

Hassan Sajjad | Scientist

<i>Current Role</i>	Scientist - manage applied machine learning group
<i>Address:</i>	Research Complex, Qatar Computing Research Institute, Hamad bin Khalifa University, 5825 Doha, Qatar
<i>Phone:</i>	+9744541548 (work), +97466863431 (cell)
<i>Email:</i>	hsajjad@qf.org.qa, hasaan.sajjad@gmail.com
<i>Webpage:</i>	https://hsajjad.github.io/
<i>Google Scholar:</i>	https://goo.gl/H5XPzb
<i>h-index</i>	22
<i>Languages</i>	Urdu (native), English (fluent), German (intermediate), Arabic (basic)

Profile

- PhD. in Computer Science, major in Natural Language Processing
- Experienced in
 - Applied machine learning and deep learning
 - Working with unstructured data both formal and informal genres
 - Handling languages with various morphological granularity and syntactic structure e.g. Russian, Arabic, German, English
- Taught courses on deep learning from the perspective of NLP and machine translation
- Technology transfer and startups with focus on applying research to real-world problems
- Experienced in working in a team and leading teams successfully
- Published work in top tier venues, such as, ACL, ICLR, AAAI, EMNLP, CSL, and CL

Interests – Research and Development, Entrepreneurial, Coaching

- Research and Development Interests
 - Natural language processing, Interpretation and manipulation of neural models, unsupervised learning, multi-task learning, transfer learning, feature selection, domain adaptation, model distillation
 - Statistical and neural machine translation, transliteration, language modeling, contextualized embeddings, dialect identification, user-generated content processing, social media analysis
- Entrepreneurial Interests: Lean startup, technology transfer, business development
- Teaching Interests: Deep learning for NLP, neural networks from scratch, basic computer science courses

Education

- **University of Stuttgart** **Germany**
PhD in Computer Science (Magna Cum Laude) **2008–2012**
Thesis: Statistical models for unsupervised, semi-supervised and supervised transliteration mining
Advisor: Prof. Dr. Hinrich Schütze
Committee: Prof. Dr. Alex Waibel, PD Dr. Helmut Schmid, Dr. Alexander Fraser, Prof. Dr. Albrecht Schmidt, Prof. Dr.-Ing Stefan Funke
- **National University of Computer and Emerging Sciences** **Pakistan**
Masters in Computer Science (CGPA 3.46) **2005–2007**
Thesis: Part-of-speech tagging for Urdu
- **National University of Computer and Emerging Sciences** **Pakistan**
Bachelors in Computer Science (CGPA 3.45) **2001–2005**
Project: Speaker recognition system

Professional Experience

- **Qatar Computing Research Institute** **Qatar**
Research Scientist **Jan. 2014–**
 - Managing applied machine learning group, commercialization of machine translation technology ([link](#)), and international collaborations – MIT and H2020 SUMMA project
 - Worked and published research work involving interpretation and manipulation of neural network models, machine translation, domain adaptation, natural language processing of informal text
 - Technology transfer to KanariAI, won several machine translation competitions such as 10th and 13th International Workshop on Spoken Language Translation for English-Arabic language pair
- **Qatar Computing Research Institute** **Qatar**
Post-doctoral Researcher **Feb. 2013–Jan. 2014**
 - Led research work on dialectal Arabic translation and shown significant improvement in quality of the Egyptian Arabic to English machine translation system.
- **DealBuzz** **Pakistan**
Co-founder **Jun. 2012–Jan. 2013**
 - Designed and built a product for innovative advertising on smart phones
 - Worked on the business process, identified potential customer segments, found niche, worked on the business model and launched the product
- **University of Stuttgart** **Germany**
Research Assistant and PhD Scholar **Feb. 2008–Oct. 2012**
 - Proposed a novel model for unsupervised transliteration mining that outperformed semi-supervised and supervised methods
 - Conducted PhD
 - Organized the machine translation reading group

