Antônio Theóphilo

e-mail: antonio.theophilo@gmail.com
https://antoniotheophilo.github.io

PRESENTATION

I am a Ph.D. student in Artificial Intelligence and Natural Language Processing, pushing the limits of AI in textual data applications.

EDUCATION

UNICAMP - University of Campinas   2018 - present
Ph.D. in Computer Science - Artificial Intelligence (Machine Learning and NLP)
Thesis: Fighting "Fake News" Through Authorship Attribution and Phylogeny Analysis
GPA: 4.0/4.0   Advisor: Anderson Rocha, Ph.D.

PUC-Rio - Pontifical Catholic University of Rio de Janeiro   08/2003 - 07/2005
M.Sc. in Computer Science - Computer Networks and Distributed Systems
Dissertation: Solutions for CORBA Firewall/NAT Traversal
GPA: 9.5/10.0   Advisors: Markus Endler, Ph.D. and Renato Cerqueira, Ph.D.

UFRN - Federal University of Rio Grande do Norte   1998 - 2002
B.Sc. in Computer Science
GPA: 9.14/10.0   Graduated with honors

EXPERIENCE - INDUSTRY/ ACADEMY

CMU - Carnegie Mellon University   Pittsburgh/USA
Visiting Scholar Researcher   March 2020 - Present
At CMU, I am developing part of my Ph.D. research (phylogeny analysis of microblog messages) in the Language Technologies Institute/School of Computer Science, under the supervision of Prof. Yulia Tsvetkov, Ph.D.

Microsoft   Redmond/USA
Ph.D. Intern   December 2019 - February 2020
At Microsoft, I conducted research and development of models in the fields of Machine/Deep Learning and Natural Language Processing to improve the Bing user experience. The models were deployed in production, with excellent results, increasing the queries coverage to a great extent. Manager: Skanda Maiya, M.Sc. Mentor: Henrique Dantas, Ph.D.

CTI - Center for Information Technology Renato Archer   Campinas/Brazil
Researcher (on leave)   July 2013 - present
My main activity is participation in researches and projects in the fields of machine learning, pattern recognition, forensics, and information security.
Relevant projects:

- Working Group "Artificial Intelligence Applied to Police Investigation": Multi-institutional working group that has the objective of applying artificial intelligence techniques to police investigations, in particular, corruption and child pornography.

- SmartDiF: Development of a tool for digital documents authentication capable of including state-of-the-art methods for solving different forensic problems.

- Colossus: Development of a system for collection, storage, visualization, and analysis of social media data to support police intelligence activities. The system was used at the Rio 2016 Olympic Games and requested by a federal government agency.

- Oráculo: Development of a specification and functional prototype for a system to collect, store, visualize, and analyze information security incidents. The project was requested by the Brazilian Federal Police and had its prototype used at the 2014 Football World Cup held in Brazil.

FIB   Bauru/Brazil
Professor   April 2019 - August 2019
Professor of Supervised Learning in a *Lato Sensu* Machine Learning graduate program.

FGV Law School   São Paulo/Brazil
Invited Professor   August 2018 - September 2018
Professor of Python Fundamentals for undergraduates Law students in a program for fostering *Legal Tech* companies and innovation initiatives.

Mackenzie Presbyterian University   São Paulo/Brazil
Invited Professor   August 2013 - December 2013
Professor of Performance Techniques in a *Lato Sensu* Computer Science graduate program.

Freeddom Technology and Services   São Paulo/Brazil
Head of Information Security and Computer Infrastructure   August 2005 - July 2013
I led a team of 12 computer science professionals responsible for all the computer infrastructure of a worldwide 24x7 mobile financial service.

PUC-Rio - Pontifical Catholic University of Rio de Janeiro   Rio de Janeiro/Brazil
Teaching Assistant   September 2004 - July 2005
Teaching Assistant in a *Lato Sensu* Software Engineering graduate program.

State Finance Department of Rio de Janeiro
Software Engineer   September 2004 - July 2005 Design and development of solution for digital signature of official documents issued by web applications. Technologies used: Java, Servlets, JSP, JavaScript, EJB, JCE, JSSE, and JBoss.

Embratel
Software Engineer   April 2004 - September 2004 Design and development of a Single Sign-On (SSO) solution for web corporate applications. Technologies used: Java, Servlets, EJB, JAAS, and JBoss.

CEF - Caixa Econômica Federal
Software Engineer   February 2003 - August 2003 Developer of Loterias Project. This project aimed at replacing all the national lottery management system, with more than 50,000 terminals spread across the country. Technologies used: EJB, Servlets, JSP, RUP, UML, JMS, XML, Swing, JBoss, and IBM MQSeries (Mainframe interface).

