Apt 19,Luna Building, Dunmurry, Belfast,UK
+447425066518
pdeka01@qub.ac.uk
p.deka@qub.ac.uk
GitHubRepository
LinkedIn Profile
HuggingFaceRepository

Dr. Pritam Deka

As a highly motivated researcher with expertise in Natural Language Processing (NLP), I am actively seeking a prestigious position in the domain of AI. I am passionate about leveraging my skills and knowledge in NLP to contribute to cutting-edge research projects in artificial intelligence. I am eager to join an organization that values innovation and provides an environment conducive to pushing the boundaries of AI. By combining my strong research background, deep understanding of NLP techniques, and dedication to continuous learning, I am confident in my ability to make significant contributions to the organization's research agenda and advance the field of AI.

Skills

- **Python**: Primary programming language
- **spaCy, scispaCy**: Used for entity recognition and keyword detection tasks
- **C++:** Taught basics of programming using C++ during my lecturer work experience
- **Soft Skills**: Team collaboration, mentoring, technical writing, cross-cultural communication
- **NLTK**: Used for sentence segmentation tasks
- **TensorFlow, PyTorch**: Used for transformer model fine-tuning tasks
- **HuggingFace,SBERT**:Working with transformer models for various domains

Experience

MAY2025-AUGUST2025

Subject Teacher/INTO Queen'sUniversityBelfast

- Deliver lectures in **Object-Oriented Programming using C++** to students.
- Design assignments, assessments, and practical coding labs with a focus on core object-oriented principles.
- Implement student engagement techniques and formative assessment tools for real-time feedback.
- Carry out other necessary duties like examination duties and grading assessments.

MAY2024-ONGOING

ResearchFellow(AI)/Queen'sUniversityBelfast

- Focusing on AI for corporate document understanding, specifically in multimodal data processing, combining text and images for insights.
- Investigating **GPT-based models** and other open-source **LLMs/VLMs** for enhancing corporate data analysis.
- Integrating **NLP** and **prompt engineering** techniques to streamline information extraction.
- Evaluating prompts and models on curated publicly available process diagram dataset.

JULY2023-JANUARY2024

Senior Research Assistant (AI-Security) / University of Southampton

- Led a team to create a comprehensive **dataset** tailored for **Named Entity Recognition (NER)** tasks in **cyber-attack attribution**.
- Developed methods and metrics to assess the quality and effectiveness of the dataset for training NER models.
- Ensured the dataset was suitable for training NER models focused on accurately identifying entities in cybersecurity attack attribution.

NOVEMBER2022

LabDemonstrator/Queen'sUniversityBelfast

- Workedasademonstratorfora**Python**based**DatabaseandProgrammingFundamentals**module supervising the lab practical classes and invigilating lab exams.

MARCH2022–SEPTEMBER2022

AIIntern/Momentone

- Worked as a part time AI intern for Momentone focusing on mental health.
- The focus of the work was to build a chatbot for assisting people who are dealing with mentalhealthissues.
- Theworkinvolvedusageof**Rasa**forthepurposeofbuildingthechatbotalong with **Huggingface** transformer models.

OCTOBER2021–DECEMBER 2021

LabDemonstrator/Queen'sUniversityBelfast

- Workedasademonstratorfora**ComputingFoundations**moduleanda**WebDevelopment**module which focused on **HTML5, CSS, JavaScript, SQL and PHP**.

FEB2021 –MAY 2021

TeachingAssistant/Queen'sUniversityBelfast

- Worked as a TA for a **Database and Programming Fundamentals** module where the primary task wasto supervisethelabsessionsandmarktheassignments.
- Themodulewasbasedon **Python**and **SQL concepts**.

AUG2018 –SEP2019

Lecturer/RoyalGlobalUniversity,Assam,India

- Worked as a Lecturer in the Department of Computer Science, responsible for teaching, assignment design, and student evaluation.
- Supervised undergraduate project work and contributed to curriculum development and departmental administration.
- Taught modules including **"Introduction to Python", "Object oriented programming with C++", "Introduction to Computing with C", "Design and analysis of algorithms"** and **"Data structures using C"**.

Education

OCTOBER2019–DEC 2024

PhD/Queen'sUniversityBelfast

- **Developed and evaluated an end-to-end health information verification pipeline**: Utilized state-of-the-art NLP techniques like **BERT** to verify **online health information**.
- **Focused on unsupervised learning**: Created algorithms that function in data-scarce environments, enabling effective verification without large, labeled datasets.
- Leveraged libraries such as **Huggingface**, **SBERT**, **NLTK**, **spaCy**, and **scispaCy** to implement and experiment with the verification pipeline.
- Fine-tuned and uploaded custom transformer models to **Huggingface** during PhD research for public use.

