



Mauro Pelucchi

● ABOUT ME

I am Senior Data Scientist and Big Data Engineer responsible for the design of the “Real-Time Labour Market Information System on Skill Requirements” for CEDEFOP (European Centre for the Development of Vocational Training).

I currently work as Head of Global Data Science at Lightcast with the goal to develop innovative models, methods, and deployments of labour market data and other data to meet customer requirements and prototype new potential solutions.

My main tasks are related to advanced machine learning modelling, labour market analyses, and the design of big data pipelines to process large datasets of online job vacancies. In collaboration with the University of Milano-Bicocca, I took part in many research projects related to the labour market intelligence systems. I collaborate with the University of Milano-Bicocca as a Lecturer for the Masters of Business Intelligence and Big Data Analytics and with the University of Bergamo as a Lecturer in Computer Engineering.

● WORK EXPERIENCE

11/2016 – CURRENT Milan, Italy

HEAD OF GLOBAL DATA SCIENCE LIGHTCAST - BURNING GLASS EUROPE SRL

SISTAL Project - Labour Market Intelligence

- Developed and maintained data pipelines for statistical labour information using Talend, Pentaho, MySQL, and Vertica.
- Implemented predictive analytics to monitor labour market trends.

CEDEFOP (European Centre for the Development of Vocational Training) - 2016-2020 “Real-time Labour Market information on Skill Requirements: Setting up the EU system for online vacancy analysis”

- Designed and implemented big data processing pipelines for analysing online job vacancies (OJVs).
- Utilized NLP models (BERT, SBERT) and ontology matching (ESCO, ISCO, O*NET) to enhance skills classification.
- Employed PySpark, AWS EMR, and Apache AirFlow for scalable data processing.

EXCELSIOR – Employment and Training Information System (UNIONCAMERE)

- Developed dashboards and data visualizations using Power BI and Tableau to track job market demand.
- Implemented Elasticsearch for fast indexing and retrieval of job postings data.

Luxembourg Institute of Socio-Economic Research (LISER)

- Conducted large-scale labour market analysis integrating ESCO, ISCO, and NACE taxonomies.
- Utilized SQL and graph databases (MongoDB) for cross-country occupational comparisons.

ATLAS – Global Data Platform for Labour Market Analysis

- Architected and implemented a cloud-based platform using AWS Lambda, Beanstalk, and AWS Batch.
- Integrated NLP pipelines with custom neural networks for job classification and skills forecasting.

EFE – Labour Market Analysis in Morocco & Algeria

- Designed and deployed data pipelines to monitor employment trends using AWS Sagemaker AI & Notebooks and Tableau.

Confindustria Lombardia – Web Data Pipeline for Labour Market Intelligence

- Implemented a web scraping framework using Selenium, Selenium Grid, and Storm Crawler to extract company job data.

Bertelsmann Institute – German Labour Market Analysis

- Applied custom ML to extract insights from job postings.
- Developed API integrations to connect various datasets into a single data ecosystem.

José Neves Foundation (Portugal) – Labour Market Intelligence

- Enhanced cross-sector analysis by integrating ISCO, ESCO, and NACE classifications.

CEDEFOP (European Centre for the Development of Vocational Training) and EUROSTAT: 2021-2024 “Towards the European Web Intelligence Hub - European system for collection and analysis of online job advertisement data (WIH-OJA)” 2021

- Designed and implemented big data processing pipelines for analysing online job vacancies (OJVs).

- Utilized NLP models (BERT, SBERT) and ontology matching (ESCO, ISCO, O*NET) to enhance skills classification.
- Employed PySpark, AWS EMR, and Apache AirFlow for scalable data processing

European Training Foundation (ETF) 2022 “BIG DATA FOR LABOUR MARKET INTELLIGENCE Online job vacancies (OJV) analysis system” Tender N°: 22ETF03 - ETF

- Designed and implemented big data processing pipelines for analysing online job vacancies (OJVs).
- Utilized NLP models (BERT, SBERT) and ontology matching (ESCO, ISCO, O*NET) to enhance skills classification.
- Employed PySpark, AWS EMR, and Apache AirFlow for scalable data processing.

