Roberta Rocca, PhD

Who am I?

Hi there! I am a researcher working at the intersection between data science, applied NLP and neuroinformatics, with international experience and publications in several high-impact scientific outlets. I have worked with small and big data across multiple domains (e.g., clinical ML, cognitive science, social sciences, humanitarian sector), done quite some applied machine learning, and contributed to open-source software. I am a passionate and avid learner, a creative problem-solver, and team worker. I am currently looking for a data science/ML/NLP role in the industry and seeking to join a dynamic, ambitious and fast-paced team.

rbrrcc@gmail.com
rbroc
RockBerta
rbroc.github.io

Skills

Programming languages: Python; SQL; R; Matlab; bash; PERL
Main libraries & frameworks: Keras; PyTorch; Scikit-learn; numpy/scipy; pandas; Git/GitHub; Docker
Areas of Expertise: Machine Learning; Statistical Modeling; Software Development; Experimental Research
Languages: Italian (native); English (fluent); Danish (fluent); French (intermediate)

Experience

Postdoctoral Researcher
December 2020 – ongoing
School of Culture and Society, Aarhus University, Denmark
Focus areas: NLP, Clinical Machine Learning, Data Science
Main responsibilities
- Engineered text- and audio-based machine learning models for mental health diagnostics
- Created and analyzed large-scale social media datasets through APIs and custom ML pipelines
- Coordinated cross-national data collection and analysis in an international consortium
- Developed out-of-the-box methodologies for large-scale analysis of multilingual text data

Postdoctoral Researcher
December 2019 – July 2022
Psychoinformatics Lab, University of Texas at Austin, USA
Focus areas: Neuroinformatics, Deep Learning, Natural Language Processing
Main responsibilities
- Co-developed Neuroscout, an open-source Python platform for end-to-end analysis of brain imaging data
  - Designed and run large-scale multi-dataset statistical analyses for workflow validation
  - QA-testing, documentation and outreach
  - Developed and implemented automated ML pipelines for video and audio annotation
- Engineered, trained and evaluated deep learning architectures for text modeling and classification
- Conducted research on model evaluation for cognitive and brain sciences
- Published research outcomes in high-impact international journals
- Co-developed and maintained *pliers* (https://github.com/PsychoInformaticsLab/pliers)

Predictive Analytics Data Fellow
June – August 2021
Centre for Humanitarian Data, United Nations
Focus area: Predictive Modeling, Dynamical Systems Analysis; Data-Driven Policy-Making
Main responsibilities
- Critically reviewed scientific literature on dynamic causal modeling
- Formulated recommendations for data-driven humanitarian policy-making using causal modeling
- Communicated findings in a blog and a an extended report

Education

PhD, Cognitive Neuroscience
September 2016 – October 2019
Aarhus University, Denmark

MSc, Cognitive Science and Computational Linguistics
September 2014 – July 2016
University of Trento, Italy

Additional research experience

Visiting Researcher, Department of Applied Mathematics and Computer Science; Technical University of Denmark
Visiting Researcher, Institute of Cognitive Science and Technologies; National Research Council, Italy
Research Assistant, Department of Experimental Psychology, University College London, United Kingdom
Research Assistant, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Scientific Output

Some of my recent academic papers

- de la Vega, A.*, **Rocca, R.* (co-first)**, Blair, R., Mentch, J., Markiewicz, C., Ghosh, S., Poldrack, R., Yarkoni, T. (2022), Neuroscout: a unified platform for generalized and reproducible fMRI research, accepted in *eLife*, pdf available at this link

- **Rocca, R.**, Yarkoni, T. Language as a fingerprint: A self-supervised approach to text-based user modeling using transformers (under review, but happy to share upon request)

- **Rocca, R.**, de la Vega, A. Evaluating the role of non-lexical markers in LLMs' language modeling behavior (same as above)

- **Rocca, R.**, Yarkoni, T. (2021). Putting psychology to the test: rethinking model evaluation through benchmarking and prediction, *Advances in Methods and Practices in Psychological Science*, pdf available online at this link

- **Rocca, R.**, (2021) Complex Systems Modeling for Humanitarian Action: Methods and Opportunities, Research Report for the United Nations' *Centre for Humanitarian Data*, pdf available online at this link

- **Rocca, R.**, Coventry, K. R., Tylén, K., Staib, M., Lund, T. E., & Wallentin, M. (2020). Language beyond the language system: dorsal visuospatial pathways support processing of demonstratives and spatial language during naturalistic fast fMRI, *NeuroImage*, pdf available online at this link

See full list of publications at: https://scholar.google.com/citations?user=EjB67tAAAAAJ&hl=en

Some of my non-academic data-related articles on news media (in Italian)

- Algoritmi di classe, *Doppiozero*, available here
- Chi ha paura dei data scientists? Numeri e pandemia, *Doppiozero*, available here
- Dati, miti, stati, *L'identità di Clio*, available here

Presentations, teaching and outreach

Over the past few years, I have participated in and organized scientific conferences and events, and taught university courses at both bachelor, masters and staff level.

In 2019, I organized a workshop on Natural Language Processing at Aarhus University, bringing together academic experts in the field from several countries.

I have taught courses in Statistical Methods (bachelor-level), R Programming (staff course at Aarhus University), Cognitive Science (master-level), Cognitive Neuroscience (bachelor- and master-level), Social and Cultural Dynamics (bachelor-level).

I have presented my work – among other venues – at the Annual Meeting of the Organization for Human Brain Mapping, at the Annual Conference of the Society for the Neurobiology of Language, at the Annual Meeting Cognitive Science Society, at the Conference of the Society for Complex Systems. I have also been invited to present my work at joint seminar series held by McGill University, Karolinska Institute and University of Toronto, and at the Centre for Humanitarian Data.

Other interests

When I am not doing science, I play bass in a band, write fiction, and watch art-house movies. I love learning new languages -- I am currently proficient in four and aiming at expanding the repertoire. I spend most of my holiday time going on high-stamina road trips and long hikes.