Silvio Amir

Curriculum Vitae

Khoury College of Computer Science
Northeastern University
Boston, MA 02115, United States
✆ +1 (512) 917 9507
✉ s.amir@northeastern.edu
🖥 https://www.silvioamir.com/

Research Topics

Social Media Analysis, Natural Language Processing, Machine Learning, and Information Retrieval

Applications in Healthcare, Public Health and the Social Sciences

Education

2018   **Ph.D. Information Systems and Computer Engineering**,
Instituto Superior Técnico da Universidade de Lisboa, Lisboa, Portugal.

Thesis   *Agile Social Media Analysis with Neural Networks*

Advisor   Mário J. Silva

Co-Advisor   Paula C. Carvalho

Funding   Ph.D. scholarship awarded by Fundação para Ciência e Tecnologia (FCT)

2009   **Master Computer Science and Engineering (MA)**,
Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal.
Specialization in Software Engineering

Thesis Title   *Desenvolvimento e Reengenharia de Aplicações Web de Suporte ao Negócio e Integração com Sistemas de Business Intelligence*

(Development and Re-Engineering of Web Applications for Business Support and Integration with Business Intelligence Systems)

Advisor   Luís Antunes

2007   **Licenciate Computer Science and Engineering (Lic)**,
Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal.

Appointments

2021   **Assistant Professor (tenure-track)**, *Khoury College of Computer Sciences*,
Northeastern University, Boston (MA), USA.

2020–2021   **Research Fellow**, *Khoury College of Computer Sciences*,
Northeastern University, Boston (MA), USA.

Advisor   Byron Wallace

2018–2020   **Research Fellow**, *Center for Language and Speech Processing*,
Johns Hopkins University, Baltimore (MD), USA.

Advisor   Mark Dredze

Refereed Publications

Conference

1. Silvio Amir, Jan-Willem van de Meent, and Byron C Wallace. "On the Impact of Random Seeds on the Fairness of Clinical Classifiers". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2021, pp. 3808–3823

2. Aaron Mueller, Zach Wood-Doughty, Silvio Amir, Mark Dredze, and Alicia Lynn Nobles. "Demographic representation and collective storytelling in the me too twitter hashtag activism movement". In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (2021), pp. 1–28

3. Paula Carvalho, Bruno Martins, Hugo Rosa, Silvio Amir, Jorge Baptista, and Mário J. Silva. "Situational Irony in Farcical News Headlines". In: *Computational Processing of the Portuguese Language*. Springer International Publishing, 2020, pp. 65–75

4. Manya Wadhwa, Silvio Amir, and Mark Dredze. "Aligning Public Feedback To Requests For Comments On Regulations.gov". In: *International Conference on Web and Social Media (ICWSM)*. 2020

5. Silvio Amir, Mark Dredze, and John W Ayers. "Mental Health Surveillance over Social Media with Digital Cohorts". In: *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*. 2019, pp. 114–120

6. Silvio Amir, Glen Coppersmith, Paula Carvalho, Mário J Silva, and Byron C Wallace. "Quantifying Mental Health from Social Media with Neural User Embeddings". In: *Journal of Machine Learning Research, W&C Track* (2017)

7. Silvio Amir, Byron C Wallace, Hao Lyu, Paula Carvalho, and Silva Mário J. "Modelling context with user embeddings for sarcasm detection in social media". In: *Proceedings of the Conference on Natural Language Learning (CoNLL)* (2016)

8. Wang Ling, Chris Dyer, Alan W Black, Isabel Trancoso, Ramon Fermandez, Silvio Amir, Luis Marujo, and Tiago Luis. "Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sept. 2015, pp. 1520–1530

9. Wang Ling, Yulia Tsvetkov, Silvio Amir, Ramon Fermandez, Chris Dyer, Alan W Black, Isabel Trancoso, and Chu-Cheng Lin. "Not All Contexts Are Created Equal: Better Word Representations with Variable Attention". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sept. 2015, pp. 1367–1372

10. Ramón Astudillo, Silvio Amir, Wang Ling, Mario Silva, and Isabel Trancoso. "Learning Word Representations from Scarce and Noisy Data with Embedding Subspaces". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Beijing, China: Association for Computational Linguistics, July 2015, pp. 1074–1084

Journal

11. Silvio Moreira, David S Batista, Paula Carvalho, Francisco M Couto, and Mario J Silva. "Tracking politics with POWER". in: *Program* 47.2 (2013), pp. 120–135