OECD/OECS – Labour Market Intelligence (2022-2024)

- Integrated big data analytics and predictive modeling to assess global employment trends.

Key Skills & Technologies Used

- **Taxonomies & Ontologies:** ISCO, ESCO, NACE, O*NET
- **Big Data & Cloud Computing:** PySpark, AWS EMR, AWS Lambda, Beanstalk, AWS Batch
- **Machine Learning & AI:** NLP, BERT, SBERT, LLM fine-tuning, OpenAI API
- **Data Engineering & Pipelines:** Apache Airflow, API Development, SQL, MongoDB, Graph Databases
- **Web Scraping & Data Extraction:** Selenium, Selenium Grid, Storm Crawler
- **Data Visualization & BI Tools:** Tableau, Power BI
- **Search & Indexing:** Elasticsearch

2002 – 2016 Italy

BUSINESS ANALYST E BUSINESS INTELLIGENCE SPECIALIST SOFTWARE ARCHITECT GRUPPO CLAS

UNIONCAMERE – Labour Market Intelligence

- Developed **data pipelines and dashboards** for real-time **job vacancy and employment trend analysis**.

Excelsior – National Employment & Training Information System

- Designed **automated data collection systems** to extract and monitor job postings.

Regione Lombardia – Education & Labour Market Department

IFTS Project – Training & Skills Forecasting

Province of Reggio Calabria – Workforce Data Monitoring

- Created **big data analytics solutions** to **track employment trends** at a local level.

ARIFIL – Regione Lombardia – Skills & Training Analysis

- Designed and implemented a **real-time labour market intelligence system**, combining **data science, AI, and econometric models**.

Survey & Data Processing for Labour Market Information Systems

- Processed and **validated large-scale employment datasets**, ensuring **data integrity, quality, and cross-country comparability**.

2010 – 2016 Italy

HEAD OF BUSINESS INTELLIGENCE & DATA INTEGRATION - PROJECT MANAGER GRUPPO FRATELLI BERETTA S.P.A

2017 – CURRENT

LECTURER UNIVERSITY MILANO BICOCCA

Lecturer at "Business Intelligence and Big Data Analytics" for "Big Data processing & Architecture" course at University Milano Bicocca

2020 – CURRENT

LECTURER UNIVERSITY OF BERGAMO

Lecturer at "Master Degree Computer Engineering" and "Mobile & Cloud Technologies" courses

2005 – 2010

SENIOR CONSULTANT GN INFORMATICA

● **EDUCATION AND TRAINING**

2016 – 2017

MASTER I LEVEL IN BUSINESS INTELLIGENCE E BIG DATA ANALYTICS Università degli Studi di Milano Bicocca

Level in EQF EQF level 7

2005 – 2016 Italy

MASTER'S DEGREE IN COMPUTER ENGINEERING University of Bergamo

Level in EQF EQF level 7

2002 – 2005 Italy

BACHELOR'S DEGREE IN COMPUTER ENGINEERING University of Bergamo

Level in EQF EQF level 6

2016 – 2016

SAS CERTIFIED PREDICTIVE MODELER USING SAS ENTERPRISE MINER 13 SAS

2017 – 2017 Italy

ISIPM-BASE (AGILE PROJECT MANAGEMENT) Istituto Italiano di Project Management

2018 – 2018

MACHINE LEARNING BY STANFORD UNIVERSITY ON COURSERA Coursera

2021 – 2021

4TH INTERNATIONAL SCHOOL ON DEEP LEARNING DeepLearn 2021 Summer School

● **LANGUAGE SKILLS**

Mother tongue(s): **ITALIAN**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C2	C2	C2	C2	C1
FRENCH	A2	A1	A1	A1	A1

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user

● **SKILLS**

Git | SQL | Docker | Java | Azure and AWS | Python | MySQL | Linux | JavaScript | PySpark Python | Microservices | MongoDB | 📺 Machine Learning | Data Scraping and Analysis | web scraping | Cloud-computing | R, R Studio, R Markdown | Artificial Intelligence | GPT