- PolyVista** **United States**
Text Analysis Consultant *Feb. 2012–Jul. 2012*
 - Consulted in hiring a team for their new office in Lahore, Pakistan
 - Streamlined product development to integrate natural language processing applications like part-of-speech tagging, term extraction into a social media analysis platform
- Microsoft Research** **United States**
Research Intern *Jul. 2011–Oct. 2011*
 - Devised an algorithm for query refinement and efficient vertical search that guides users to their target result by engaging them into a dialog
 - Granted a patent and published the research work
- Center for Research in Urdu Language Processing, NUCES** **Pakistan**
Research Assistant *Feb. 2006–Feb. 2007*
 - Worked on Pan Asian Networking Localization project
 - Supervised and conducted on-field survey to measure technology dissemination in schools

Publications

Under-review Conference and Journal Papers

Firoj Alam, **Hassan Sajjad**, Muhammad Imran, and Ferda Ofli. "Standardizing and Benchmarking Crisis-related Social Media Datasets for Humanitarian Information Processing". In: *Underreview*. ([link](#)). 2020.

Firoj Alam et al. "Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society". In: *Underreview*. ([link](#)). 2020.

Prakhar Ganesh et al. "Compressing Large-Scale Transformer-Based Models: A Case Study on BERT". In: *Underreview*. ([link](#)). 2020.

Hassan Sajjad, Fahim Dalvi, Nadir Durrani, and Preslav Nakov. "Poor Man's BERT: Smaller and Faster Transformer Models". In: *Underreview*. ([link](#)). 2020.

Peer-reviewed Journal Papers

Yonatan Belinkov, Nadir Durrani, Fahim Dalvi, **Hassan Sajjad**, and James Glass. "On the Linguistic Representational Power of Neural Machine Translation Models". In: *Computational Linguistics (CL)* (2020). ([link](#)).

Abdul Rafae, Asim Karim, **Hassan Sajjad**, Faisal Kamiran, and Jia Xu. "A Clustering Framework for Lexical Normalization of Roman Urdu". In: *Natural Language Engineering (NLE)* (2020).

Shafiq Joty, Nadir Durrani, **Hassan Sajjad**, and Ahmed Abdelali. "Domain Adaptation using Neural Network Joint Model". In: *Computer Speech and Language* 45.C (2017). ([link](#)). ISSN: 0885-2308.

Hassan Sajjad, Helmut Schmid, Alexander Fraser, and Hinrich Schütze. "Statistical models for unsupervised, semi-supervised and supervised transliteration mining". In: *Computational Linguistics* 43.2 (2017). ([link](#)).

Walid Magdy, **Hassan Sajjad**, Tarek El-Ganainy, and Fabrizio Sebastiani. "Bridging social media via distant supervision". In: *Social Network Analysis and Mining* 35.5 (2015). ([link](#)).

Peer-reviewed Conference Papers

Fahim Dalvi, **Hassan Sajjad**, Nadir Durrani, and Yonatan Belinkov. "Analyzing Redundancy in Pretrained Transformer Models". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ([link](#)). Online, Nov. 2020.

Nadir Durrani, **Hassan Sajjad**, Fahim Dalvi, and Yonatan Belinkov. "Analyzing Individual Neurons in Pre-trained Language Models". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online, Nov. 2020.

John Wu, Yonatan Belinkov, **Hassan Sajjad**, Nadir Durrani, Fahim Dalvi, and James Glass. "Similarity Analysis of Contextual Word Representation Models". In: *Proceedings of the Annual Conference of the Association for Computational Linguistics (ACL)*. [\(link\)](#). US, June 2020.

Nadir Durrani, Fahim Dalvi, **Hassan Sajjad**, Yonatan Belinkov, and Preslav Nakov. "One Size Does Not Fit All: Comparing NMT Representations of Different Granularities". In: *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. [\(link\)](#). Minneapolis, US, June 2019.

Hamdy Mubarak, Ahmed Abdelali, **Hassan Sajjad**, Younes Samih, and Kareem Darwish. "Highly Effective Arabic Diacritization using Sequence to Sequence Modeling". In: *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. [\(link\)](#). Minneapolis, US, June 2019.

D. Anthony Bau*, Yonatan Belinkov*, **Hassan Sajjad**, Fahim Dalvi, Nadir Durrani, and James Glass. "Identifying and Controlling Important Neurons in Neural Machine Translation". In: *International Conference on Learning Representations (ICLR)*. [\(link\)](#). New Orleans, US, May 2019.

