

31 Broomlea Crescent, Edinburgh,  
Scotland, Great Britain  
EH12 7NR

## HARIS BIN ZIA

+44 07498300955  
[harisbinzia@gmail.com](mailto:harisbinzia@gmail.com)  
<https://github.com/harisbinzia>

### EDUCATION

<b>London, United Kingdom</b>	<b>Queen Mary University of London</b>	<b>Spring 2021 - Present</b>
<ul style="list-style-type: none"><li>• PhD (Computer Science)</li><li>• Doctorate Coursework: Deep Learning for NLP; Deep Learning for Audio and Music; Computational Creativity.</li><li>• Thesis: Protecting audiences from online harms on social media.</li></ul>		
<b>Lahore, Pakistan</b>	<b>Information Technology University</b>	<b>Fall 2016 - Spring 2018</b>
<ul style="list-style-type: none"><li>• Master of Science (Computer Science). <b>CGPA: 3.80</b></li><li>• Graduate Coursework: Computational Linguistics; Speech Processing.</li><li>• Thesis: <i>Guftaar Shanaas</i>: Multi-Speaker Urdu Speech Recognition.</li></ul>		
<b>Islamabad, Pakistan</b>	<b>National University of Computer and Emerging Sciences</b>	<b>Fall 2011 - Spring 2015</b>
<ul style="list-style-type: none"><li>• Bachelor of Science (Computer Science). <b>CGPA: 3.48</b></li><li>• Undergraduate Coursework: Artificial Intelligence; Information Retrieval; Design and Analysis of Algorithms.</li></ul>		
<b>Independent Coursework:</b>		
<ul style="list-style-type: none"><li>• <a href="#">Deep Learning, a 5-course specialization by deeplearning.ai on Coursera.</a></li><li>• <a href="#">Machine Learning by Stanford University on Coursera.</a></li><li>• <a href="#">Python for Everybody, a 5-course specialization by University of Michigan on Coursera.</a></li></ul>		

### EMPLOYMENT

<b>Intern</b>	<b>Telefonica Research</b>	<b>November 2022 - January 2023</b>
<ul style="list-style-type: none"><li>• Worked on Content Moderation in Decentralized Social Networks e.g., Mastodon</li></ul>		
<b>Research Associate</b>	<b>Information Technology University</b>	<b>August 2018 - December 2020</b>
<ul style="list-style-type: none"><li>• Research Associate at Center for Speech and Language Technologies (CSaLT), worked on text simplification, lexical normalization of Roman Urdu (Latin script Urdu) and Urdu TTS.</li></ul>		
<b>Graduate Fellow</b>	<b>Information Technology University</b>	<b>September 2016 - June 2018</b>
<ul style="list-style-type: none"><li>• Research Assistant at CSaLT, worked on Urdu ASR, Pronunciation Modeling and Text Segmentation.</li><li>• Teaching Assistant: Discrete Mathematics (Fall 2016), Linear Algebra (Spring 2017), Software Testing (Summer 2017), Natural Language Processing (Fall 2017) and Speech Processing (Spring 2018).</li></ul>		
<b>Solution Expert</b>	<b>Telenor Group</b>	<b>November 2015 - August 2016</b>
<ul style="list-style-type: none"><li>• Developed integration components using Java, Business Process Execution Language (BPEL), Oracle Service Bus (OSB) and Oracle Data Integrator (ODI) as per Application Integration Architecture (AIA).</li></ul>		
<b>Intern</b>	<b>Bentley Systems</b>	<b>June 2014 - August 2014</b>
<ul style="list-style-type: none"><li>• Wrote automation scripts to some of regression testing tasks and also contributed towards writing tests using GoogleTest for C++ APIs.</li></ul>		