Workshop

12. Darius Irani, Avyakta Wrat, and Silvio Amir. "Early Detection of Online Hate Speech Spreaders with Learned User Representations—Notebook for PAN at CLEF 2021". In: *CLEF 2021 Labs and Workshops, Notebook Papers*. Ed. by Guglielmo Faggioli, Nicola Ferro, Alexis Joly, Maria Maistro, and Florina Piroi. CEUR-WS.org, Sept. 2021. URL: http://ceur-ws.org/Vol-2936/paper-174.pdf

13. Silvio Amir, Ramon F Astudillo, Wang Ling, Mário J Silva, Isabel Trancoso, and Rua Alves Redol. "INESC-ID at SemEval-2016 Task 4-A: Reducing the Problem of Out-of-Embedding Words". In: *Proceedings of SemEval* (2016), pp. 238–242

14. Pedro Saleiro, Silvio. Amir, Mário J. Silva, and Carlos Soares. "Tracking Political Opinion on the Web". In: *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. Oct. 2015

15. Silvio Amir, Wang Ling, Ramón Astudillo, Bruno Martins, Mario J. Silva, and Isabel Trancoso. "INESC-ID: A Regression Model for Large Scale Twitter Sentiment Lexicon Induction". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 613–618

16. Ramón Astudillo, Silvio Amir, Wang Ling, Bruno Martins, Mario J. Silva, and Isabel Trancoso. "INESC-ID: Sentiment Analysis without Hand-Coded Features or Linguistic Resources using Embedding Subspaces". In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Denver, Colorado: Association for Computational Linguistics, June 2015, pp. 652–656

17. Silvio Amir, Miguel B. Almeida, Bruno Martins, João Filgueiras, and Mário J. Silva. "TUGAS: Exploiting unlabelled data for Twitter sentiment analysis". In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland: Association for Computational Linguistics and Dublin City University, Aug. 2014, pp. 673–677

18. Joao Filgueiras and Silvio Amir. "POPSTAR at replab 2013: Polarity for reputation classification". In: *CLEF 2013 Eval. Labs and Workshop Online Working Notes*. 2013

19. Silvio Moreira, João Filgueiras, Bruno Martins, Francisco Couto, and Mário J. Silva. "REACTION: A naive machine learning approach for sentiment classification". In: *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Atlanta, Georgia, USA: Association for Computational Linguistics, June 2013, pp. 490–494

20. Silvio Moreira, David Batista, Paula Carvalho, Francisco M Couto, and Mário J Silva. "POWER-Politics Ontology for Web Entity Retrieval". In: *Advanced Information Systems Engineering Workshops*. Springer. 2011, pp. 489–500

Invited Talks and Seminars

2021 **Towards Precision Social Media Analysis**, *Speed Data-ing by NULab for Texts, Maps and Networks*, College of Social Sciences and Humanities Northeastern University, Boston (MA), USA.
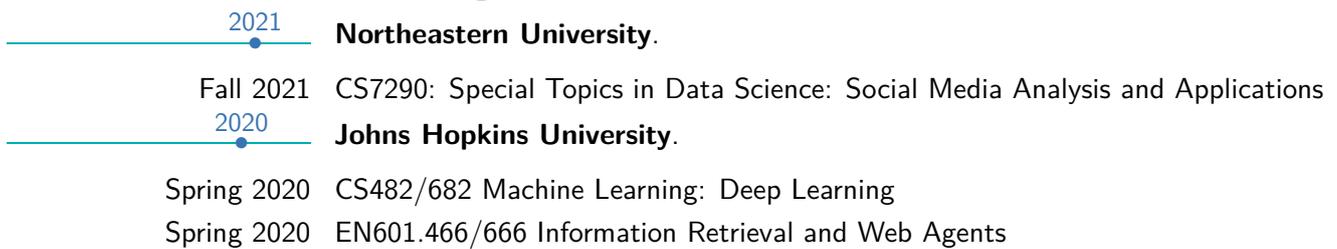
2021 **User Representations for Human-Centered AI**, *Khoury College of Computer Sciences*, Northeastern University, Boston (MA), USA.