JUNE2014–JULY2016

MTech(I.T.)/TezpurUniversity,Assam,India

- Using a probabilistic model the competition between two languages was modeled based on certain factors which can determine the coexistence or the death of one language.
- The model has an extension which uses **Markov Chain modeling**.
- For experiments, I used **Matlab** as the programming language.

JUNE2009–JULY2013

BE(CSE)/GauhatiUniversity,Assam,India

- Developed a notepad application having a reminder and to-do option using the **Android SDK** based on **Java**.

AcademicPapers

- **Unsupervised Keyword Combination Query Generation from Online Health Related Content for Evidence-Based Fact Checking (P.Deka,** A.Jurek-Loughrey, D.P**)** which got the best paper award in **iiWAS21**. The DOI of the paper: 10.1145/3487664.3487701
- **Improved Methods To Aid Unsupervised Evidence-Based Fact Checking For Online Health News (P.Deka,** A.Jurek-Loughrey, D.P**)** published in **Journal of Data Intelligence**.The DOI of the paper: 10.26421/JDI3.4-5
- **Evidence Extraction to Validate Medical Claims in Fake News Detection (P.Deka,**A.Jurek-Loughrey, D.P**)** in **HIS2022.** The DOI of the paper: 10.1007/978-3-031-20627-6_1
- **Multiple Evidence Combination for Fact-Checking of Health-Related Information (P.Deka,** A. Jurek-Loughrey, D.P**)** in **BioNLP2023** @ **ACL2023**. The DOI of the paper: 10.18653/v1/2023.bionlp-1.20
- **BERT-based Language Identification in Code-Mix Kannada-English Text at the CoLI-Kanglish Shared Task@ICON2022 (P.Deka,**N.J.Kalita,S.K.Sarma**)** published in **ICON2022**. TheURLof the paper: https://aclanthology.org/2022.icon-wlli.3
- **BERT Based Language Identification in Code-Mixed English-Assamese Social Media Text (**NJ Kalita, P Deka, V Chennareddy, SK Sarma**)** published in **Springer Algorithms for Intelligent Systems.** The DOI of the paper: https://doi.org/10.1007/978-981-99-1620-7_14
- **PD-AR at ArAIEval Shared Task: A BERT-Centric Approach to Tackle Arabic Disinformation (**Pritam Deka, Ashwathy Revi**)** published in ArabicNLP 2023. The DOI of the paper: 10.18653/v1/2023.arabicnlp-1.57
- **A New Probabilistic Model of Language Competition, Endangerment and Extinction (P. Deka,** S.K. Sinha**)** in **IJCA**. The DOI of the paper: 10.5120/ijca2016909958
- **AttackER: Towards Enhancing Cyber-Attack Attribution with a Named Entity Recognition Dataset (P. Deka et. al)** in **WISE2024**. The DOI of the paper: https://doi.org/10.1007/978-981-96-0576-7_20
- **Structured Extraction from Business Process Diagrams Using Vision-Language Models (P. Deka and B. Devereux)** Accepted for presenting and publication at ACM SAC 2026.

AcademicActivities

- Released fine-tuned language models in HuggingFace repository which can be accessed at shorturl.at/egjsM
- Created two online demos based on part of my PhD work which can be accessed at shorturl.at/DX347and shorturl.at/eimuz
- The demos are created using **Python** and **Gradio** and uses several libraries such as the state-of-the-art NLP libraries like **sentence transformers**, **Huggingface**.
- Was a reviewer for the **ICON2021** conference which was held in NIT Silchar, Assam, India.
- Was a guest lecturer for a module of Deep Learning at Queen's University Belfast.
- Was a reviewer for **The 2024 ACM/SIGAPP Symposium on Applied Computing.**
- Program Committee member for **The 2023 International Conference on Natural Language Processing (ICON)**

OutreachActivities

- Was a speaker for a webinar on"**Research and Innovation on Information Technology**" at Universitas Islam Negeri Walisongo, Indonesia.
- Gave a radio talk at **University of Calicut** regarding fake health information.
- Was a main speaker for **ElevateNI 2023** held in Queen's University Belfast and delivered a talk on "**The role of large language models in AI**".
- Was a speaker at AICON 2024 held at Titanic Belfast and delivered a talk on "**AI-Powered Information Extraction**".
- Was a main speaker at the **Vision and Lanuage Symposium 2025** held at Queen's University Belfast.

Referees

- Dr.AnnaJurek-Loughrey
  SeniorLecturer,SchoolofEEECS
  Queen's University Belfast
  Email: a.jurek@qub.ac.uk
  Tel:+44(0)2890974484
- Dr. Deepak Padmanabhan
  SeniorLecturer,SchoolofEEECS
  Queen's University Belfast
  Email:D.Padmanabhan@qub.ac.uk
  Tel: +44 (0)28 9097 4874