

Simone BALLOCCU



PROFILE

Computer scientist with 6+ years of experience in NLP & AI research, and author of 17 peer-reviewed NLP publications. I was involved in several EU-funded projects, including Horizon 2020, ERC and EU Erasmus Traineeship. I implemented and evaluated several AI resources for behaviour change, including conversational agents and datasets for nutrition counselling and motivational interviewing.

CONTACT DETAILS

@balloccu@ufal.mff.cuni.cz
☎ +39 345 595 1827
🌐 uccollab.github.io
📍 Room 230, IMPAKT Building,
V Holešovičkách 2, 180 00 Praha 8,
Prague, CZ

PERSONAL INFORMATION

Citizenship: **Italian**
Languages: **Italian** (Native),
English (IELTS-C1)

SKILLS

- Python, Java, C, R, Scala, Javascript
- Machine learning model training, analysis, bias and fairness assessment.
- Implementation and evaluation of NLP systems in the context of behaviour change, counselling and mental health.

RESEARCH EXPERIENCE

POSTDOC 2023.07–pres.
📍 Charles University (CZ)
➤ **ERC-2021-STG, NG-NLG project**
♦ **Topics:** Neuro-symbolic NLG approaches; Large Language Models controllability.

VISITING RESEARCHER 2022.05–2022.10
📍 Philips Research (NL)
➤ **Horizon 2020, Philhumans project**
♦ **Topics:** Use of Large Language Models in Counselling; AI-driven activity tracking for behaviour change.

VISITING RESEARCHER 2020.07–2021.05
📍 Philips Research (NL)
➤ **Horizon 2020, Philhumans project**
♦ **Topics:** AI-oriented counselling datasets.

ASSISTANT RESEARCHER 2018.12–2019.09
📍 University of Cagliari (IT).
➤ **Sardegna Ricerche, DoUtDes project**
♦ **Topics:** AI-driven marketing campaigns for companies

ASSISTANT RESEARCHER 2017.01–2019.09
📍 University of Cagliari (IT).
➤ **Sardegna Ricerche, OKGraph project**
♦ **Topics:** Unsupervised NLP algorithm for knowledge extraction from free text.

EDUCATION

PHD in Artificial Intelligence 2019.10–2023.05
📍 University of Aberdeen (UK).
➤ **Marie Skłodowska-Curie ESR - Philhumans project (ESR-7)**
➤ Thesis title: *Emotion-aware and tailored diet coaching with Natural Language Generation.*

M.Sc in Computer Science 2017.10–2019.09
📍 University of Cagliari (IT).
➤ Thesis title: *Unsupervised Hypernym Discovery from plain text.*

B.Sc in Computer Science 2013.10–2017.09
📍 University of Cagliari (IT).
➤ Thesis title: *Unsupervised learning of IS-A relations from plain text.*

TEACHING

TEACHING ASSISTANT Winter 2023 - Spring 2024
📍 Charles University (CZ).
➤ **NPFL099: Statistical Dialogue Systems**
➤ **Responsibilities:** Marking and running lectures.

DEMONSTRATOR Spring 2021 & Spring 2022
📍 University of Aberdeen (UK).
➤ **CS551H: Natural Language Generation**
➤ **Responsibilities:** Demonstrating and running lab practicals.

- Extrinsic and Intrinsic NLP system evaluation; Human-in-the-loop and Expert-in-the-loop evaluation.
- LLMs prompt engineering including zero and few-shot prompting, prompt injection and red-teaming.

INVITED LECTURER

Spring 2020 & Spring 2021

📍 University of Cagliari (IT).

> SC/o132/EN: Natural Language Processing

> Responsibilities: Running invited lectures and related labs.

MENTORING

Experience co-supervising students at both Bachelor's and Master's level

| | |
|--------------------------------------|-----------|
| 📍 Charles University (CZ) | present |
| > 1 Bachelor's student (in progress) | |
| > 1 Master's student (in progress) | |
| 📍 University of Aberdeen (UK) | 2023-2024 |
| > 1 Master's student | |
| 📍 University of Utrecht (NL) | 2021-2022 |
| > 1 Master's student | |
| 📍 University of Cagliari (IT) | 2020-2021 |
| > 2 Bachelor's students | |

ORGANISATIONAL EXPERIENCE

| | |
|--|------|
| CO-CHAIR | 2024 |
| <ul style="list-style-type: none">• The 2nd Workshop on Practical LLM-assisted Data-to-Text Generation (Practical D2T 2024 @ INLG 2024)• The International Workshop on Human-Centered Modeling and Adaptation for Digital Transformation (HUMAD 2024 @ ACM UMAP 2024)• The 4rd Workshop on Human Evaluation of NLP Systems (HumEval 2024 @ LREC-COLING 2024) | |
| ORGANIZING COMMITTEE MEMBER | 2023 |
| <ul style="list-style-type: none">• The 3rd Workshop on Human Evaluation of NLP Systems (HumEval 2023 @ RANLP 2023)• The 16th International Natural Language Generation Conference (INLG 2023)• The 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDial 2023) | |

PROPOSALS

META LLAMA IMPACT GRANT (AWAITING FEEDBACK)

2024

Title: *LLama 2 for efficient communication with autistic students*

Members: *Charles University (CZ), Institute of Neuropsychiatric Care (CZ), University of Aberdeen (UK), LuxAI (LU).*

Amount: \$500,000.

Role: Coordinated proposal writing.

EPSRC (AWAITING FEEDBACK)

2023

Title: *Using Large Language Models in Dietary Chatbots*

Members: *University of Aberdeen (UK), Charles University (CZ), University of Manchester (UK), University of Stirling (UK).*

Amount: £1,193,716.

Role: Contributed to proposal writing.

RESEARCH AND PUBLICATIONS

Peer-reviewed publications: 17

Google h-index: 7

Google i-10 index: 4

Citations: 149

Link to the complete list of publications: [Google Scholar](#)

> Selected peer-reviewed publications

- **"Leak, Cheat, Repeat: Data Contamination and Evaluation Malpractices in Closed-Source LLMs"**

Simone Balloccu, Patrícia Schmidtová, Mateusz Lango, Ondřej Dušek.

In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), Mar 2024

CORE Rank: A

🏆 **Best Non Publicized Paper Award**

- **Are Experts Needed? On Human Evaluation of Counselling Reflection Generation.**

Zixiu Wu, **Simone Balloccu**, Ehud Reiter, Rim Helaoui, Diego Reforgiato Recupero, Daniele Riboni.

In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Jul 2023.

CORE Rank: A*

- **Anno-MI: A Dataset of Expert-Annotated Counselling Dialogues.**

Zixiu Wu, **Simone Balloccu**, Vivek Kumar, Rim Helaoui, Ehud Reiter, Diego Reforgiato Recupero, Daniele Riboni.

In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2022.

CORE Rank: B