

TARIQ ALHINDI

www.cs.columbia.edu/~tariq

500 W 120th St, New York, NY 10027

✉ tariq@cs.columbia.edu ◊  [Tariq60](#) ◊  [tariqalhindi](#)

RESEARCH INTERESTS

Natural Language Processing: including Argument Mining & Discourse Analysis, Misinformation & Propaganda Techniques Detection, Automatic Fact-Checking, and Arabic Natural Language Processing.

EDUCATION

Ph.D. in Computer Science - NLP *2017 - 2022 (expected)*

Columbia University, New York, NY, USA

Thesis: Pragmatics-Infused Models for Argumentation and Misinformation Detection

Committee: Smaranda Muresan (advisor), Kathleen Mckweon, Julia Hirschberg, Daniel Preoțiu-Pietro (Bloomberg), and Preslav Nakov (QCRI).

MPhil in Computer Science - NLP *2017 - 2020*

Columbia University, New York, NY, USA

Coursework: NLP, Machine and Deep Learning, NLP in Social Context, Learning Theory

Candidacy Exam: A Review of Fact-Checking, Fake News Detection and Argumentation Mining.

Dual M.Sc. in Computer Science and Engineering Management *2011 - 2013*

Syracuse University, Syracuse, NY, USA

Coursework: Algorithms, Databases, AI, NLP, Project Management, Statistics, Economics

B.Sc. in Computer Engineering *2005 - 2010*

King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

PUBLICATIONS

11. **Tariq Alhindi**, Brennan Xavier McManus and Smaranda Muresan. [What to Fact-Check: Guiding Check-Worthy Information Detection in News Articles through Argumentative Discourse Structure](#). Proceedings of the The 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL) 2021.
10. **Tariq Alhindi**, Amal Alabdulkarim, Ali Alshehri, Muhammad Abdul-Mageed and Preslav Nakov. [AraStance: A Multi-Country and Multi-Domain Dataset of Arabic Stance Detection for Fact Checking](#). Proceedings of The 4th Workshop on NLP4IF: Censorship, Disinformation, and Propaganda, co-located with NAACL 2021.
9. **Tariq Alhindi** and Debanjan Ghosh. [“Sharks are not the threat humans are”: Argument Component Segmentation in School Student Essays](#). Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), with EACL 2021.
8. **Tariq Alhindi**, Smaranda Muresan and Daniel Preoțiu-Pietro. [Fact vs. Opinion: the Role of Argumentation Features in News Classification](#). Proceedings of the 28th International Conference on Computational Linguistics (COLING) 2020.
7. Christopher Hidey, Tuhin Chakrabarty, **Tariq Alhindi**, Siddharth Varia, Kriste Krstovski, Mona Diab and Smaranda Muresan. [DeSePtion: Dual Sequence Prediction and Adversarial Examples for Improved Fact-Checking](#). Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL) 2020.

6. **Tariq Alhindi**, Jonas Pfeiffer and Smaranda Muresan. [Fine-Tuned Neural Models for Propaganda Detection at the Sentence and Fragment Levels](#). Proceedings of The 2nd Workshop on NLP4IF: Censorship, Disinformation, and Propaganda, co-located with EMNLP 2019.
5. **Tariq Alhindi**, Savvas Petridis and Smaranda Muresan. [Where is your Evidence: Improving Fact-checking by Justification Modeling](#). Proceedings of the First Workshop on Fact Extraction and VERification (FEVER), co-located with EMNLP 2018.
4. El Moatez Billah Nagoudi, AbdelRahim Elmadany, Muhammad Abdul-Mageed and **Tariq Alhindi**. [Machine Generation and Detection of Arabic Manipulated and Fake News](#). Proceedings of the 5th Arabic Natural Language Processing Workshop, co-located with COLING 2020.
3. Amal Alabdulkarim and **Tariq Alhindi**. [Spider-Jerusalem at SemEval-2019 Task 04: Hyperpartisan News Detection](#). Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval), co-located with NAACL 2019.
2. Tuhin Chakrabarty, **Tariq Alhindi** and Smaranda Muresan. [Robust Document Retrieval and Individual Evidence Modeling for Fact Extraction and Verification](#). Proceedings of the First Workshop on Fact Extraction and VERification (FEVER), co-located with EMNLP 2018.
1. Elena Musi, **Tariq Alhindi**, Manfred Stede, Leonard Kriese, Smaranda Muresan and Andrea Rocci. [A Multi-layer Annotated Corpus of Argumentative Text: From Argument Schemes to Discourse Relations](#). Language Resources and Evaluation Conference (LREC). 2018.

RESEARCH EXPERIENCE

Graduate Research Assistant

Sep 2017 - present

Department of Computer Science, Columbia University. New York, NY, USA

Advisor: [Smaranda Muresan](#)

Fact-Checking Developed computational models for end-to-end fact-checking (document retrieval, evidence sentence selection and textual entailment) part of the Fact-Extraction and Verification (FEVER) shared tasks. Focused on computational models for selecting candidate sentences and textual entailment models for predicting the label of a claim (Supported, Refuted, Not Enough Info) given retrieved evidence from Wikipedia. Developed adversarial claims for fact-checking by adding temporal reasoning and lexical substitution using rule-based alterations, counter-fitted word embeddings, and WordNet. Developed models to capture more nuanced labels for claim veracity (true, mostly true, half true, false, mostly false and pants on fire). Modeled the human-provided justifications to predict the factuality of claims from Politifact.com that are tagged with these six degrees of truth. Released LIAR-PLUS, a fact-checking corpus with the automatically extracted justifications.