Fahim Dalvi*, Nadir Durrani*, **Hassan Sajjad***, Yonatan Belinkov, D. Anthony Bau, and James Glass. "What is one Grain of Sand in the Desert? Analyzing Individual Neurons in Deep NLP Models". In: *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*. [\(link\)](#). Honolulu, US, Mar. 2019.

Fahim Dalvi, Nadir Durrani, **Hassan Sajjad**, and Stephan Vogel. "Incremental Decoding and Training Methods for Simultaneous Translation in Neural Machine Translation". In: *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*. [\(link\)](#). New Orleans, US, June 2018.

Yonatan Belinkov, Nadir Durrani, Fahim Dalvi, **Hassan Sajjad**, and James Glass. "What do Neural Machine Translation Models Learn about Morphology?". In: *Proceedings of the 55th Conference of the Association for Computational Linguistics (ACL)*. [\(link\)](#). Vancouver, Canada, Aug. 2017.

Hassan Sajjad, Fahim Dalvi, Nadir Durrani, Ahmed Abdelali, Yonatan Belinkov, and Stephan Vogel. "Challenging Language-Dependent Segmentation for Arabic: An Application to Machine Translation and Part-of-Speech Tagging". In: *Proceedings of the 55th Conference of the Association for Computational Linguistics (ACL)*. [\(link\)](#). Vancouver, Canada, Aug. 2017.

Dat Tien Nguyen, Kamla Al-Mannai, Shafiq Joty, **Hassan Sajjad**, Muhammad Imran, and Prasenjit Mitra. "Robust Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks". In: *Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM)*. [\(link\)](#). Montreal, Canada, May 2017.

Yonatan Belinkov, Lluís Màrquez, **Hassan Sajjad**, Nadir Durrani, Fahim Dalvi, and James Glass. "Evaluating Layers of Representation in Neural Machine Translation on Part-of-Speech and Semantic Tagging Tasks". In: *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP)*. [\(link\)](#). Taipei, Taiwan, Nov. 2017.

Fahim Dalvi, Nadir Durrani, **Hassan Sajjad**, Yonatan Belinkov, and Stephan Vogel. "Understanding and Improving Morphological Learning in the Neural Machine Translation Decoder". In: *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP)*. [\(link\)](#). Taipei, Taiwan, Nov. 2017.

Hassan Sajjad et al. "Eyes Don't Lie: Predicting Machine Translation Quality Using Eye Movement." In: *Proceedings of the 15th Annual Conference of the North American Chapter of the Association of Computational Linguistics: Human Language Technologies (NAACL-HLT)*. [\(link\)](#). San Diego, US, June 2016.

Nadir Durrani, **Hassan Sajjad**, Shafiq Joty, and Ahmed Abdelali. "A Deep Fusion Model for Domain Adaptation in Phrase-based MT". In: *Proceedings of the 26th International Conference on Computational Linguistics (COLING)*. [\(link\)](#). Osaka, Japan, Dec. 2016.

Hassan Sajjad, Francisco Guzmán, and Stephan Vogel. "An Empirical Study: Post-editing Effort for English to Arabic Hybrid Machine Translation". In: *Proceedings of the Association for Machine Translation in the Americas (AMTA)*. ([link](#)). Austin, US, Oct. 2016.

Shafiq Joty, **Hassan Sajjad**, Nadir Durrani, Kamla Al-Mannai, Ahmed Abdelali, and Stephan Vogel. "How to Avoid Unwanted Pregnancies: Domain Adaptation using Neural Network Models". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ([link](#)). Lisbon, Portugal, Sept. 2015.

Abdul Rafae, Abdul Qayyum, Muhammad Moeen Uddin, Asim Karim, **Hassan Sajjad**, and Faisal Kamiran. "An Unsupervised Method for Discovering Lexical Variations in Roman Urdu Informal Text." In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ([link](#)). Lisbon, Portugal, Sept. 2015.

Walid Magdy, **Hassan Sajjad**, Tarek El-Ganainy, and Fabrizio Sebastiani. "Distant Supervision for Tweet Classification Using YouTube Labels". In: *Proceedings of the Ninth International AAAI Conference on Web and Social Media (ICWSM)*. ([link](#)). Oxford, UK, May 2015.