RESEARCH PROJECTS

Déjà Vu: Feature-Space-Time Coherence from Heterogeneous Data for Media Integrity Analytics and Interpretation of Events   2017 - present
Project funded by FAPESP - São Paulo Research Foundation - that aims at developing techniques capable of taking high-level conclusions of events from heterogeneous media (Internet and private data) using Artificial Intelligence. I participate in this project as a Ph.D. student in the research and development of social media textual analyses.

DeepEyes: Visual Computing and Machine Intelligence Techniques for Digital Forensics and Electronic Surveillance   2016 - present
Project funded by CAPES - Coordination for the Improvement of Higher Level Education - to develop solutions for many forensic problems. My participation in this project is in the research of authorship attribution of microblog messages.

Research and Development of Forensic Techniques in Digital Media   2015 - present
Project coordinated by me at CTI Renato Archer aiming to produce research and development of forensic techniques for high-level analyses of digital media, especially, texts.

Integrade   2004 - 2005
This project, headed by USP - University of São Paulo - and with other member universities as PUC-Rio, targeted the development of a grid computing middleware to coordinate

the processing of idle machines distributed around the globe. The project used CORBA for communication with the remote agents, and my participation was in studying and solving the issues of firewall traversal of these elements, proposing better ways to deal with this scenario. This work was part of my M.Sc. dissertation.

LANGUAGES

English: Full proficiency (C1 Level by TOEFL ITP Test - 630/677)

French and Spanish: Basic skill in reading and listening

Portuguese: Native speaker

RELEVANT PUBLICATIONS

JOURNALS & MAGAZINES

- ROCHA, ANDERSON; SCHEIRER, WALTER J.; FORSTALL, CHRISTOPHER W.; CAVALCANTE, THIAGO; **THEOPHILO, ANTONIO**; SHEN, BINGYU; CARVALHO, ARIADNE R. B.; STAMATATOS, EFSTATHIOS. Authorship Attribution for Social Media Forensics. IEEE T-IFS - Transactions on Information Forensics and Security, v.12, p.5-33, 2017. IEEE. ISSN: 1556-6013. DOI: `http://dx.doi.org/10.1109/TIFS.2016.2603960` . 97 citations according to Google Scholar.

- REZENDE, EDMAR; RUPPERT, GUILHERME; **THEOPHILO, ANTONIO**; TOKUDA, ERIC K.; CARVALHO, TIAGO. Exposing Computer Generated Images by Using Deep Convolutional Neural Networks. Signal Processing: Image Communication, v.66, p. 113-126, 2018. Elsevier. ISSN: 0923-5965. DOI: `https://doi.org/10.1016/j.image.2018.04.006` . 12 citations according to Google Scholar.

CONFERENCE PAPERS

- SCHWARZ, STEPHANE; **THEOPHILO, ANTONIO**; ROCHA, ANDERSON. EMET: Embeddings From Multilingual-Encoder Transformer for Fake News Detection. *To Appear in* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, Barcelona, Spain. IEEE.

- **THEOPHILO, ANTONIO**; PEREIRA, LUÍS A. M.; ROCHA, ANDERSON. A Needle in a Haystack? Harnessing Onomatopoeia and User-specific Stylometrics for Authorship Attribution of Micro-messages. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, Brighton, United Kingdom. IEEE. ISSN: 1520-6149. DOI: `https://doi.org/10.1109/ICASSP.2019.8683747` . 1 citation according to Google Scholar.

- DE MENDONÇA, RICARDO RESENDE; DE FRANCO ROSA, FERRUCIO; **COSTA, ANTONIO CARLOS THEOPHILO**; BONACIN, RODRIGO; JINO, MARIO. OntoCexp: A Proposal for Conceptual Formalization of Criminal Expressions. 16th International Conference on Information Technology - New Generations (ITNG), p.43-48, 2019. Advances in Intelligent Systems and Computing, vol 800. Springer. ISBN: 978-3-030-14069-4. DOI: `https://doi.org/10.1007/978-3-030-14070-0_7` . 3 citations according to Google Scholar.

- REZENDE, EDMAR; RUPPERT, GUILHERME; CARVALHO, TIAGO; **THEOPHILO, ANTONIO**; RAMOS, FABIO; DE GEUS, PAULO. Malicious Software Classification Using VGG16 Deep Neural Network's Bottleneck Features. Information Technology - New Generations. p.51-59. Springer. ISSN: 978-3-319-77027-7. DOI: `https://doi.org/10.1007/978-3-319-77028-4_9` . 12 citations according to Google Scholar.