2017 **Empowering Social Media Analysis for Social Good**, *CLSP Seminar*, Johns Hopkins University, Baltimore (MD), USA.
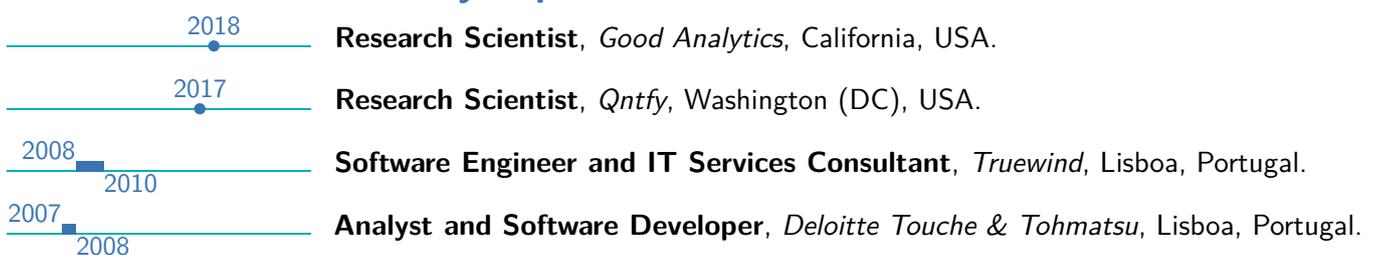
2016 — **Representation Learning for Agile Social Media Analysis**, *SIGAI CNC*, Northeastern University, Boston (MA), USA.

2015 — **Towards Social Media Analysis for Low-Resource Languages**, *Priberam Machine Learning Lunch Seminars*, Instituto Superior Técnico, Lisboa, Portugal.

2014 — **How to build a Twitter Sentiment Classifier: part II**, *Sapo Labs Talks*, Sapo Labs Lisboa, Lisboa, Portugal.

2014 — **How to build a Twitter Sentiment Classifier**, *Sapo Labs Talks*, Sapo Labs Lisboa, Lisboa, Portugal.

2014 — **Análise de Sentimento em Social Media**, *Workshop em Jornalismo Computacional*, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal.

Teaching

2021 — **Northeastern University**.

Fall 2021 — CS7290: Special Topics in Data Science: Social Media Analysis and Applications

2020 — **Johns Hopkins University**.

Spring 2020 — CS482/682 Machine Learning: Deep Learning

Spring 2020 — EN601.466/666 Information Retrieval and Web Agents

Research Experience

2019 — **Visiting Researcher**, *School of Medicine*, University of California San Diego, La Jolla, San Diego (CA), USA.

Advisor — John W. Ayers

2016 / 2018 — **Research Assistant**, *CCIS - Machine Learning Lab*, Northeastern University, Boston (MA), USA.

Advisor — Byron Wallace

2016 — **Visiting Researcher**, *School of Information*, University of Texas at Austin, Austin (TX), USA.

Advisor — Byron Wallace

2013 / 2018 — **Research Assistant**, *Data Management and Information Retrieval Group (DMIR)*, INESC-ID Lisboa, Lisboa, Portugal.

Advisor — Mário J. Silva

2011 / 2013 — **Junior Research Assistant**, *Large-Scale Informatics Systems Lab (LASIGE)*, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal.

Advisor — Mário J. Silva

Industry Experience

2018 — **Research Scientist**, *Good Analytics*, California, USA.

2017 — **Research Scientist**, *Qntfy*, Washington (DC), USA.

2008 / 2010 — **Software Engineer and IT Services Consultant**, *Truewind*, Lisboa, Portugal.

2007 / 2008 — **Analyst and Software Developer**, *Deloitte Touche & Tohmatsu*, Lisboa, Portugal.

Service

Peer-Reviewing

Reviewer and secondary reviewer for ACL, NAACL, EACL, EMNLP, NeurIPS, IJCAI and IJCNLP conferences, and for the Springer Machine Learning Journal

Mentoring

2020-2021 Avyakta Wrat, *Indian Institute of Technology Bombay*

2020-2021 Darius Irani, *Johns Hopkins University*

Fall 2019 Kennedy Thompson, *Garrison Forest School/Johns Hopkins University* (Women In Science and Engineering Program)

2019 **Women In Science and Engineering Program (WISE)**,
Johns Hopkins University, Baltimore (MD), USA.

Conference Organization

2017 **Volunteer**, *Machine Learning for Healthcare*,
Northeastern University, Boston (MA), USA.

2015 **Volunteer**, *Conference on Empirical Methods in Natural Language Processing*,
Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal.

Summer Schools

2017 **Laboratory Monitor**, *Lisbon Machine Learning Summer School (LxMLS 2017)*,
Instituto Superior Técnico, Lisboa, Portugal.

2015 **Laboratory Monitor**, *Lisbon Machine Learning Summer School (LxMLS 2015)*,
Instituto Superior Técnico, Lisboa, Portugal.

2014 **Laboratory Monitor**, *Lisbon Machine Learning Summer School (LxMLS 2014)*,
Instituto Superior Técnico, Lisboa, Portugal.

Volunteer Work

2011 **European Voluntary Service Program**, *Multikultura Egyesület*, Hungary.
2012 Worked as a volunteer in a refugee camp in Bicske and in a day care center in a problematic neighborhood of Budapest, teaching English and basic computer skills.