Nadir Durrani, **Hassan Sajjad**, Shafiq Joty, Ahmed Abdelali, and Stephan Vogel. "Using Joint Models for Domain Adaptation in Statistical Machine Translation". In: *Proceedings of the 15th Machine Translation Summit (MT Summit XV)*. ([link](#)). Florida, USA, Nov. 2015.

Ahmed Abdelali, Francisco Guzman, **Hassan Sajjad**, and Stephan Vogel. "The AMARA Corpus: Building Parallel Language Resources for the Educational Domain". In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*. ([link](#)). Reykjavik, Iceland, May 2014.

Mohammad Moeen Uddin, Mohammad Imran, and **Hassan Sajjad**. "Understanding Types of Users on Twitter". In: *Proceedings of the 6th ASE International Conference in Social Computing (SocialCom)*. ([link](#)). Stanford, USA, May 2014.

Nadir Durrani, **Hassan Sajjad**, Hieu Hoang, and Philipp Koehn. "Integrating an Unsupervised Transliteration Model into Statistical Machine Translation". In: *Proceedings of the 15th Conference of the European Chapter of the ACL (EACL)*. ([link](#)). Gothenburg, Sweden, Apr. 2014.

Kareem Darwish, **Hassan Sajjad**, and Hamdy Mubarak. "Verifiably Effective Arabic Dialect Identification." In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ([link](#)). Doha, Qatar, Oct. 2014.

Hassan Sajjad, Kareem Darwish, and Yonatan Belinkov. "Translating Dialectal Arabic to English". In: *Proceedings of the 51st Conference of the Association for Computational Linguistics (ACL)*. ([link](#)). Sofia, Bulgaria, Aug. 2013.

Hassan Sajjad, Alexander Fraser, and Helmut Schmid. "A Statistical Model for Unsupervised and Semi-supervised Transliteration Mining". In: *Proceedings of the 50th Conference of the Association for Computational Linguistics (ACL)*. ([link](#)). Jeju, Korea, July 2012.

Hassan Sajjad, Patrick Pantel, and Michael Gamon. "Underspecified Query Refinement via Natural Language Question Generation". In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*. ([link](#)). Mumbai, India, Dec. 2012.

Hassan Sajjad, Alexander Fraser, and Helmut Schmid. "An Algorithm for Unsupervised Transliteration Mining with an Application to Word Alignment". In: *Proceedings of the 49th Conference of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*. ([link](#)). Portland, OR, USA, June 2011.

Hassan Sajjad, Nadir Durrani, Helmut Schmid, and Alexander Fraser. "Comparing Two Techniques for Learning Transliteration Models Using a Parallel Corpus". In: *Proceedings of 5th International Joint Conference on Natural Language Processing (IJCNLP)*. ([link](#)). Chiang Mai, Thailand, Nov. 2011.

Nadir Durrani, **Hassan Sajjad**, Alexander Fraser, and Helmut Schmid. "Hindi-to-Urdu Machine Translation through Transliteration". In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*. ([link](#)). Uppsala, Sweden, July 2010.

Hassan Sajjad and Helmut Schmid. "Tagging Urdu Text with Parts of Speech: A Tagger Comparison". In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. ([link](#)). Athens, Greece, Apr. 2009.

Peer-reviewed Demo Conference Papers

Hamdy Mubarak, Ahmed Abdelali, Kareem Darwish, Mohamed Eldesouki, Younes Samih, and **Hassan Sajjad**. "A System for Diacritizing Four Varieties of Arabic". In: *In Proceedings of the Empirical Methods in Natural Language Processing (EMNLP)*. ([link](#)). Nov. 2019.

Fahim Dalvi et al. "NeuroX: A Toolkit for Analyzing Individual Neurons in Neural Networks". In: *AAAI Conference on Artificial Intelligence (AAAI)*. ([link](#)). Honolulu, USA, Jan. 2019.

Fahim Dalvi et al. "QCRI Live Speech Translation System". In: *In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. ([link](#)). Apr. 2017.

Renars Liepins et al. "The SUMMA Platform Prototype". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. ([link](#)). Apr. 2017.

Peer-reviewed Workshop Papers

Hassan Sajjad, Nadir Durrani, Fahim Dalvi, Yonatan Belinkov, and Stephan Vogel. "Neural Machine Translation Training in a Multi-Domain Scenario". In: *Proceedings of the 14th International Workshop on Spoken Language Translation (IWSLT)*. ([link](#)). Tokyo, Japan, Dec. 2017.

