PRADIPTA DEB

Research Engineer - ML, Privacy & Cybersecurity

☎ (+33) 752115162
✉ deb.pradipta@gmail.com
www.pradiptadeb.com
/pradipta-deb
Google Scholar
Lille, France

ABOUT ME

With more than ten years of expertise, I, Pradipta Deb, specialize in privacy-preserving AI and large-scale machine learning systems. I have demonstrated expertise in designing and implementing commercial machine learning systems that handle over six billion daily interactions. I'm knowledgeable about creating innovative solutions in the fields of computer vision, NLP, and LLMs with an emphasis on security and privacy protection. I'm a distinguished technical leader with expertise in leading cross-organizational projects through the AI Committee of M3AAWG and managing AI teams. I'm a seasoned researcher who has worked on optimization algorithms, speech synthesis, and privacy-preserving machine learning.

EXPERIENCE

M3AAWG
AUG 2024 – PRESENT

Vice Chair, Artificial Intelligence Committee
GLOBAL

- Working with international group of 50+ organizations in developing AI/ML standards for cyber abuse prevention
- Established framework for responsible LLM deployment in email security, adopted by major industry players
- Coordinating cross-industry initiatives for secure and privacy-preserving AI in cybersecurity
- Authored technical guidelines for large language model deployment in anti-abuse system
- Provided talks and tutorial sessions regarding LLM and it's safe uses

Hornetsecurity (formerly Vade)
AUG 2022 – PRESENT

Research Engineer
HEM, FRANCE

- Leading AI research initiatives for email security platform protecting 6B+ mailboxes globally
- Developed and deployed LLM-based detection system for Business Email Compromise (BEC) attacks, achieving 95% detection rate
- Implemented fine-tuned language models for sophisticated threat detection (spam, phishing, impersonation attempts)
- Pioneered integration of Vision LLMs (Pixtral) for webpage scam detection, reducing false positives by 40%
- Architected scalable ML pipeline processing billions of daily emails while maintaining GDPR compliance
- Designed agent-based architecture for autonomous threat investigation and response
- Mentored and established best practices for LLM deployment, monitoring in production within the hornet's research team.

Contract.Fit
MAY 2021 – JUL 2022

Senior AI Engineer
BRUSSELS, BELGIUM

- Led development of ML systems for information extraction from diverse document types (invoices, insurance forms, medical records, purchase orders)
- Implemented transformer-based NLP models for multi-language document classification across 140+ document types
- Developed computer vision solutions for detecting embedded objects (ID cards, medical insurance cards, driving licenses)
- Architected end-to-end ML pipeline processing 100K+ daily documents with 98% accuracy while prioritizing automation
- Served as Lead Developer and Tech Analyst for 4 major clients, bridging technical solutions with business requirements
- Collaborated with sales/delivery teams to create implementation roadmaps and ensure smooth project deployment
- Led team of 4 AI engineers and supervised interns, delivering projects worth €2M+ while maintaining 99.9% SLA compliance

Team Magnet, INRIA

OCT 2019 – FEB 2021

Research and Development Engineer

LILLE, FRANCE

- Developed collaborative computation platform for privacy-preserving statistical analysis using U-statistics
- Implemented decentralized learning framework with differential privacy, reducing privacy budget by 45%
- Created domain-specific language for privacy policy specification in collaboration with PhD researchers
- Optimized homomorphic encryption for low-resource devices, reducing computation overhead by 60%

MLT Lab, DFKI

OCT 2016 – DEC 2018

Research Assistant

GERMANY

- Led development of MaryTTS, improving voice quality by 40% through advanced neural synthesis techniques
- Achieved 2nd place in Blizzard Challenge 2018 and participated in Blizzard 2017 for speech synthesis
- Developed multi-speaker, multilingual speech synthesis system using deep learning for master thesis
- Innovated hybrid approach for grapheme-to-phoneme conversion, achieving state-of-the-art accuracy for 5 languages
- Created automated testing framework reducing development cycle time by 50%

