

SHANTANU NATH

RESEARCH FELLOW · COMPUTATIONAL LINGUISTICS

33, Viale verona, Trento, Italy

☎ (+39)3887206711 | ✉ shantanu.cse@std.cu.ac.bd | 📱 shantanu778 | 🌐 shantanu778 | 🎓 Shantanu

Objectives

My enthusiasm lies in Natural Language Processing, Machine Learning, Multimodal Learning, and intersection between AI and social science. My research interests focus on exploring the synergy between AI and human culture, leveraging advanced computational methods to address complex societal challenges. I also aspire to contribute into cutting-edge research that addresses critical challenges in Explainability, Trustworthiness and Reasoning in AI.

Research Experience

CIMEC (Center for Mind/Brain Science), University of Trento

Trento, Italy

RESEARCH FELLOW

Feb. 2024 -

- Working within the project “Repetition Experiments as Probes on Linguistics Analysis and Integration (REPLAI) funded by MUR and European Union within the programme PRIN 2022, Principal Investigator prof. Roberto Zamparelli

Fondazione Bruno Kessler (FBK) in Information and Communication Technology

Povo, Italy

NATURAL LANGUAGE PROCESSING INTERN

Oct 2022 - Dec 2022

- Investigated the correlation between confidence scores and evaluation metrics to assess the quality of automatic captions generated by an Automatic Speech Recognition model and their translations in seven languages by specific Machine Translation (MT) models.
- Acquired skills in reporting and presenting research findings through clear, concise documentation and presentations.

University of Groningen

Groningen, The Netherlands

RESEARCH TRAINEE

Jun 2022 - Aug 2022

- Investigated entity-centric event sequences to identify and highlight the presence of perspectives. **Supervisor:** Prof. Tommaso Caselli
- Investigated the representation of models after further pre-training with different data sizes and the number of epochs.

Eutech Systems Limited

Dhaka, Bangladesh

MACHINE LEARNING ENGINEER

Nov 2020 - Jul 2021

- Investigated, designed and implemented machine learning applications to solve business problems.
- Lead a cross-functional team in the execution of the project, demonstrating strong leadership and effective collaboration to achieve project milestones.

Education

University of Groningen

Groningen, The Netherlands

RESEARCH MASTER IN LANGUAGE AND COMMUNICATION TECHNOLOGIES

Sept. 2021 - Nov. 2023

- Relevant Courses:** Machine Learning, Deep Learning, Natural Language Processing, Language Technology Project, Hand Written Recognition
- Grade:** 7.4/10

University of Trento

Trento, Italy

MASTER OF COGNITIVE SCIENCE

Sept. 2022 - Oct. 2023

- Relevant Courses:** Grounded Language Processing, Foundation of Cognitive Neuroscience and Psychology
- Master Thesis:** Evaluation of Coreference Resolution Models on Different Domain and Text Genres; **Supervisor:** Prof. Giuseppe Riccardi
- Grade:** 105/110

University of Chittagong

Chattogram, Bangladesh

BACHELOR OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

Jan. 2014 - Oct. 2018

- Bachelor Thesis:** English-to-Bangla Neural Machine Translation Using Transfer Learning
- Grade:** 3.32/4.00

Publications

BD-NLP at SemEval-2024 Task 2: Investigating Generative and Discriminative Models for Clinical Inference with Knowledge Augmentation 📄

Shantanu Nath, and Ahnaf Mozib Samin

In Proceedings of the The 18th International Workshop on Semantic Evaluation (SemEval-2024), Association for Computational Linguistics

GPL at SemEval-2023 Task 1: WordNet and CLIP to Disambiguate Images 📄

Shibingfeng Zhang, **Shantanu Nath**, and Davide Mazzaccara

In Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics

Chest X-ray abnormalities localization via ensemble of deep convolutional neural networks 📄

V.T. Pham, C.M. Tran, S. Zheng, T.M. Vu and **S. Nath**

2021 International Conference on Advanced Technologies for Communications (ATC)

Poster Presentation

S. Nath, G. Minnema, T. Caselli

An Event-Driven Multilingual Football Corpus for Detecting Perspectives and Polarity 📄

The 32nd Meeting of Computational Linguistics in The Netherlands (CLIN32), Tilburg, The Netherlands.

