LUCAS WEBER

✉ lucas.weber@upf.edu | 🔗 LinkedIn | ⌗ LucWeber | 🌐 webpage |

Education

Ph.D. Computational Linguistics and NLP   2019 – 2024
University Pompeu Fabra   *Barcelona, ES*
- Thesis: *Interpreting the Learning Dynamics of Language Models*
- DTCL-Scholarship
- supervision by:
    * Dr. Elia Bruni (University of Osnabrück)
    * Dr. Dieuwke Hupkes (FAIR)

RMSc. Brain and Cognitive Sciences; Track: Cognitive Science   2016 – 2018
University of Amsterdam   *Amsterdam, NL*
- with distinction ('cum laude') | equiv. GPA (US): 4.0
- Thesis-project: *MRC-CBU, University of Cambridge*
    * Empirical paper, Review paper

BSc. Psychology   2013 – 2016
University of Vienna   *Vienna, AT*
- equiv. GPA (US): 3.9
- ERASMUS-exchange: *University of Groningen*

Talks and Presentations

Fraunhofer IIS: Invited talk   2024
- Presenting work on efficient benchmarking

Samaya AI: Invited talk   2024
- Presenting work on *tinyBenchmarks*

DLBCN 2023 | *Poster presentation*   2023
- Presenting work on *Mind the instructions*

GenBench @ EMNLP 2023 | *Poster presentation*   2023
- Presenting work on *Mind the instructions* and *The ICL consistency test*

CoNLL 2023 | *Oral presentation*   2023
- Presenting work on *Mind the instructions: a holistic evaluation of consistency and interactions in prompt-based learning*
- Honourable mention (rated as one of the 4 best papers at the conference)

ENS Paris, Invited talk at CoML   2022
- Presenting work on *Teaching with language-commentaries*

ELLIS-NLP Workshop 2021 | *Poster-presentation*   2021
- Presenting work on *Meeting Formal Linguistics and Optimization*

EVIL (UPF/FAIR-internal meeting): Invited talk   2021
- Presentation *Similarity Probing in the Linguistic Task-Space*

EACL 2021 | *Oral-presentation*   2021
- Presenting work on project *Language Modelling as a Multi-Task Problem*

TAB (Seminar series at UPF/UvA/CIMeC) | *Oral presentation*   2021
- Presentation *Language Modelling as a Multi-Task Problem*

EVIL (UPF/FAIR-internal meeting) | *Oral presentation*   2019
- Presentation *Continual Learning in Language Emergence in artificial conversational agents*

Notable publications

- Weber, L., Jumelet, J., Bruni, E. and Hupkes, D. (2024) *Interpreting Language Models via Task Spaces*, ACL 2024

- Maia Polo, F., Weber, L., Choshen, L., Sun, Y., Xu, G. and Yurochkin, M. (2024) *tinyBenchmarks: evaluating LLMs with fewer examples*, ICML 2024

- Weber, L., Bruni, E. and Hupkes, D. (2023) *The ICL consistency test*, GenBench 2023

- Weber, L., Bruni, E. and Hupkes, D. (2023) *Mind the instructions: a holistic evaluation of consistency and interactions in prompt-based learning*, CoNLL 2023

- Weber, L., Jumelet, J., Bruni, E. and Hupkes, D. (2021) *Language Modelling as a Multi-Task Problem*, EACL 2021

Pre-prints / Under review

- Maia Polo, F., Xu, R., Weber, L., Mírian, S., Bhardwaj, O., Choshen, L., Flavio Melo de Oliveira, A., Sun, Y. and Yurochkin, M.(2024) *Efficient multi-prompt evaluation of LLMs*, under Review

- Weber, L., Jumelet, J., Michel, P., Bruni, E. and Hupkes, D. (2023) *Curriculum Learning with Adam: The Devil Is in the Wrong Details*, arXiv

Open-source contributions

tinyBenchmarks   2024
- Subsamples (n=100) of the MMLU, open LLM leaderboard, helm-lite and AlpacaEval datasets paired with an item-response model to do efficient benchmarking for LLMs. Find on the huggingface hub at tinyBenchmarks/*.

ICL consistency test   2023
- A NLP task for the consistency evaluation of LLMs. Find the code on github.

ICL consistency dataset   2023
- A NLP dataset the consistency evaluation of LLMs. Find on the huggingface hub at LucasWeber/icl_consistency_test.

2048 RL environment   2020
- A OpenAI gym environment to train agents to play the game of 2048 and visualise their play. Find on the code on github.

Teaching

Computational Semantics | *Graduate Teaching Assistance*   2021 - 2022
- DTCL, University Pompeu Fabra
- Language of instruction: English
- Content: Introduction to NLP methods for linguistics research

Recursos TIC | *Teaching Assistance*   2021 - 2022
- DTCL, University Pompeu Fabra
- Language of instruction: Spanish/Catalan
- Content: IT-resources for linguists

Descripció i Anàlisi 2 | *Teaching Assistance*   2020 – 2022
- DTCL, University Pompeu Fabra
- Language of instruction: English/German/Catalan

- Content: Phonology/Morphology/comp. Linguistics

Descripció i Anàlisi 1 | *Teaching Assistance*   2020
- DTCL, University Pompeu Fabra
- Language of instruction: English/German
- Content: Syntax

Foundations of Neural and Cognitive Modelling | *Graduate Teaching Assistance*   2017
- IIS, University of Amsterdam
- Language of instruction: English
- Content: Dynamical systems

Organization

UR-LING workshop | *Lead organizer*   2021
- United workshop for different sub-fields of linguistics at University Pompeu Fabra

Supervision

Thesis projects | *Co-supervision*   2022
- Leon Schmid at IKW, University of Osnabrück
- Simon Helling at IKW, University of Osnabrück

Skills

Languages:
German (native);
English (fluent);
Dutch (fluent);
Spanish (fluent);
French (basic);
Catalan (basic);
Latin
Programming languages: Python, MATLAB, R
Frameworks: Pytorch, Tensorflow; DeepSpeed, fairscale, FSDP
Relevant skills: High-performance computing (HPC), CUDA, multi-GPU / multi-machine training, model sharding, data parallel