

الجمهورية الجزائرية الديمقراطية الشعبية
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
وزارة التعليم العالي والبحث العلمي
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH



HIGHER NATIONAL SCHOOL OF COMPUTER
SCIENCE -08 MAI 1945- SIDI BEL ABBÈS -

Multi-Modal Data Extraction and Analysis
of Digital Labour-Market Intermediaries

Presented by **AYOUB FRIHAOUI**

THESIS

To obtain the diploma of **Master**

Field: **Computer Science**

Speciality: **Artificial Intelligence and Data Science (AIDS)**

Supervised By

MOHAMMED BEKKOUCHE

OLIVIER PONS

LÉA LIMA

Submission Date: **Sep, 2025**

In front of the jury composed of:

BELKACEM KHALDI
BADIA KLOUCHE

Academic Year: 2024/2025

Abstract

Digital platforms increasingly mediate access to employment opportunities, yet their role in shaping labor market inequalities remains insufficiently understood. This thesis presents a comprehensive analysis of Digital Labor Market Intermediaries (DLMIs) across France and Brazil, focusing on care work platforms where gendered and racialized inequalities are particularly pronounced. Through innovative methodological approaches combining large-scale web scraping, multimodal AI inference, and explainable machine learning techniques, the research reveals systematic patterns of algorithmic bias in worker representation, pricing mechanisms, and platform success factors.

This internship developed a scalable data engineering and AI pipeline to study how digital intermediation platforms shape labor market inequalities. By reverse engineering diverse platform APIs, the system automated extraction and quality validation of tens of thousands of caregiver profiles across multiple countries. Integrated multimodal AI models inferred rich demographic attributes, enabling large-scale, explainable analysis of algorithmic bias in pricing, temporal advantages, and geographic concentration. The project creates a foundation for real-time monitoring, fairness-aware recommendation systems, and evidence-based policy guidance to promote accountability and equity in the digital economy.

Keywords: Digital Labor Markets, Algorithmic Bias, Data Engineering, Multimodal AI, Platform Economics, Web Scraping, Care Work, Explainable Machine Learning

ملخص

تزايد وساطة المنصات الرقمية في الوصول إلى فرص العمل، ومع ذلك يبقى دورها في تشكيل عدم المساواة في سوق العمل غير مفهوم بشكل كافٍ. تقدّم هذه الأطروحة تحليلاً شاملاً لوسطاء سوق العمل الرقمي في فرنسا والبرازيل، مع التركيز على منصات عمل الرعاية حيث تكون أوجه عدم المساواة المرتبطة بالنوع الاجتماعي والانتماء العرقي بارزة بشكل خاص. من خلال مقاربات منهجية مبتكرة تجمع بين تقنيات تجريف الويب واسعة النطاق، واستدلال الذكاء الاصطناعي متعدد الوسائط، وأساليب التعلم الآلي القابل للتفسير، تكشف هذه الدراسة عن أنماط منهجية من التحيز الخوارزمي في تمثيل العمال، وآليات التسعير، وعوامل نجاح المنصات.

طوّر هذا التدريب مهّداً لهندسة بيانات ونظام ذكاء اصطناعي قابلين للتوسع لدراسة كيفية تشكيل منصات الوساطة الرقمية لعدم المساواة في سوق العمل. من خلال إجراء هندسة عكسية لواجهات برمجة تطبيقات منصات متنوعة، أتاح النظام أتمتة استخراج والتحقق من جودة عشرات الآلاف من ملفات مقدمي الرعاية عبر عدة دول. استنتجت نماذج ذكاء اصطناعي متعددة الوسائط متكاملة سمات ديموغرافية غنية، مما مكن من إجراء تحليل واسع النطاق وقابل للتفسير للتحيز الخوارزمي في التسعير، والمزايا الزمنية، والتركيز الجغرافي. يؤسس المشروع قاعدة قوية للمراقبة الآنية، وأنظمة توصية تراعي العدالة، وإرشادات سياسية قائمة على الأدلة لتعزيز المساءلة والإنصاف في الاقتصاد الرقمي.

الكلمات المفتاحية:

سوق العمل الرقمي، التحيز الخوارزمي، هندسة البيانات، الذكاء الاصطناعي متعدد الوسائط، اقتصاد المنصات، تجريف الويب، عمل الرعاية، التعلم الآلي القابل للتفسير