Judith Jeyafreeda Andrew

judithjeyafreeda@gmail.com | +33 6 99 34 23 39

RESEARCH PROFILE

Experienced researcher specializing in trustworthy artificial intelligence for textual and multimodal data, with proven expertise in clinical NLP, multimodal learning, bias detection, and low-resource language processing. Strong publication record in top-tier venues (ACM TOIS, LREC-COLING, Clinical NLP Workshop) and extensive teaching experience across undergraduate and graduate programs.

Research Interests: Natural Language Processing • Clinical NLP • Explainable AI • Low-Resource Learning • Temporal Information Extraction

CURRENT POSITION

Research Engineer | April 2026 – Present

CHU Bordeaux

- Working on the ARIPPA project to determine if the duration of the prescribed antibiotics for patients follow the medical rule
- I am applying Deep Learning methods and exploring LLM for information extraction from text

EDUCATION

PhD in Computer Science | 2017–2020
Université de Caen Normandie, France

- **Thesis:** Task-Oriented Web Page Segmentation
- **Supervisors:** Dr. Gaël Dias, Dr. Stéphane Ferrari, Dr. Fabrice Maurel
- **Key Contributions:** Developed multimodal clustering approaches for web page segmentation combining textual, visual, and structural features; published in ACM Transactions on Information Systems (TOIS)

Master in Decision Support and Business Intelligence | 2016–2017
École Centrale Paris (now CentraleSupélec), Paris-Saclay

- Specialization in Machine Learning and Data Analytics
- Internship at LIMSI (now LISN) on Named Entity Recognition

Bachelor of Engineering in Computer Science | 2012–2016
PSG College of Technology, India

- First Class with Distinction

PREVIOUS POSITIONS

Postdoctoral Researcher | June 2023 - November 2025

Université Paris Cité, INSERM UMRS 1138, Paris, France

Research Focus: Temporal information extraction from French clinical texts for rare pediatric diseases

Key Achievements:

- Developed LLM-based approaches for extracting temporal entities and relations from pediatric clinical narratives

- Published 2 papers at international venues (CL4Health@LREC-COLING 2024, JAMIA Open 2025)
- Collaborated with clinicians at Gustave Roussy and Necker Hospital on real-world clinical applications
- Contributed to shared tasks (ChemoTimelines 2024) demonstrating practical NLP deployment

Technical Skills Applied: Large Language Models (GPT, LLaMA), Fine-tuning, Temporal Relation Extraction, French Clinical NLP, Low-Resource Adaptation

Postdoctoral Researcher | September 2021 – June 2023

University of Manchester, United Kingdom

Research Focus: Clinical NLP for de-identification and named entity recognition from outpatient letters

Key Achievements:

- Developed privacy-preserving de-identification methods for clinical documents (paper under review in JMIR AI)
- Implemented NER systems for extracting medical entities from unstructured clinical texts
- Supervised 1 MSc student on de-identification methods and 7 undergraduate students on web design projects
- **Teaching:** Web Design for undergraduate students (practical sessions)

Technical Skills Applied: Clinical NER, De-identification, Privacy-Preserving NLP, BERT-based Models, Annotation Guidelines

ATER (Attaché Temporaire d'Enseignement et Recherche) | 100% | September 2020 – August 2021

Université de Caen Normandie, France

Responsibilities:

- **Teaching (128 hours/year):** Introduction to Programming (C, Python), Web Technologies (HTML, CSS, JavaScript, PHP)
- **Research:** Continued NLP research on information extraction and document understanding
- **Student Level:** 1st year Bachelor students in Computer Science

Teaching Approach: Hands-on practical sessions with emphasis on problem-solving and real-world applications

TEACHING EXPERIENCE

University-Level Teaching (Total: ~250+ hours)

Teaching at University of Manchester (2021–2023)

- **Web Design** – Undergraduate level · **Project Supervision:**
  - Supervised 7 undergraduate students on web design group project ○ Supervised 1 MSc student on Masters thesis (de-identification of clinical texts)

Teaching at Université de Caen Normandie (2017–2021)

- **Python Programming** – 1st year Bachelor (64 hours/year, 2020–2021)
- **Web Technologies** (HTML, CSS, JavaScript, PHP) – 1st year Bachelor (48hours/year, 2018–2021)

Teaching Philosophy

- **Active Learning:** Emphasize hands-on implementations and practical problem-solving
- **Research Integration:** Connect fundamental concepts to cutting-edge research and real-world applications
- **Inclusive Pedagogy:** Adapt teaching methods to diverse student backgrounds and learning styles
- **Practical Skills:** Balance theoretical foundations with practical programming and deployment skills

Language Capability: Fluent in English for teaching; can teach all courses in English