Programming Skills

- **NLP:** Text processing, NLTK, Spacy, sentiment analysis, named entity recognition, Attention Mechanism, Transformers, LLM, T5, BERT, GPT, Image classification, Object detection, YOLO, Faster R-CNN
- **Deep learning:** Pytorch, Keras, Tensorflow, Hugging face
- **Machine learning:** NumPy, SciPy, Matplotlib, OpenCV, scikit-learn, pandas, R and RStudio
- **Web Development:** Mysql, PostgreSQL, Django and Django Rest Framework, HTML, CSS, & JavaScript
- **Others:** Complex Problem Solving, Team Building and Support, Organizational Skills, Github, Linux Development Environment

Honors & Awards

INTERNATIONAL

- 2021 **Erasmus Mundus Scholarship**, Higher Educational Institute, EU
- 2019 **Global Finalist NASA Space App Challenge 2019, NASA**, Finished in the worldwide top-20 by creating a Simulation Game In2theBeginning based on the James Webb Telescope

Projects

Automatic Dataset Translation is Suboptimal: Can Quality Estimation Help? 📄

- The project aims to demonstrate the positive effects of Quality Estimation (QE) on system performance.
- A comparison is made between a system that uses data with low QE scores and one that uses increasingly higher QE scores. In this study, we found that the subset of high-quality instances can be used to create better models than when using the entire training set.

A comparison study of pre-trained and randomly initialized models on image and sequence data 📄

- In this study, the effects of pre-training of deep learning models are investigated. A computer vision model is performed on the identification of leaf diseases on Cassava plants with a pre-trained ResNet-34 model.
- Additionally, a sentiment analysis task is on Coronavirus related tweets using a pre-trained BERT model and an untrained version of BERT.

Offline Handwriting Recognition: Deep Neural Network Approaches 📄

- Evaluate several approaches of digitizing offline handwritten text on two data sets, the Dead Sea Scrolls (DSS) and the IAM data set. With a focus on Deep Learning including ARU-net, CNN, InceptionRes-NetV2, and encoder-decoder architecture
- We were able to achieve Character Error Rates (CER) as far as 5% for distinct tasks.

Is Prompt-based Learning Method a Better Solution for Aspect Detection? 📄

- We made a comparison among multiple prompt learning methods including: tuning-free prompting, fixed-prompt LM tuning and Prompt Tuning using GPT-2 and GPT-Neo.
- We achieved BLEU score of 37.59 using Fixed-prompt LM tuning with GPT-neo and 0.68 BERT F1 score using GPT-2 as base model.

LIST OF REFEREES

Prof. Roberto Zamparelli, PhD

Associate professor

Department of Psychology and Cognitive Science

E-mail: roberto.zamparelli@unitn.it

Tel. +390464 808613

Prof. Giuseppe Riccardi

Full Professor of Computer Science and Information Engineering

Founder and Head of the Signals and Interactive Systems Lab,

Department of Information Engineering and Computer Science,

University of Trento, Italy

Email: giuseppe.riccardi@unitn.it



We hereby certify that **NATH SHANTANU** born in CHATTOGRAM - BANGLADESH - on 27th August 1995, completed his studies in this University and received the **Master Degree in Cognitive Science**(LM-55 - Cognitive Sciences) Computational and theoretical modelling of language and cognition curriculum. We also certify that he passed the following exams with marks expressed in a scale from 18 to 30.

The Master is a TWO - year program.

The course is offered entirely in English.

DSS	EXAM	Credits	MARK	Honour	KEY
ING-INF/05	Grounded Language Processing	9	28		
NN	Internship	15	Pass		
M-PSI/01	Foundation of Cognitive Psychology and Neuroscience	9	19		
M-PSI/02	Basic Statistics	5	26		M
L-LIN/01	Corpus Linguistics	5	28		M
ING-INF/05	Deep Learning	5	25		M
ING-INF/05	Handwriting Recognition	5	30		M
M-PSI/02	Language Technology Project	5	28		M
ING-INF/05	Learning from Data	5	27		M
L-LIN/01	Linguistics Analysis	10	27		M
ING-INF/05	Machine learning	5	28		M
ING-INF/05	Natural Language processing	5	28		M
M-PSI/02	Research training	10	28		M
PROFIN_S	Master Thesis	30	Pass		
		Tot. credits			
		123			

Keys:

DSS	= Disciplinary and scientific sector
AP	= Approved exam
OBG	= Recognized Open Badge
CMPA	= Post-graduate credits already attained, transferrable to current degree course
A	= Recognition from previous career
ITA	= Recognition of credits earned at other Italian universities
M	= International Exchange

On **11th October 2023** he passed his degree examination with **105 / 110**.

Pursuant to the article 40 of the D.P.R. No. 445 of 28 December 2000, this certificate is issued for use abroad only.

The present certificate is issued on unstamped paper for the use foreseen by law. For different use a revenue stamp is to be applied.

TRENTO, 22/11/2023



UNIVERSITÀ
DI TRENTO

20232857141/2023

Student ID number: **239086**



A blue ink signature of Dott. Michele Dell'Orco.

