** Multi-modal LLMs, Vision-Language Foundation Models, Contrastive Representation Learning, Self-supervised Learning.
- **Additional Interests:** Visual Question Answering, Open-Vocabulary Object Detection.

### • Max Planck Institute / Saarland University

Oct. 2014 – Sept. 2017

MSc in Computer Science

### • Amirkabir University of Technology

Oct. 2009 – Feb. 2014

BSc in Computer Software Engineering

## PUBLICATIONS

---

### • Mitigate the Gap: Improving Cross-Modal Alignment in CLIP [\[Paper\]](#) [\[Code\]](#)

ICLR 2025

- Reduced the modality gap via sharing parameters in the vision and language towers. Furthermore, proposed a semantically-regularized intra-modality contrastive loss objective function.
- As a result of our modifications, the average cosine similarity of image-text pairs increased by more than 0.2. Moreover, the zero-shot classification as well as linear probing improved by 5%. Additionally, the multi-modal retrieval results got enhanced by 2%.

### • ArtQuest: Countering Hidden Language Biases in ArtVQA [\[Paper\]](#) [\[Code\]](#) [\[Presentation\]](#)

WACV 2024

- Discovered the hidden language bias in the ArtVQA benchmark datasets using CLIP and T5 models.
- Curated a novel dataset for ArtVQA that overcomes the language bias. Our work also provides various baselines using CLIP for our proposed dataset.

### • PubMedCLIP [\[Paper\]](#) [\[Code\]](#) [\[Presentation\]](#)

EACL 2023

- Fine-tuned CLIP using image-caption pairs from PubMed articles for the task of medical visual question answering.
- Our model gained an absolute increase of 5-10% accuracy in comparison to the previous pre-trained encoders.

## PROFESSIONAL EXPERIENCE

---

### • Jina AI Research Scientist Internship

Jan. 2025 – Jun. 2025

Berlin

- Working on multimodal LLMs with high-resolution image, visual token compression and sparse attentions.

### • Scientist Researcher

Aug. 2022 – Present

AI Service Center, Berlin/Brandenburg

- Developing novel AI models in the field of vision-language foundation model pre-training and publishing at top-tier conferences such as EACL and WACV.
- Providing consultations about multi-modal deep learning, especially on contrastive learning, open-vocabulary object detection, video representation learning and visual question answering.

### • Data Scientist

Jan. 2020 – Jul. 2022

Data4Life, Berlin

- Developed BERT-based NLP solutions for text classification as well as analytic solutions on clinical EHR data.
- Scientific management of projects for anonymizing clinical discharge letters using statistical NLP.
- Worked on medical visual question answering using Bilateral-Branch Networks for handling long-tail distributions.

### • Software Engineer

Aug. 2017 – Dec. 2019

Data4Life, Berlin

- Developed react-based front-end and scala back-end for web applications concerning zero-knowledge storage of patients' medical data.
- Developed proof-of-concepts using NodeJS for pseudo-anonymization of patients' medical data compliant to HIPAA.
- Strong knowledge of microservice architecture design as well as REST API.

### • SAP Security Research Internship

Apr. 2017 – Jul. 2017

Karlsruhe

- Worked on privacy-preserving NLP with a focus on differential-privacy for text classification in author detection.

## TECHNICAL SKILLS AND INTERESTS

---

**Languages:** Python, Java, NodeJS, Scala, Javascript.

**Libraries:** PyTorch, HuggingFace, PyTorch Lightning, Tensorflow, Numpy, Pandas, SKlearn, Wandb.

**Dev Tools:** VScode, Git, Github, Docker, IntelliJ

**Databases:** MongoDB, Relational Database(MySQL, PostgreSQL), SAP HANA In-Memory Database

**Soft Skills:** Problem Solving, Getting Things Done, Self-learning, Presentation, Adaptability, Teamwork, Creativity

## HONORS

---

- **Ranked 3rd in the International ImageCLEF VQA-Med Challenge** *Sep. 2021*
  - Developed an ensemble model based on Bilateral-Branch Networks (BBN) with cumulative learning for medical visual question answering with long-tail answer distribution. Our ensemble model achieved an accuracy improvement of 5% in comparison to the baseline BBN model.
- **Rewarded with the student fellowship for high-ranked students at Saarland University** *Nov. 2016*