RESEARCH EXPERIENCE G CONTRIBUTIONS

Multimodal Learning

- Developed self-organized multi-objective clustering for multimodal web page segmentation
- Combined textual, visual, and structural features for document understanding
- **Skills:** Multi-objective optimization, clustering algorithms, feature fusion

Clinical NLP with Limited Resources

Low-Resource Adaptation:

- Evaluated LLMs for temporal entity extraction in rare diseases context (CL4Health 2024)
- Demonstrated that smaller fine-tuned models compete with large LLMs (ChemoTimelines 2024)
- **Skills:** Transfer learning, few-shot learning, domain adaptation, clinical data annotation

Temporal Information Extraction:

- Developed methods for extracting temporal relations from French clinical narratives
- Published in JAMIA Open (2025), manuscript in preparation for IEEE JBHI
- **Skills:** Temporal reasoning, relation extraction, sequence labeling

Bias Detection & Fairness

Shared Task Participation:

- Machine Learning for Explainable Detection of Online Sexism (SemEval 2023)
- Homophobia/Transphobia detection using GPT models (LT-EDI Workshop 2023)
- **Skills:** Bias detection, fairness metrics, explainability methods, GPT fine-tuning

Information Extraction

Legal & Web Documents:

- Automatic extraction of entities and relations from legal documents (Named Entities Workshop 2018)
- Web page segmentation for non-visual accessibility (PACLIC 2019, HDI@ICDAR 2019)

PUBLICATIONS

Peer-Reviewed Journal Articles (4)

1. Srivatsa Ramesh, Gaël Dias, **Judith Jeyafreeda Andrew**, Sriparna Saha, Fabrice Maurel, Stéphane Ferrari. "Multimodal Web Page Segmentation Using Self-organized Multi-objective Clustering." *ACM Transactions on Information Systems (TOIS)*, 40(3), 59:1-59:49, March 2022.
2. **Judith Jeyafreeda Andrew**, Juliette Poitier, Nicolas Garcelon, Anita Burgun, Marc Vincent. "Using Large Language Models for Temporal Relation Extraction from Pediatric Clinical Reports." *JAMIA Open*, ooaf121, 2025. https://doi.org/10.1093/jamiaopen/ooaf121

3. Coppola C, **Andrew JJ**, Ruggieri M, La Spina M, La Bianca MR and Leonardi S (2026) Machine learning prediction of oxygen therapy in pediatric Mycoplasma pneumoniae pneumonia. Front. Digit. Health 8:1755878. doi: 10.3389/fdgth.2026.1755878
4. **Judith Jeyafreeda Andrew**, Juliette Poitier, Nicolas Garcelon, Anita Burgun, Marc Vincent. "From Annotation to Adaptation: Extracting Temporal Relations in French Clinical Narratives." BMC Medical Research Methodology. DOI : https://doi.org/10.1186/s12874-026-02867-4 May 2026

Conference Proceedings

4. **Judith Jeyafreeda Andrew**, Marc Vincent, Anita Burgun, Nicolas Garcelon. "Evaluating LLMs for Temporal Entity Extraction from Pediatric Clinical Text in Rare Diseases Context." *Proceedings of CL4Health @ LREC-COLING 2024*, pages 145-152, Torino, Italia, 2024.
5. Nesrine Bannour, **Judith Jeyafreeda Andrew**, Marc Vincent. "Team NLPeers at Chemotimelines 2024: Evaluation of two timeline extraction methods." *Proceedings of the 6th Clinical NLP Workshop*, 2024.
6. **Judith Jeyafreeda Andrew**. "Machine Learning for Explainable Detection of Online Sexism." *SemEval-2023*, 2023.
7. **Judith Jeyafreeda Andrew**. "Using GPT model for recognition of Homophobia/Transphobia detection from social media." *LT-EDI Workshop*, 2023.
8. **Judith Jeyafreeda Andrew**, Stéphane Ferrari, Fabrice Maurel, Gaël Dias, Emmanuel Giguet. "Web Page Segmentation for Non Visual Skimming." *PACLIC 33*, Hakodate, Japan, 2019.
9. Fabrice Maurel, Gaël Dias, Stéphane Ferrari, **Judith Jeyafreeda Andrew**, Emmanuel Giguet. "Concurrent Speech Synthesis to Improve Document First Glance for the Blind." *HDI@ICDAR 2019*, pages 10-17, Sydney, Australia, 2019.
10. **Judith Jeyafreeda Andrew**, Stéphane Ferrari, Fabrice Maurel, Gaël Dias, Emmanuel Giguet. "Model- driven Web Page Segmentation for Non Visual Access." *6th International Conference of PACLING 2019*, Hanoï City, Vietnam, 2019.
11. **Judith Jeyafreeda Andrew**. "Automatic extraction of entities and relation from legal documents." *Seventh Named Entities Workshop*, pages 1-8, 2018.
12. **Judith Jeyafreeda Andrew**. "JudithJeyafreeda_StressIdent_LT-EDI@EACL2024: GPT for stress identification." *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 173-176, St. Julian's, Malta, 2024.