Tata Consultancy Services

JUL 2012 – SEP 2015

Software Engineer

INDIA

- Developed and maintained enterprise applications for Fortune 500 clients including Johnson & Johnson and ALCOA
- Specialized in full-stack development using Java, C++, and Python technologies
- Collaborated in cross-functional teams to deliver high-quality software solutions
- Implemented automated testing frameworks improving code coverage by 40%

EDUCATION

Saarland University

JULY, 2019

MS, Computer Science

SAARLAND, GERMANY

West Bengal University of Technology

AUGUST, 2012

BTech., Information Technology

WEST BENGAL, INDIA

TECHNICAL EXPERTISE

AI/ML Expertise
Large Language Models (LLaMA, GPT) · Transformer Architectures · Neural Networks · Computer Vision · NLP · Speech Processing · Document AI · GraphML

ML Engineering
Model Deployment · A/B Testing · Model Monitoring · MLOps · Distributed Training · Performance Optimization

Privacy & Security
Differential Privacy · Homomorphic Encryption · Federated Learning · GDPR Compliance · Secure ML Pipeline Design

ML Frameworks
PyTorch · TensorFlow · Hugging Face · Fastai · scikit-learn · MLflow

Cloud & Infrastructure
Docker · Kubernetes · Azure ML · AWS SageMaker · CI/CD · Git

Data Engineering
Python · SQL · Pandas · NumPy · MongoDB · GraphDB · Data Pipeline Design

Leadership
Team Management · Technical Mentorship · Project Planning · Cross-functional Collaboration · Research Direction

LANGUAGES

English (Professional) · French (Intermediate) · German (Elementary) · Bengali (Native) · Hindi (Professional)

TALKS AND TUTORIALS

- **Large Scale Private Text Generation and Information Retrieval with Large Language Models**
  M3AAWG 63rd General Meeting, *Lisbon, Portugal (Feb 2025)*
- **Mastering Large Language Models**
  M3AAWG 61st General Meeting, *Vienna, Austria (June 2024)*
- **Demystifying Large Language Models Applied to Anti-Abuse**
  M3AAWG 60th General Meeting, *San Francisco, USA (Feb 2024)*
  Blog post: Demystifying Large Language Models: An Introduction

SELECTED PUBLICATIONS

- Le Maguer, S., Steiner, I., Tombini, F., **Deb, P.**, Basu, M., & Kröger, I. (2018). **Agile MaryTTS Architecture for the Blizzard Challenge 2018**
- Basu, M., **Deb, P.**, & Garai, G. (2014). **Hybrid of Particle Swarm Optimization and Simulated Annealing for Multidimensional Function Optimization.** International Journal of Information Technology, 20(1), 112–20
- **Deb, P.**, & Basu, M. (2014). **Next Generation Sequencing Using Ant Colony Optimization Algorithm and Binary Particle Swarm Optimizers.** International Journal of Information Technology, 20(1)
- Banerjee, C., Kundu, A., Basu, M., **Deb, P.**, Nag, D., & Dattagupta, R. (2013). **A Service Based Trust Management Classifier Approach for Cloud Security.** In 2013 15th International Conference on Advanced Computing Technologies (ICACT), 1-5
- **Deb, P.**, & Basu, M. (2013). **An Analogy of Algorithms for Tagging of Single Nucleotide Polymorphism and Evaluation Through Linkage Disequilibrium.** International Journal of Computer Engineering & Technology, 4(4)

AWARDS

- **IEEE International Conference on Advanced Computing Technologies** - Best Paper Award (2013)
- **Tata Consultancy Services Limited** - Kudos Certification (2012-13)
- **Government of India** - National Merit Scholarship (2006)

SELECTED CERTIFICATIONS

- **AI for Medicine Specialization** (Coursera—DL.AI, 2020)
  - AI for Medical Diagnosis
  - AI for Medical Prognosis
  - AI for Medical Treatment
- **Mathematics for Machine Learning Specialization** (Imperial College London, 2019)
- **Deep Learning Specialization** (Coursera—DL.AI, 2018)

Additional certifications, publications and project details are available at pradiptadeb.com