Dat Tien Nguyen, Shafiq Joty, Muhammad Imran, **Hassan Sajjad**, and Prasenjit Mitra. "Applications of Online Deep Learning for Crisis Response Using Social Media Information". In: *Proceedings of the 4th International Workshop on Social Web for Disaster Management (SWDM)*. ([link](#)). Indianapolis, US, Oct. 2016.

Wajdi Zaghouani, Ahmed Abdelali, Francisco Guzmán, and **Hassan Sajjad**. "Normalizing Mathematical Expressions to Improve the Translation of Educational Content". In: *Proceedings of the AMTA 2016 Workshop Semitic Machine Translation (SeMaT)*. ([link](#)). Austin, US, Oct. 2016.

Francisco Guzmán, Ahmed Abdelali, Irina Temnikova, **Hassan Sajjad**, and Stephan Vogel. "How do Humans Evaluate Machine Translation". In: *Proceedings of the Workshop on Machine Translation (WMT)*. ([link](#)). Lisbon, Portugal, Sept. 2015.

Kamla Al-Mannai, **Hassan Sajjad**, Alaa Khader, Fahad Al Obaidli, Preslav Nakov, and Stephan Vogel. "Unsupervised word segmentation improves dialectal Arabic to English machine translation". In: *Proceedings of the Workshop of Arabic Natural Language Processing (ANLP)*. ([link](#)). Doha, Qatar, Oct. 2014.

Francisco Guzmán, **Hassan Sajjad**, Stephan Vogel, and Ahmed Abdelali. "The AMARA Corpus: Building Resources for Translating the Web's Educational Content". In: *Proceedings of the 10th International Workshop on Spoken Language Technology (IWSLT)*. ([link](#)). Heidelberg, Germany, Dec. 2013.

Shared-task Workshop Papers

Xian Li et al. "Findings of the First Shared Task on Machine Translation Robustness". In: *Proceedings of the Fourth Conference on Machine Translation, Volume 2: Shared Task Papers*. ([link](#)). Florence, Italy, Aug. 2019.

Houda Bouamor and **Hassan Sajjad**. "H2@BUCC18: Parallel Sentence Extraction from Comparable Corpora Using Multilingual Sentence Embeddings". In: *Proceedings of the 11th Workshop on Building and Using Comparable Corpora (BUCC)*. ([link](#)). Miyazaki, Japan, May 2018.

Nadir Durrani, Fahim Dalvi, **Hassan Sajjad**, and Stephan Vogel. "QCRI's Machine Translation Systems for IWSLT'2016". In: *Proceedings of the 13th International Workshop on Spoken Language Translation (IWSLT)*. ([link](#)). Seattle, USA, Dec. 2016.

Mohamed Eldesouki, Fahim Dalvi, **Hassan Sajjad**, and Kareem Darwish. "QCRI @ DSL 2016: Spoken Arabic Dialect Identification Using Textual". In: *Proceedings of the 3rd Workshop on NLP for Similar Languages, Varieties and Dialects*. ([link](#)). Osaka, Japan, Dec. 2016.

Houda Bouamor, **Hassan Sajjad**, Nadir Durrani, and Kemal Oflazer. "QCMUQ@QALB-2015 Shared Task: Combining Character level MT and Error-tolerant Finite-State Recognition for Arabic Spelling Correction". In: *Proceedings of the Workshop of Arabic Natural Language Processing (ANLP)*. [\(link\)](#). Beijing, China, July 2015.

Hassan Sajjad et al. "QCN Egyptian Arabic to English Machine Translation System for NIST OpenMT15". In: *Workshop of NIST OpenMT15*. [\(link\)](#). Washington DC, US, June 2015.

Nadir Durrani, Helmut Schmid, Alexander Fraser, **Hassan Sajjad**, and Richárd Farkas. "Munich-Edinburgh-Stuttgart Submissions of OSM Systems at WMT13". In: *Proceedings of the Eighth Workshop on Statistical Machine Translation (WMT)*. [\(link\)](#). Sofia, Bulgaria, Aug. 2013.