Under Review

Judith Jeyafreeda Andrew. "Toward Verifiable Clinical NLP via Provenance-First Architectures and Temporal Knowledge Graphs" [Under Review, NPJ Digital Medicine]

Shared Task Participation (4)

- **Dravidian-CodeMix-FIRE2020:** Sentiment Analysis (FIRE 2020, Hyderabad)
- **DravidianLangTech-EACL2021:** Offensive language detection (EACL 2021)
- **TamilNLP-ACL2022:** Emotion Analysis in Tamil (ACL 2022, Dublin)
- **ChemoTimelines 2024:** Timeline extraction from clinical texts (Clinical NLP Workshop 2024)

INVITED TALKS & SEMINARS

1. "Perspectives in Clinical NLP – privacy and temporality" – ISPED Bordeaux, 2025
2. "Extracting temporal relations from clinical texts" – Telecom Paris, 2025
3. "Temporality in Clinical Narratives" – GdR Traitement Informatique des Données de Santé, Paris, October 2025
4. "Using LLM for extracting temporal entities and relations in French Clinical Text" – GdR MADICS, Toulouse, May 2025
5. Clinical NLP – IRIT, Toulouse, 2025
6. Clinical NLP – LS2N, Nantes, 2025
7. Clinical NLP – LISN, Paris Saclay, 2025
8. Clinical NLP – LORIA, Nancy, January 2026

Impact: Established visibility in French NLP research community through invited seminars at major research laboratories

TECHNICAL SKILLS

Programming Languages

- **Expert:** Python, R
- **Proficient:** C, C++, Java, SQL
- **Familiar:** JavaScript, PHP, HTML/CSS

AI/ML Frameworks & Tools

- **Deep Learning:** PyTorch, TensorFlow, Hugging Face Transformers
- **NLP:** spaCy, NLTK, Stanford CoreNLP, Stanza
- **ML Tools:** scikit-learn, WEKA, Pandas, NumPy
- **Visualization:** Matplotlib, Seaborn, Tableau
- **Version Control:** Git, GitHub

Specialized AI Expertise

- **Large Language Models:** GPT, LLaMA, BERT, CamemBERT, fine-tuning, prompt engineering
- **Multimodal Learning:** Vision-language models, cross-modal fusion, attention mechanisms
- **Clinical NLP:** Named entity recognition, temporal relation extraction, de-identification
- **Bias & Fairness:** Bias detection, fairness metrics, adversarial debiasing
- **Low-Resource NLP:** Transfer learning, few-shot learning, domain adaptatio

PROFESSIONAL SERVICE & COMMUNITY ENGAGEMENT

Programme Committee / Reviewing

- Regular Reviewer: LT-EDI Workshop, Dravidian CodeMix workshops
- Reviewer: Workshop on Computational Linguistics for Health, 2025 and 2026

Volunteer Activities

- Volunteer at EACL 2021
- Volunteer at ACL 2022

Professional Membership

- ATALA (Association pour le Traitement Automatique des Langues) – Active Member

Research Community Contributions

- Participated in PhD Symposium, 18th International Conference on Web Engineering (WebSci 2018)
- Attended International Research Summer School on Statistics for Data Science (S4D), Caen, June 2019
- Poster presentation at GdR MADICS, Toulouse, France. 2025
- Poster Presentation at CORIA-TALN, Marseille, France. 2025

INTERNATIONAL MOBILITY COLLABORATION

France (Primary Location)

- PhD at Université de Caen Normandie (3 years)
- ATER at Université de Caen Normandie (1 year)
- Postdoc at Université Paris Cité (2 years)

United Kingdom

- Postdoc at University of Manchester (2 years)

India

- Undergraduate studies at PSG College of Technology (4 years)

International Research presentations

- Multiple invited seminars across French research laboratories (2025)
- Conference presentations in Australia, Japan, Vietnam

International Collaborations

- Collaboration with hospitals in Catania and Trapani, Italy. Starting in 2025

Language Skills

- **Tamil:** Native
- **English:** Fluent (C2 level) – Capable of teaching all courses in English
- **French:** Intermediate (C1 level) – Actively improving

Interests

- Hiking – Forest, Mountains and Lakes
- Yoga and Pilates
- Traveling and learning about new cultures
- Thriller Novels and Movies