### PUBLICATIONS

- Jiahui He\*, **Haris Bin Zia\***, Aravindh Raman, Ignacio Castro, Nishanth Sastry, Gareth Tyson, Flocking to Mastodon: Tracking the Great Twitter Migration. In Proceedings of the 23<sup>rd</sup> ACM Internet Measurement Conference (**IMC 2023**), Oct 24-26, Montreal, Canada.
- Ishaku Hassan Anaobi, Aravindh Raman, Ignacio Castro, **Haris Bin Zia**, Damilola Ibosiola, Gareth Tyson, Will Admins Cope? Decentralized Moderation in the Fediverse. In Proceedings of the 2023 ACM Web Conference (**WWW 2023**), April 30-May 04, Texas, USA.
- **Haris Bin Zia**, Aravindh Raman, Ignacio Castro, Ishaku Hassan Anaobi, Emiliano De Cristofaro, Nishanth Sastry, Gareth Tyson, Toxicity in the Decentralized Web and the Potential for Model Sharing. In Proceedings of the 2022 ACM SIGMETRICS/ IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (**SIGMETRICS/ PERFORMANCE 2022**), June 06-10, Mumbai, India.
- **Haris Bin Zia**, Ignacio Castro, Arkaitz Zubiaga, Gareth Tyson, Improving Zero-shot Cross-lingual Hate Speech Detection with pseudo-label fine-tuning of Transformer Language Models. In Proceedings of the 16<sup>th</sup> International AAAI Conference on Web and Social Media (**ICWSM 2022**), June 06-09, Georgia, USA.
- Ishaku Hassan Anaobi, Aravindh Raman, Ignacio Castro, **Haris Bin Zia**, Emiliano De Cristofaro, Nishanth Sastry,

---

Gareth Tyson, Exploring content moderation in the decentralised web: the pleroma case. In Proceedings of the 17<sup>th</sup> International Conference on emerging Networking EXperiments and Technologies (**CoNEXT 2021**), December 07-10, Munich, Germany.

- **Haris Bin Zia**, Ignacio Castro, Gareth Tyson, Racist or Sexist Meme? Classifying Memes beyond Hateful. In Proceedings of the 5<sup>th</sup> Workshop on Online Abuse and Harms (**WOAH 2021**), August 01-06, Online.
- Muhammad Qasim, **Haris Bin Zia**, Awais Athar, Tania Habib, Agha Ali Raza, Personalized weather information for low-literate farmers using multimodal dialog systems. In International Journal of Speech Technology (**IJST 2021**).
- Namooos Hayat Qasmi, **Haris Bin Zia**, Awais Athar, Agha Ali Raza, SimplifyUR: Unsupervised Lexical Text Simplification for Urdu. In 12<sup>th</sup> International Conference on Language Resources and Evaluation (**LREC 2020**), Marseille, France.
- Agha Ali Raza, Awais Athar, Shan Randhawa, Zain Tariq, Muhammad Bilal Saleem, **Haris Bin Zia**, Umar Saif, Roni Rosenfeld, Rapid Collection of Spontaneous Speech Corpora using Telephonic Community Forums. In 19<sup>th</sup> INTERSPEECH event (**INTERSPEECH 2018**), September 02-06, Hyderabad, India.
- **Haris Bin Zia**, Agha Ali Raza, Awais Athar, Urdu Word Segmentation using Conditional Random Fields. In 27<sup>th</sup> International Conference on Computational Linguistics (**COLING 2018**), August 20-26, New-Mexico, USA.
- **Haris Bin Zia**, Agha Ali Raza, Awais Athar, PronouncUR: An Urdu Pronunciation Lexicon Generator. In 11<sup>th</sup> International Conference on Language Resources and Evaluation (**LREC 2018**), May 07-12, Miyazaki, Japan.

---

#### HONORS AND AWARDS

- Awarded scholarship for doctoral degree at Queen Mary University of London.
- Awarded LDC Data Scholarship worth USD \$4000 for my master's thesis (Fall 2017).
- Awarded Graduate Fellowship for master's degree at Information Technology University.
- Awarded NICT scholarship for bachelor's degree at National University of Computer and Emerging Sciences.
- Inscribed in Dean's List of Honors for excellent academic achievement in the semesters: Fall 2012, Spring 2013, Spring 2014, Fall 2014, Spring 2015, Fall 2016, Spring 2017, Fall 2017 and Spring 2018.

---

#### LANGUAGES AND TECHNOLOGIES

- Python; C++; Keras; PyTorch; Hugging Face; spaCy, NLTK
- IPython Notebook; PyCharm; CMU Sphinx; KALDI