# Proyag Pal

Edinburgh, UK
☎ (+44)07513660849
✉ proyag.pal@ed.ac.uk
proyag.github.io
www.linkedin.com/in/proyag-pal
github.com/Proyag

## Experience

Jun 2020 – Oct 2020 Amsterdam
**Data Engineer**, *TAUS*.
Working on the EU-funded ParaCrawl project to collect parallel corpora from large-scale web crawls.
- Optimised, maintained, and ran a highly scalable processing pipeline to extract, translate, align, and clean parallel corpora obtained through web crawling.

Feb 2020 – Apr 2020 Lisbon
**Junior AI Researcher**, *Unbabel*, Applied AI.
Machine translation and quality estimation for customer-facing products.
- Built domain-specific machine translation models.
- Built quality estimation models to skip human post-editing for high-quality MT output.

Feb 2018 – Jan 2020 Geneva
**Fellow in Neural Machine Translation**, *World Intellectual Property Organization (WIPO)*, Advanced Technology Applications Center.
Development and maintenance of WIPO Translate and related NLP tools and technologies.
- *WIPO Translate*: Built, improved, evaluated and deployed domain-specific neural and statistical machine translation models using the Marian and Moses toolkits.
- *IPCCAT*: Developed neural text classification systems for patent categorisation.
- Developed a system using BERT/XLM/LASER sentence representations and Faiss indexes to retrieve semantically similar content from large collections of text.

Sep 2017 – Dec 2017 Edinburgh
**Research Assistant**, *University of Edinburgh*, ILCC, School of Informatics.
- Worked on developing isiXhosa-English machine translation to facilitate doctor-patient communication in health centres in South Africa.
- Collected corpora released as a public resource.

## Education

2020 – 2023 Edinburgh
**PhD Informatics**, *University of Edinburgh*, in progress.
Low-resource machine translation. Supervised by Dr. Kenneth Heafield and Dr. Alexandra Birch.

2016 – 2017 Edinburgh
**M.Sc. Informatics**, *University of Edinburgh*, with Distinction.
*Selected Courses:* Machine Translation, Accelerated Natural Language Processing

2011 – 2016 Kolkata
**B.Sc. & M.Sc. Computer Science**, *St. Xavier's College*.
*Selected Courses:* Artificial Intelligence, Data Mining & Warehousing, Computer Architecture

## Master's Projects

Jun 2017 – Aug 2017
**Reward Augmented Maximum Likelihood to Improve Neural Machine Translation Training**, *University of Edinburgh*, supervised by Dr. Kenneth Heafield.
- Used reinforcement learning - inspired task rewards to augment the training objective.
- Improved upon the University of Edinburgh's strong baseline by 1.07 BLEU.

Aug 2015 – May 2016
**Permutation Flow Shop Scheduling using Natural Algorithms**, *St. Xavier's College, Kolkata*, supervised by Prof. Siladitya Mukherjee.
- Optimization of makespan in permutation flow shop scheduling, using genetic algorithms.

## Programming

**Python**, *advanced*, with PyTorch, NumPy, sklearn.
**C++**, *intermediate*.
**Julia, Perl, Bash, Docker, LaTeX**.

## Languages

**English, Bengali**, *Native/Bilingual*.
**French**, *Conversational*.
**Chinese (Mandarin)**, *Basic*.
**Hindi**, *Fluent*.