Hassan Sajjad, Svetlana Smekalova, Nadir Durrani, Alexander Fraser, and Helmut Schmid. "QCRI-MES Submission at WMT13: Using Transliteration Mining to Improve Statistical Machine Translation". In: *Proceedings of the Eighth Workshop on Statistical Machine Translation (WMT)*. [\(link\)](#). Sofia, Bulgaria, Aug. 2013.

Marion Weller et al. "Munich-Edinburgh-Stuttgart Submissions at WMT13: Morphological and Syntactic Processing for SMT". In: *Proceedings of the Eighth Workshop on Statistical Machine Translation (WMT)*. [\(link\)](#). Sofia, Bulgaria, Aug. 2013.

Hassan Sajjad et al. "QCRI at IWSLT 2013: Experiments in Arabic-English and English-Arabic Spoken Language Translation". In: *Proceedings of the 10th International Workshop on Spoken Language Technology (IWSLT)*. [\(link\)](#). Heidelberg, Germany, Dec. 2013.

Teaching Experience

- **NLP with Python** **Qatar**
• [Carnegie Mellon University in Qatar](#) 20th Oct. - 15th Nov. 2020
As part of the NLP with Python course, delivered 6 hours lectures covering the fundamentals of deep learning with practical exercises
- **Deep Learning for NLP** **Germany**
• [University of Duisburg-Essen](#) 9-13th Sept. 2019
15 hours course covering the fundamentals of deep learning from intuition to practical exercises
- **Deep Learning for NLP** **Sri Lanka**
• [International Spring School, University of Moratuwa](#) 16-21st Mar. 2019
15 hours course on the basics of deep learning with hands-on exercises involving several NLP tasks
- **From Theory to Practice: Deep Learning for NLP** **Germany**
• [University of Duisburg-Essen](#) 8-14th Apr. 2018
15 hours course covering theory of deep learning models from intuition to practical exercises in Keras ([content](#))
- **Deep Learning for Machine Translation** **Germany**
• [9th Computational linguistics fall school, DGfS](#) 11-22nd Sept. 2017
15 hours course teaching fundamentals of deep learning and its application to machine translation ([content](#))
- **Probability and Statistics** **Pakistan**
• [National University of Computer and Emerging Sciences](#) Sept. 2004–Dec. 2004
Teaching assistant
- **Calculus** **Pakistan**
• [National University of Computer and Emerging Sciences](#) Feb. 2003–Jun. 2003

Teaching assistant

- **Introduction to Natural Language Processing**

- [ADDO AI](#)

- Course content design and delivery advisor

Pakistan

Nov. 2019–to date

Skills

- Programming languages: Python, Perl, Shell, Keras, PyTorch
- Tools and libraries: Transformers, Nematus, OpenNMT, Moses, Scikit-learn, NLTK

Patent

- [Search system with query refinement](#) (US Patent Number 9767144B2)
- Method and system for diacritizing Arabic text (pending)

Grants

- Qatar Science and Technology Park Accelerator Program 2016 – Grant of 100K USD accelerator funding for Customized Machine Translation

Research Competitions

- Workshop on Building and using Comparable Corpora (2018), French-English parallel corpus extraction
- **Best system** – International Workshop on Spoken Language Technology (2016), Arabic-English translation task
- **Second best system and best system description paper** – NIST Open Machine Translation Evaluation Workshop (2015), Egyptian Arabic to English translation task
- **Second best system** – Shared Task on Automatic Arabic Error Correction (2015)
- **Best system** – International Workshop on Spoken Language Technology (2013), Arabic-English translation task
- **Second best constrained system** – Eighth Workshop on Statistical Machine Translation (2013), English-Russian translation task

Open Source Contributions

- NeuroX - Interpretation and manipulation of deep models ([link](#))
- Transliteration Mining Tool ([link](#))
- Linguistic Resources
 - The QCRI Educational Domain corpus – a collection of parallel corpora of the education domain in multiple languages ([link](#))
 - English-Hindi and English-Arabic transliteration gold set for transliteration mining evaluation ([link](#))

- Most widely-used part of speech tagset of Urdu and annotated corpus of 100,000 words

