Guokan Shang

LLMs, NLP, ML, AI
Research Scientist, Ph.D.

🔗 **https://shang.tech/aboutme**
✉ guokan.shang@hotmail.com
🗺 French, English, Chinese
🏠 Guyancourt, France
📅 https://calendly.com/guokan-shang

💼 Work experience & Internships

➤ **Lead Research Scientist**, MBZUAI France Lab, Mohamed bin Zayed University of Artificial Intelligence

- Conducting research on Generative AI.   Nov. 2023 - current
- Developing LLMs.

➤ **Chargé d'Enseignement - Temps Incomplet**, École Polytechnique   Sep. 2023 - current

- Teaching in part-time NLP/LLMs modules of the lectures (cours magistraux) and the lab sessions (travaux pratiques) of the following courses: MAP670N, INF582.

➤ **Research Scientist**, LINAGORA Labs   Dec. 2016 - Nov. 2023

- Our goal is to develop a state-of-the-art and open-source conversational AI platform: LinTO. I'm in charge of developing NLP pipelines and APIs, e.g., named entity recognition, keyphrase extraction, topic modeling, and summarization.
- Writing of research proposals to public funding organizations, e.g., ANR, Bpifrance.
- Conducting research related to LLM, NLP for conversations, and NLG evaluation, in collaboration with industrial and academic partners, under various funded research projects. (see ⚗ section)
- Publishing and presenting scientific articles at top-tier international conferences and journals.
- Advising / mentoring students, engineers and interns. (see 👥 section)

➤ **Ph.D. Candidate**, DaSciM team, École Polytechnique   Dec. 2016 - Jan. 2021

- Topic: *Spoken Language Understanding for Abstractive Meeting Summarization*
- Under the supervision of Michalis Vazirgiannis and Jean-Pierre Lorré, funded by French CIFRE Ph.D. fellowship and LINAGORA. I was also advised by Antoine J.-P. Tixier during this period.

➤ **Research Intern**, DaSciM team, École Polytechnique   Apr. 2016 - Sep. 2016

- Topic: *Approximate Reinforcement Learning with Mixture of Experts*
- Under the supervision of Nikos Tziortziotis.

➤ **Back-end Developer Intern**, Tessi   Feb. 2015 - Jul. 2015

- Topic: *Redesign and Migration of a Document Generation System in Service-oriented Architectures*

➤ **Research Intern**, MI&T Lab, Harbin Institute of Technology   Feb. 2013 - Jul. 2013

- Topic: *Statistical Machine Translation for Multilingual Information Retrieval*

🎓 Education

2016–2021	Ph.D.	Computer Science, Data and Artificial Intelligence	École Polytechnique
2015–2016	Master	Machine Learning, Information and Content	Paris-Saclay University
2013–2015	Engineer	Informatics and Information Systems	University of Technology of Troyes
2009–2013	Bachelor	Computer Science and Technology	Harbin Institute of Technology

⚗ Research projects & activities

➤ Member of SUMM-RE ANR project (2020–2024), with LPL, IRIT, LINAGORA, LIX.
The objective is to combine expertise in theories of discourse interpretation with recent developments in distant supervision to improve the automatic production of meeting summaries and minutes from spoken data. Additionally, we are building a 100 hour audio-video-text annotated corpus of multi-party, meeting-like interactions in French.

➤ Member of LinTO Bpifrance project (2018–2021), with IRIT, LAAS-CNRS, LINAGORA, LIX, Zelros. The goal was to develop a privacy-aware AI-powered open-source conversational meeting assistant for professional use in a corporate environment.

➤ Member of OpenPaaS:NG Bpifrance project (2015–2019), with LINAGORA, LIX, Loria, Nexedi, XWiki. The aim was to offer an open source alternative to Microsoft 365 & Google Suite, providing all the functionalities for enterprise collaborative work: mail box, instant messaging, video conferencing, etc.

➤ I serve as a long-term reviewer of the ACL Rolling Review (ARR) since Dec. 2021, and I have contributed > 20 reviews.

📄 Research papers

1. Shang, G, W Ding, Z Zhang, A Tixier, P Meladianos, M Vazirgiannis, and JP Lorré (2018). Unsupervised Abstractive Meeting Summarization with Multi-Sentence Compression and Budgeted Submodular Maximization. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics - **ACL 2018***.

2. Shang, G, A Tixier, M Vazirgiannis, and JP Lorré (2020). Energy-based Self-attentive Learning of Abstractive Communities for Spoken Language Understanding. In: *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing - **AACL-IJCNLP 2020***.

3. Shang, G, A Tixier, M Vazirgiannis, and JP Lorré (2020). Speaker-change Aware CRF for Dialogue Act Classification. In: *Proceedings of the 28th International Conference on Computational Linguistics - **COLING 2020***.

4. Kamal Eddine, M, G Shang, A Tixier, and M Vazirgiannis (2022). FrugalScore: Learning Cheaper, Lighter and Faster Evaluation Metrics for Automatic Text Generation. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics - **ACL 2022***.

5. Guo, Y, G Shang, V Rennard, M Vazirgiannis, and C Clavel (2023). Automatic Analysis of Substantiation in Scientific Peer Reviews. In: ***Findings** of the Association for Computational Linguistics: **EMNLP 2023***.

6. Kamal Eddine, M, G Shang, and M Vazirgiannis (2023). DATScore: Evaluating Translation with Data Augmented Translations. In: ***Findings** of the Association for Computational Linguistics: **EACL 2023***.

7. Rennard, V, G Shang, D Grari, J Hunter, and M Vazirgiannis (2023). FREDSum: A Dialogue Summarization Corpus for French Political Debates. In: ***Findings** of the Association for Computational Linguistics: **EMNLP 2023***.

8. Rennard, V, G Shang, J Hunter, and M Vazirgiannis (2023). Abstractive Meeting Summarization: A Survey. *Transactions of the Association for Computational Linguistics - **TACL Journal**, co-first authorship*.

</> Technical skills

➤ Python, Scikit-learn, Keras, Transformers, Apache Spark | spaCy, NLTK, Gensim, SentenceTransformers| Git, SQL, HTML, Linux | Celery, Redis, Flask, Plotly Dash, etc.

➤ 4th contributor of an open source repository on GitHub 🖧: DesignPatternsPHP with > 21.3k ★

👥 Advisees

Doctoral students

➤ Virgile Rennard (École Polytechnique & LINAGORA), *meeting summarization*   Sep. 2021 - current
(1 long paper at TACL Journal)

➤ Yanzhu Guo (École Polytechnique), *NLG evaluation*   Oct. 2022 - current
(1 pre-print)

➤ Moussa Kamal Eddine (École Polytechnique), *NLG evaluation*   Feb. 2021 - Dec. 2022
(2 long papers at EACL Findings and ACL)

Interns

➤ Jules Peyrat & Rayan Autones (LINAGORA), *dialogue summarization*   Jun. 2023 - Sep. 2023
(polytechnicien 2nd year project)