Head of Division

Dott. Michele Dell'Orco

Teaching and Students Services Division
Psychology and Cognitive Science Area

The above information is extracted from the University of Trento database. For official purposes this certificate has to be produced in Italian and English version.



UNIVERSITÀ DI TRENTO

Repubblica Italiana / *Italian Republic*
in nome della Legge / *in the name of the Law*

Noi Professor Flavio Deflorian, Rettore dell'Università degli Studi di Trento,
visti gli attestati degli studi compiuti e il risultato dell'esame di laurea superato in questa Università l'11 ottobre 2023

*We Professor Flavio Deflorian, Rector of the University of Trento,
given the academic records achieved in the graduate program and the results of the graduation examination passed at this University on October 11th, 2023*

conferiamo a / *confer upon*

SHANTANU NATH

nato a Chattogram (Bangladesh) il 27 agosto 1995 / *born in Chattogram (Bangladesh) on August 27th, 1995*

la Laurea Magistrale / *the degree of Master of Science*
in

SCIENZE COGNITIVE / *COGNITIVE SCIENCE*

Classe LM-55 Scienze cognitive / *Class LM-55 Cognitive Science*

Indirizzo Computational and Theoretical Modelling of Language and Cognition / *Track Computational and Theoretical Modelling of Language and Cognition*

Il presente Diploma è valido a tutti gli effetti di legge / *This Certificate is a legally valid document*

Il Direttore del
CIMeC - Centro Interdipartimentale Mente/Cervello
The Director of the
CIMeC - Centre for Mind/Brain Sciences

Yuri Bortoluzzi

Il Rettore
The Rector

Flavio Deflorian

Il Direttore Generale
The General Director

Felice

BOLO ARRETRATO VENT'ANNI
AUTORIZZAZIONE N. 1866/2014 D.D. 26.12.2014 RELASCIATA DALL'AGENZIA DELLE ATTIVITÀ DI TRENTO
TIPOGRAFIA DELL'UNIVERSITÀ DEGLI STUDI DI TRENTO
TRENTO, 15/09/2023



Transcript of Records

Name : S. Nath
Date of birth : 27 August 1995
Student number : 4998405

Code	Title	Date	Mark	Achieved (ECTS)
LTR013M05	Basic Statistics	02/11/2021	7.1	5
LIX016M05	Learning From Data	08/11/2021	7.3	5
WMAI010-05	Machine Learning	25/01/2022	7.5	5
LTR024M05	Corpus Linguistics	05/03/2022	7.6	5
LTR026M10	Linguistic Analysis	14/03/2022	7.3	10
WMAI017-05	Deep Learning	06/04/2022	6.5	5
LIX001M05	Natural Language Processing	14/04/2022	7.8	5
LIX025M05	Language Technology Project	17/06/2022	7.7	5
WMAI019-05	Handwriting Recognition	07/07/2022	8.5	5
LTR00001I	LOT Summer School 2022: Early Perception and Production of Sounds and Words	15/07/2022	P	1
LTR00002I	LOT Summer School 2022: Modern Approaches to the Assessment of Language Impairments	15/07/2022	P	1
LTR00003I	LOT Summer School 2022: Historical Linguistics	15/07/2022	P	1
LTR028M10	Research Training 1	31/08/2022	7.5	10
XXB00001I	Grounded Language Processing	31/08/2023	8.0	9
XXB00002I	Foundation of Cognitive Psychology and Neuroscience	31/08/2023	6.0	9
XXB00003I	Internship	31/08/2023	P	15
LTR999M30	Research Master's Thesis: Linguistics, Including Thesis Class	01/11/2023	7.5	30

Total ECTS credits: 126 (1 ECTS credit = 28 hours of study)

Average grade (weighted average): 7.4

Explanation of grades	
10 = excellent	6 = sufficient
9 = very good	5-1 = fail
8 = good	P = pass
7 = more than sufficient	EX = exemption

Groningen, 12 March 2024



4998405
 S. Nath

050 363 8004
 www.rug.nl/insandouts

Broerstraat 5
 Groningen
 PO Box 72
 9700 AB Groningen
 The Netherlands

www.rug.nl/SIA

Date
 12 March 2024

Examination statement

The Board of the University of Groningen hereby certifies that

Student number:	4998405
Name:	Shantanu Nath
Date of birth:	27 August 1995
Has passed master examination:	Master Linguistics (research) (Language and Communication Technologies) Full-time Groningen
At the faculty:	Faculty of Arts
On:	1 November 2023

The Board of the University of Groningen
 on its behalf:

Head of the Student Information and Administration,

T.K. Idema