Claim Detection and Analysis of Argument and Discourse Structures Analyzed the correlation between Rhetorical Structure Theory (RST) discourse relations and argument schemes in a set of micro-text essays. Trained a BiLSTM model on six argumentative datasets for cross-domain claim detection using dropout, word-level attention and features from the ArgLex lexicon. Developed an RNN+BERT model for news vs. opinion articles classification using the argumentative types of sentences. Released a multi-layer annotated dataset of fact-checked climate change articles and argument structure to enable further study of the relation between argumentation and fact-checking. Currently developing a model for determining what sentences to fact-check in the climate change domain using argumentation and discourse features.

Detection of Propaganda Techniques and Argumentative Fallacies Developed a sentence-level propaganda detection model by fine-tuning a BERT classifier with modified probability thresholds for label prediction, and a fragment-level tagging model of eighteen propaganda techniques by training a BiLSTM-CRF tagger with stacked Flair, Glove, and Urban embeddings along with one-hot encoded features. Currently developing a taxonomy of fallacy for fact-checking in the climate change

domain by analyzing the relations of fact-checkers remarks and fact-checked segments to understand the type of argumentative fallacies in segments, which will provide grounds for automation in the future.

AI Research Intern

Jun 2021 - Sep 2021

AI News, Bloomberg LP. New York, NY, USA

Mentors: [Hao Zhang](#), [Debanjan Mahata](#)

Worked in the AI News group on developing adversarial attacks on sentiment models for financial text and improving these models through data augmentation and adversarial training. Implemented retrofitting for BERT by adding orthogonal transformation on the word embeddings to have more efficient and more successful attacks, and to have more robust sentiment models.

Research Intern

Jun 2020 - Aug 2020

NLP and Speech Team, Educational Testing Service. Princeton, NJ, USA

Mentor: [Debanjan Ghosh](#)

Worked on Claim and Premise Token Detection from middle and high school students essays. Developed CRF model using lexico-syntactic features and BERT embeddings. Fine-tuned BERT-based classifier with adaptive pre-training on GRE essays. Experimented with joint learning of sentence-level and token-level models in addition joint learning of claim and premise token detection.

Research Intern

Jun 2018 - Aug 2018

ProfX Search Team, Jet.com (Walmart Labs). Hoboken, NJ, USA

Mentor: [Aliasgar Kutiyawala](#)

Worked on Query and Stock Keeping Unit (SKU) understanding for product search in eCommerce. Extracted product entities automatically along with dictionaries creation and entity tagging for SKU understanding. Clustered SKUs based on semantic similarity of titles to expedite quality assurance process. Prepared a Summarization framework setup for automatic generation of SKU titles with products name and descriptive features.

Research Associate

Sep 2014 - Aug 2017

Center for Complex Engineering Systems, KACST. Riyadh, KSA

(Research Affiliate) Center for Complex Engineering Systems, MIT. Cambridge, MA, USA

Advisor: [Anas Alfaris](#)

Analyzed Arabic tweets for event detection and social patterns using LDA topic modeling. Developed City Vibes, neighborhoods classifier based on Geo-tagged tweets, applied on Riyadh, Saudi Arabia. Worked on Author disambiguation of Web of Science publication data. Designed and developed front-end web tools, visualizations and tangible user interfaces for urban systems.

OTHER WORK EXPERIENCE

Software Developer and Production Support

Feb 2014 - Aug 2014

Information Technology Department, Riyadh Bank. Riyadh, Saudi Arabia

Manager: [Tariq Al-Issa](#)

KIOSK production support team lead, and web services development.

Software Test Engineer

Jun 2013 - Dec 2013

Blue Highway Inc. Syracuse, NY, USA

Manager: [Albert Goldfain](#)

Software testing of Welch Allyn plug-in code before release. Working on ontology development and web development for multiple projects.

SKILLS

Languages	Arabic (Native), English (Proficient)
Programming	Python (sklearn, nltk, spaCy, keras, pytorch, gensim, transformers, textattack) R, JavaScript, HTML, CSS, Java, C
Tools	Unix, Git, Jupyter, Rstudio, MySQL, L ^A T _E X, Microsoft Office
Teaching	Teaching Assistant - Fall 2018 - COMS 4111 - Introduction to Databases Teaching Assistant - Spring 2019 - COMS 4705 - Natural Language Processing
Reviewing	SemEval 2019, JASIST 2019, NLP4IF 2019-2021 ACL 2020, NAACL 2021, EMNLP 2021, ACL 2022 (ARR)

GRANTS AND AWARDS

ACL 2020 Outstanding Reviewer Awarded to the top 10% of reviewers	<i>2020</i>
Amazon Student Bursary FEVER workshop at EMNLP 2018	<i>2018</i>
KACST Scholarship Ph.D. Scholarship for Academic Merit	<i>2017-2022</i>
King Abdullah Scholarship Program M.S. Scholarship for Academic Merit	<i>2011-2013</i>
KFUPM Honors Degree Graduating with Honors Degree	<i>2010</i>