Honors and Awards

- Machine translation technology transfer to [KanariAI](#) (June 2020)
- Media coverage: 1 billion tokens translated by our [machine translation system](#) (Aug. 2020). ([link1](#)), ([link2](#))
- Panelist on discussing NLP Research In Pakistan: Building Synergies and Collaborations at 7th International Conference on Language and Technology (Feb. 2020). ([link](#))
- Media coverage: MIT News covers our work on analyzing and controlling deep models (Feb. 2019) ([link](#))
- Best Innovation Award at the Annual Research Conference 2018, Qatar for our speech translation system ([link](#))
- Media coverage: MIT News covers our work on analyzing representations in deep models (Dec. 2017) ([link](#))
- Qatar Science and Technology Park speed pitching (Nov. 2015) – selected among the top 3 teams out of 26 teams
- Second position in the Qatar Annual Research Conference 2014 for a student research project
- Google sponsorship for participation in the Second Lisbon Machine Learning School
- Full PhD merit-based scholarship from HEC-DAAD for doctoral studies in Germany (4 years)
- Four year scholarship for the bachelors program

Invited Talks and Keynotes

- Hidden Linguistics in Deep NLP Models. Heinrich-Heine Universität Düsseldorf, Germany (Oct. 2020)
- Interpreting Deep NLP Models. University of Edinburgh, UK (Apr. 2020)
- Interpreting Deep NLP Models. University of Sheffield, UK (Mar. 2020)
- Efficient Transfer Learning of Pre-trained Model. 7th International Conference on Language and Technology ([link](#)), UET, Pakistan (Feb. 2020) (**Keynote**)
- Interpreting Deep NLP Models: A Case Study on Neural Machine Translation at Google, Salesforce, Bosch, Facebook and Amazon, US (Apr. 2019)
- Hidden Linguistics in Deep NLP Models. Symposium on Natural Language Processing. University of Moratuwa, Sri Lanka (Mar. 2019) (**Keynote**)
- Machine Translation in the Real World. King's College London, UK (Mar. 2019)
- Analyzing Individual Neurons in Deep NLP Models. University of Melbourne, Melbourne, Australia (Feb. 2019)
- Analyzing Individual Neurons in Deep NLP Models. Thomson Reuters, Toronto, Canada (Feb. 2019)
- What do Neural Machine Translation Models Learn about Morphology. Macquarie University, Sydney, Australia (Apr. 2017)
- What do Neural Machine Translation Models Learn about Morphology. University of Melbourne, Melbourne, Australia (Apr. 2017)

- From Phrase-based to Neural Machine Translation. Workshop on Semitic Machine Translation, AMTA, Austin, US (Nov. 2016) (**Keynote**)
- Deep Learning – Neural Machine Translation. Sixth Conference on Language and Technology (CLT16), Lahore, Pakistan (Nov. 2016) (**Keynote**)
- Content Model Applications for Promoting Local Language Content. Workshop on Facilitating Local Language Content Access and Generation using Human Language Technologies, UET, Lahore, Pakistan (Aug. 2015) (**Keynote**)
- Statistical Machine Translation for Community Service: Translating Educational Content. Fifth Conference on Language and Technology (CLT14), Karachi, Pakistan (Nov. 2014) (**Keynote**)
- Separating Transliterations from Translations in Transliteration Mining Context. FBK, Trento, Italy (Oct. 2012)
- Unsupervised Transliteration Mining. School of Science and Engineering, Lahore University of Management and Sciences, Pakistan (Apr. 2012)
- Unsupervised Transliteration Mining, Punjab University College of Information Technology, Pakistan (Apr. 2012)

Professional Services

- Senior PC: AAAI 2021
- Workshop/shared task organization: MT robustness 2019/20 (WMT), BlackBoxNLP 2020 (EMNLP)
- Conference Area chair: ACL 2020 (NLP applications)
- Conference reviewer: regular reviewer at ACL/EMNLP/NAACL/EACL/COLING
- Conference reviewer: ICLR 2020, NIPS 2020, AAAI 2019/20, IJCAI 2018/19, EAMT 2017
- Workshop reviewer: IWSLT, WANLP
- Journal reviewer: TACL, ML, NLE, LRE, PAMI, PLOS-ONE
- Journal review editor: [Language and computation](#)
- **Challenge22**: proposal reviewer for phase 1 and phase 2 (2017)
- Judge at the annual Hackathon at Carnegie Mellon University - Qatar (27-28th Jan. 2017)
- Judge at the first Alice Competition at Carnegie Mellon University - Qatar (23rd May 2016)