- DAHER, J.; PASTI, R.; **THEOPHILO, ANTONIO**; SILVA, R. Evolutionary Combinations of Text Classifiers: Investigating the Diversity of Contexts. II Symposium on Knowledge Discovery, Mining and Learning, 2014, São Carlos, Brazil. Proceedings of the II Symposium on Knowledge Discovery, Mining and Learning - KDMiLe, 2014.

- **COSTA, ANTONIO THEOPHILO**; ENDLER, MARKUS; CERQUEIRA, RENATO. Evaluation of Three Approaches for CORBA Firewall/NAT Traversal. On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE.

OTM 2005. Lecture Notes in Computer Science, v.3761, p. 923-940, Springer Berlin Heidelberg, 2005. DOI: `https://doi.org/10.1007/11575801_1` . 10 citations according to Google Scholar.

- NASCIMENTO, FERNANDO NEY DA COSTA; **THEOPHILO COSTA JÚNIOR, ANTONIO CARLOS**; DE PAULA, VIRGINIA C. C.; LEMOS DE SOUZA FILHO, GUIDO. Using Acme to Specify the Software Architecture of the OpenH323 Open Source Code. 2nd WSEAS International Conference On Multimedia, Internet And Video Technologies - WSEAS ICOMIV 2002, Skiathos - Grécia.

- **THEÓPHILO COSTA JÚNIOR, ANTÔNIO CARLOS**; NASCIMENTO, FERNANDO NEY DA COSTA; LEITE, LUIZ EDUARDO CUNHA; LEMOS DE SOUZA FILHO, GUIDO. Implementation of a Programming Interface for Multicast Services (Implementação de uma Interface de Programação para Serviços de Multicast). II Workshop RNP2 - High-speed Metropolitan Networks Working Group Workshop (Workshop Consórcio ReMAV), 2000.

REVIEWER	*IEEE Transactions on Information Forensics and Security - T-IFS (Impact Factor: 6.211)*
	IEEE International Workshop on Information Forensics and Security - WIFS
	Brazilian Symposium in Information Security and Computational Systems - SBSeg

AWARDS / GRANTS	*Google LARA - Latin America Research Awards*	2020
	Google LARA - Latin America Research Awards	2019
	Google LARA - Latin America Research Awards	2018
	Research Internships Abroad Grant - São Paulo Research Foundation - FAPESP	2020-21
	Ph.D. Research Grant - São Paulo Research Foundation - FAPESP	2018-21
	M.Sc. Grant - National Council for Scient. and Technol. Devel. - CNPq	2004-05
	M.Sc. Grant - Coord. for the Improv. of Higher Educ. Personnel - CAPES	2003-04
	Distinguished Student Award - Brazilian Society for Computing/SBC	2002
	Student Merit Medal - Best Computer Science Undergraduate - UFRN	2002

CERTIFICATIONS	*TOEFL ITP - ETS (Score: 630/677 - Level C1)*	2018
	Sun Certified Programmer for the Java Platform	2003
	University of Cambridge FCE - First Certificate in English (Level B2)	1995

TECHNOLOGIES *Programming Languages:* Python, C, C++, Java, Objective-C, Perl, Bash, Lua, R, Octave, Assembly
Machine Learning: scikit-learn, NLP (gensim, NLTK), TensorFlow/Keras, PyTorch, Theano/Lasagne, Caffe
Databases: MySQL (Standalone and Cluster), Oracle, SQL, Hibernate, JDBC, ODBC, MongoDB, BerkeleyDB, HDF5, Elasticsearch
Virtualization: Xen, Docker, VirtualBox
Information Security: SSH, SSL, JAAS, JSSE, JCE, Symmetric and Asymmetric Cryptography, Hashing, Digital Signature, Firewalls (IPTables e PF), VPN, IPSec
Operating Systems: Linux, OpenBSD, FreeBSD, Mac OS X, IOS, Windows
Typesetting: LaTeX
Misc: Cobbler, Puppet, Fabric, eXtreme Programming, UML, RUP, MDA, Optimization and Operational Research

ADDITIONAL INFORMATION

- Google Scholar: `https://scholar.google.com/citations?user=k5aJUOQAAAAJ`

- LinkedIn: `https://www.linkedin.com/in/antoniotheophilo/`

- ORCID: `http://orcid.org/0000-0003-1408-0745`