● **MANAGEMENT AND LEADERSHIP SKILLS**

Project Management and team lead

● **PUBLICATIONS**

Pelucchi M., Psaila G. and Toccu M. (2017). Building a Query Engine for a Corpus of Open Data. In Proceedings of the 13th International Conference on Web Information Systems and Technologies - Volume 1: WEBIST, ISBN 978-989-758-246-2, pages 126-136. DOI: 10.5220/0006308801260136

Pelucchi M., Psaila G. and Toccu M. (2017). The Challenge of using Map-reduce to Query Open Data. In Proceedings of Special Session in Knowledge Discovery meets Information Systems (KomIS) at DATA2017, ISBN , pages 0-0. DOI: 10.5220/0006487803310342

Nobani, N.; Pelucchi, M.; Perico, M.; Scrivanti, A. and Vaccarino, A. (2021). GRASP: Graph-based Mining of Scientific Papers. In Proceedings of the 10th International Conference on Data Science, Technology and Applications - DATA, ISBN 978-989-758-521-0, pages 176-183. DOI: 10.5220/0010518901760183

Pelucchi M., Psaila G. and Toccu M. (2018) - Enhanced Querying of Open Data Portals Pelucchi, Mauro (et al.) Web Information Systems and Technologies 13th International Conference, WEBIST 2017, Porto, Portugal, April 25–27, 2017, Revised Selected Papers - Editors: Majchrzak, T.A., Traverso, P., Krempels, K.-H., Monfort, V. (Eds.) ISBN 978-3-319-93527-0

Pelucchi M., Psaila G. and Toccu M. (2018) - Hadoop vs. Spark: Impact on Performance of the Hammer Query Engine for Open Data Corpora – Algorithms, vol. 11 ISBN: 1999-4893, DOI: 10.3390/a11120209

Marrara, Stefania, Mauro Pelucchi, and Giuseppe Psaila. "Blind Queries Applied to JSON Document Stores." Information 10.10 (2019): 291.

Silvia Dusi, Claudia Graziani, Mario Mezzanzanica, Mauro Pelucchi (2019) - Informal Employment: Can Big Data Improve Knowledge on This Phenomenon? - Assessing Informal Employment and Skills Needs: Approaches and Insights from Regional and Local Labour Market Monitoring – Hampp, Rainer - ISBN: 3957102596

New insights on skills from the application of economic complexity to job postings data in the UK - Duncan Brown, Mauro Pelucchi, Elena Magrini, Anna Clara Gatti and Simone Perego (Brighton - 2022)

Web Job Vacancies per l'analisi della transizione digitale nel Made in Italy. - Una prima applicazione di artificial intelligence ai fabbisogni di competenze digitali in Puglia (Osservatorio regionale dei sistemi di istruzione e formazione in Puglia)

Trainito, Roberto, Anna Clara Gatti, and Mauro Pelucchi. "Le macro-competenze come dimensione per la progettazione formativa." (2022).

Montalbano, Giuseppa, et al. "Conoscere le competenze richieste dal mercato del lavoro. Un approfondimento di ricerca sulle macro-competenze digitali, hard skills, soft skills." (2023).

Porcelli, Rita, Valentina Ferri, and Anna Clara Gatti. "La domanda di competenze nell'ambito della mobilità elettrica. Gli indici macro-competenze digitali, hard skills, soft skills." (2023).

Brown, Duncan, Elena Magrini, and Mauro Pelucchi. "Predicting Skill Shortages with Real-Time Data: Using Online Job Adverts to Predict UK and Italian Employer Perceptions." Shortages of Labour and Skills. Nomos Verlagsgesellschaft mbH & Co. KG, 2024.

Gvasalia, Sophie, Mauro Pelucchi, and Anna Clara Gatti. "GREEN AND DIGITAL SKILLS ANALYSIS IN 4AFRICAN COUNTRIES." (2024).

Pelucchi, Mauro. "PATHWAYS OF GREENING LABOUR MARKETS: OPPORTUNITIES AND CHALLENGES." (2023).

Gvasalia, Sophie, et al. "Decoding AI's Evolution Using Big Data: A Methodological Approach." (KDIR 2024)

Camassa, Stefania, et al. "Gli effetti dell'AI sulle competenze: come decifrarli con l'analisi degli annunci di lavoro." (2024).