Advising

External PhD Committee

- External Ph.D. Advisor: Samreen Kazi, IBA, Pakistan. Topic: SQUAD for Urdu Language
- External Ph.D. Advisor: Reem Suwaileh, Qatar University, Qatar. Topic: Automatic Geo-location tagging
- External Ph.D. Committee Member: Jamal Abdul Nasir. Lahore University of Management Sciences, Pakistan. 2014

Masters and Bachelor Advising

- Laura Lehmann, LMU Munich, Germany (External examiner of Bachelor thesis titled MulTraLit: a multilingual transliteration system)
- Kamela Ali Al Mannai, HBKU, Qatar (External advisor, Master thesis on Dialect Identification)

- Abdul Rafae, LUMS, Pakistan (External advisor, later did PhD from the Hunter College City University, New York)
- Fahim Dalvi (Research intern, graduated from Stanford University, now working at QCRI)
- Dat Tien Nguyen (Research intern, now a PhD student at the University of Amsterdam)
- Kamela Ali Al Mannai (Research intern, Qatar University)
- Alaa Khader (Research intern, Carnegie Mellon University Qatar)

Participations

- Deep Learning for Machine Translation, Winter School, Dublin, Ireland (18-24th Oct. 2015)
- Dagstuhl Seminar on Statistical Techniques for Translating to Morphologically Rich Languages (2-7th Feb. 2014)
- Taming the social web, 2nd Lisbon Machine Learning School (19-25th Jul. 2012)
- 5th MT Marathon, Le Mans France (13-17 Sept. 2010). Worked with Adam Lopez on Minimum Risk Decoding in Cdec
- Summer School in Asian Language Processing (Jun.-Aug. 2006)

International Collaborations

- **Massachusetts Institute of Technology (Oct. 2016 – Mar. 2020)**
Interpreting deep neural models
 - Role: managed and led the machine translation research agenda from the QCRI side
 - Worked on the analysis and interpretation of neural machine translation in its ability to learn various language phenomena
- **Scalable Understanding of Multilingual Media (Feb. 2016 – Feb. 2019)**
EU Horizon 2020 funded Big Data project aimed at creating a platform to automate the analysis of media streams across several languages
 - Role: managed Arabic machine translation component. Responsible for research and development deliverables
 - Delivered an Arabic to English neural machine translation system
 - Proposed and delivered a language-independent pipeline for Arabic processing
- **Carnegie Mellon University – Qatar (Apr. 2015 – Dec. 2017)**
Natural Language Processing of Dialectal Arabic
 - Role: a team player and a manager
 - Conducted research on Arabic error correction, dialect identification and spelling standardization
 - Ranked second on an Arabic error correction task
- **Karlsruhe Institute of Technology (Mar. 2016 – Jun. 2016)**
German to Arabic Lecture Translation
Project was motivated from the refugee crisis in Europe. It aimed to provide real-time German to Arabic translation to Arabic students in a university hall
 - Role: a team player, responsible for German-Arabic data extraction, and translation system
 - Faced the challenge of little to no parallel data available for German to Arabic
 - Adapted data generation and modeling strategies to build a usable system

- System deployed in university halls

References

Prof. Dr. Hinrich Schütze
Chair of Computational Linguistics
Director, Center for Information and Language Processing
University of Munich, Oettingenstr 67
80538 München, Germany
Phone: +49 89 2180 9720-1
Email: hinrich@hotmail.com

Dr. Preslav Nakov
Principal Scientist
Qatar Computing Research Institute
HBKU, Doha, Qatar 5825
Phone: +974 4454 1733
Email: pnakov@hbku.edu.qa

Prof. Dr. Alexander M. Fraser
Center for Information and Language Processing
Ludwig-Maximilians-Universität München
University of Munich, Oettingenstr 67
80538 München, Germany
Phone: +49 89 2180 9710
Email: fraser@cis.uni-muenchen.de

Dr. Lluís Màrquez
Principal Applied Scientist
Amazon
Email: lluismv@amazon.com