

# Bohua Peng

Ethnicity: Chinese Degree: Master Tel: +86 18800295676  
E-mail: bp1119@ic.ac.uk  
Personal Page: <https://pengbohua.github.io/Pengbohua/>  
Specialism: Artificial Intelligence and Machine Learning  
Research Interests: Knowledge Graph Representation; QA; Multimodality;



## EDUCATION

Imperial College London Master Sep.2019 – Nov.2021  
Major: Artificial Intelligence and Machine Learning  
Modules: NLP, Computational Optimization, Computer Vision, Reinforcement learning, Advanced Database, Variational Inference and Gaussian Process etc.  
Rank: Distinction 72

## PUBLICATIONS

**Bohua Peng**, Zhiyuan Liu et al. UniLink: Unifying Link Prediction and Entity Linking with Contrastive Learning of Bidirectional Transformer Encoders. (Under Review)

**Bohua Peng**, Mobarakol Islam, Mei Tu. Angular Gap: Reducing the Uncertainty of Image Difficulty through Model Calibration. (Accepted by ACM Multimedia 2022)

Yujin Wang, Jing Zhang, **Bohua Peng** et al. SRCB at SemEval-2022 Task 5: Pretraining Based Image to Text Late Sequential Fusion System for Multimodal Misogynous Meme Identification. (ACL Semantic Evaluation Workshop)

## SELECTED PROJECTS

- **Misogyny detection (SemEval Task5 Top1)** 2021,12-2022,02
  - Background: Weak semantic consistency and domain shift exist in this image-text classification.
  - To handle statistical bias between train and test distribution, we aggregate CLIP and a decision tree algorithm XGBoost as late-stage interaction in our final submission.
  - I have tried finetuning CLIP and Zero-Shot with Prompt Tuning. I have also tried UNITER as pretrained VL model and ELECTRA as PLM.
- **Conditional visual question generation system (Duolingo)** 2020,6 – 2020,12
  - To generate questions based on CEFR level, I use two VAEs to fuse questions, answers and CEFRs. This is a B plan for the final design using GPT-Neo.
  - I investigate posterior collapse with KL term annealing, FreeBits and BN-VAE.
- **Evaluating humor in edited news (Semeval2020 Task7)** 2019,12- 2020,02
  - Background: Compare the funniness of the edited versions of the original headline.
  - Data augmentation: I apply NER with BERT and CRF. I finetune a GPT-2 to generate context.
  - I finetune RoBERTa with MarginLoss and BCE against multi-head GRU, XGBoost and SVM.

## WORK EXPERIENCE

THUNLP (Tsinghua)	RA	May.2022 – Now
Samsung Research	Engineer	Nov.2021 – May. 2022
Duolingo	Remote Intern	July.2021 – Oct.2021
Samsung	Staff	Oct.2017 – August.2019

## PROGRAMMING&LANGUAGE

- IELTS 7.5 Reading 8.5 Listening 8.5 Writing 6.5 Speaking 6, GRE 316.
- Proficient: Huggingface Transformers, SentenceBERT, Detectron, YOLO, timm, EasyOCR, PaddleOCR, PyTorch, Python, Bash, Latex, Docker, Django, labelstudio, labelme.
- Competitive: Virtuoso, ElasticSearch, C++, Tensorflow, PaddlePaddle, CSS, html5, MySQL.

## AWARDS

SemEval 2022 Task5 Top1 over 500+ teams  
National Scholarship 2016  
Third Prize of the Third National IoT Competition over 1000+ teams  
Second Prize of 2016 Shanghai Dancing Competition