

# Simone BALLOCCU



## PROFILE

Computer scientist with 6+ years of experience in NLP & AI research, and author of 19 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

- LAB LEADER** **2024.10–present**  
📍 *TU Darmstadt (DE)*  
> **NLP for expert domains (ExpNLP) group**  
◊ **Topics:** agents for expert-AI cooperation
- POSTDOC** **2023.07–2024.09**  
📍 *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)*  
> **NPFLo99: Statistical Dialogue Systems**  
> **Responsibilities:** Marking and running lectures.

- 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.

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

INVITED LECTURER **Spring 2020 & Spring 2021**  
 📍 *University of Cagliari (IT).*  
 > **SC/0132/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**
  - > 2 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**

- **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**

- **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)**

## RESEARCH AND PUBLICATIONS

---

Peer-reviewed publications: 19

Google h-index: 8

Google i-10 index: 5

Citations: 207

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