● CONFERENCES AND SEMINARS

The Challenge of using Map-reduce to Query Open Data - Special Session in Knowledge Discovery meets Information Systems (KomIS) at DATA2017, Madrid 24 Luglio 2017

Expert for the thematic sessions “Big data analysis techniques and machine learning” @ European big data hackathon follow up: from competition to cooperation organised by Cedefop and Eurostat, in Thessaloniki on 18 and 19 September (<http://www.cedefop.europa.eu/en/events-and-projects/events/prototype-production-0>)

“Data pre-processing” and “Text processing and information extraction” @ “Real-time labour market information and skill requirements - Setting up the infrastructure for EU system” by Cedefop and ESS-Net (network of European Statistical Systems) 20-21 March 2018, Milano, Italy

“Data pre-processing” and “Text processing and information extraction” @ “Expert Workshop - Setting up the infrastructure for EU system” by Cedefop, 12-13 September 2018, University of Exeter, Exeter (UK)

Lecturer @ “How to navigate Cedefop’s OJV data using Big data test infrastructure?” by EUROSTAT, 11-12 June 2019, Thessaloniki (Greece)

“Key points on method and technical approach to data treatment” @ “Expert workshop: exploring the Cedefop EU system for OJV analysis to update ICEs on current stage and next steps of the project” by Cedefop, 13- June 2019, Thessaloniki (Greece)

OJA system set up for the Real-time Labour Market information on Skill Requirements by Eurostat-DIGIT-Cedefop 14 October 2019, Luxembourg

“Description of the functional architecture of the CEDEFOP system on online ads for job vacancies and skills analysis” @ “Meeting of the Task Force on Trusted Smart Statistics” by EUROSTAT and ESS-Net (network of European Statistical Systems) 15-16 October 2019, Luxembourg

“EUROSTAT Advanced Training on Big Data Architecture and Project” by EUROSTAT and CEDEFOP April/May 2021 (Remote)

“ETF project Big Data for Labour Market Intelligence (LMI)” by ETF June 2021 (Remote)

“Big Data and Big Data Ingestion” – EUROPYTHON 2021 Interactive Session

Mauro Pelucchi, Ettore Colombo “ESTP Training Program - OJAs web scraping” by EUROSTAT April 2022 (Remote)

Mauro Pelucchi, Ettore Colombo “ESTP Training Program - OJAs web scraping” by EUROSTAT October 2022 (Koln - DE)

Mauro Pelucchi, Ettore Colombo “ESTP Training Program - OJAs web scraping” by EUROSTAT October 2023 (Koln - DE)

Mauro Pelucchi, Ettore Colombo “ESTP Training Program - OJAs web scraping” by EUROSTAT October 2024 (Koln - DE)

Mauro Pelucchi - Egypt - 2nd National Forum on the Future of Work - 7 March 2023

Ettore Colombo, Simone Perego, Paolo Invernizzi, Mauro Pelucchi - Web Intelligence Hub Conference - EUROSTAT - 12 June 2023

Elena Magrini, Mauro Pelucchi - 2023 joint Cedefop-OECD symposium - 15/16 June 2023 Thessaloniky

Anna Gatti, Simone Perego, Mauro Pelucchi - XLIV Conferenza scientifica annuale AISRe "Europa e Mediterraneo tra transizioni e conflitti. Opportunità e rischi per regioni e territori" - Napoli - 6-8 September 2023

Mauro Pelucchi - UNESCO Conference on Education Data and Statistics - 6-9 February 2024

Mauro Pelucchi - Skills Intelligence @ CEPS Brussels - 14 February 2024

Mauro Pelucchi - IA and future of work - Science PO Toulouse - 29 March 2024

Mauro Pelucchi - World Employment Confederation Conference - 23-25 April 2024

Mauro Pelucchi - A workforce transformed: jobs and skills in the age of AI - Brussels 26 June 2024

Mauro Pelucchi - Expert workshop on 'The Employment Dimension of the Green Economy: Measuring Green Jobs in the EU' (JRC - DG Employment) - 18 September 2024

Mauro Pelucchi - Human-centred digital transitions: skill mismatches in European workplaces - 11/12 December 2024
