



Mariia Levchenko

Email address: marylevchenko@gmail.com | LinkedIn: [maria-levchenko-39953284](https://www.linkedin.com/in/maria-levchenko-39953284) | GitHub: [mary-lev](https://github.com/mary-lev) | Kaggle: [marialevchenko](https://www.kaggle.com/marialevchenko)

SKILLS

Large Language Models | Machine Learning, NLP | Object-Oriented Programming | Python | Microservices | XML technologies: XML, XSLT, XPath, XQuery, TEI | Linked Data (RDF, SPARQL, Triple Stores) | Git | Linux | Kubernetes and Docker | Relational databases | Web Development | React, Node.js | Django | OpenAPI (Swagger) | API Integrations | TEI / XML | Deep Learning frameworks: PyTorch, Tensorflow, Keras | OCR Solutions | n8n

PROJECTS

01/05/2023 – CURRENT

Leggo Manzoni Project – Research & Development | Bologna University

- Developed a TEI-based digital edition integrating commentaries and multilingual translations.
- Automated pipelines for aligning 40 critical commentaries and translations in 10+ languages.
- Built a Node.js web application to enhance user interaction and present the project online, providing smooth navigation between text, comments and translations.

Link <https://projects.dharc.unibo.it/leggomanzoni/>

10/07/2022 – 31/10/2023

Stand for Ukraine – Volunteer Developer

- Developed and maintained an asynchronous Telegram chat bot using Python, asyncio/aiogram, and PostgreSQL to support fundraising initiatives for Ukrainian refugees.

01/06/2025 – CURRENT

DiScEPT: Digital Scholarly Editions and Aligned Translations Platform / Istituto Italiano di Studi Germanici (Roma, Italy)

I am building the universal backend application for automatic translation alignment with the TEI converter.

WORK EXPERIENCE

 **WELLTORY INC. (MOBILE WELLNESS APP)** – REDWOOD CITY, UNITED STATES

Website <https://welltory.com/>

SOFTWARE ENGINEER – 19/12/2020 – 10/01/2025

Welltory operates as a health and fitness application for smartphones, serving 8 million users with the ability to integrate data from multiple apps and devices to provide valuable health and lifestyle insights.

- Contributed to the development of a microservices platform for personalized message generation.
- Enhanced the message generation pipeline to improve performance and speed, positively impacting the user experience and increasing company revenue.
- Developed key performance metrics to facilitate consistent monitoring of system performance.
- Build and maintained AI-based automation tools.

 **EPAM SYSTEMS** – MOSCOW, RUSSIA

Website <https://www.epam.com/>

SOFTWARE ENGINEER – 09/09/2019 – 09/12/2020

Company: A Leading American Custom Software Manufacturer

- Developed an OCR pipeline using PaddleOCR with fine-tuned models and heuristics to extract structured data from medical record tables.
- Managed end-to-end machine learning pipelines, including data preprocessing, model training, and evaluation for document structuring tasks.

PYTHON DEVELOPER – 01/05/2019 – 01/09/2019

- Sole software engineer for a no-code platform used in online UX/UI design and content creation courses, handling both development and maintenance.
- Designed and implemented an innovative automation system for collecting and processing user feedback, improving the efficiency of feedback analysis.
- Created a comprehensive system of 4 indexes to accurately grade student assignments, improving the grading process and providing valuable insights for course improvement.

● **PUBLICATIONS**

2025

[TEI Encoding as a Unified Structure for Multilingual Digital Editions: The LeggoManzoni Case Study \(with Beatrice Nava, Ersilia Russo\)](#)

AIUCD-2025: The 14th Annual Conference of the Association of Digital Humanities and Digital Culture. 11-13th June, Verona, Italy.

2025

[Computational Analysis of Literary Communities: Event-Based Social Network Study of St. Petersburg 1999-2019](#)

Journal of Computational Literary Studies.

2025

[Evaluating Named Entity Recognition Models for Russian Cultural News Texts: From BERT to LLM](#)

2025

[Historical Russian TrOCR Model for Civil Script \(ru-trocr-1700s\)](#)

TrOCR model with CER 1.69%

2024

[Automatic Translation Alignment Pipeline for Multilingual Digital Editions of Literary Works](#)

Proceedings of the Computational Humanities Research Conference 2024. Aarhus, Denmark, 2024.

2024

[Mapping Literary Space: A Social Network from the Timeline of Cultural Events](#)

DH2024. Book of Abstracts. Pp. 423-424.

2024

[Commentare – Leggo Manzoni. Quaranta commenti alla Quarantana](#)

Manzoni e Leopardi in digitale. Idee e proposte per la scuola. Bologna, Clueb, 2024.

2024

[Literary Events in Saint Petersburg \(1999-2019\) from SPbLitGuide Newsletters](#)

Dataset

● **CONFERENCES AND SEMINARS**

13/02/2025 – 14/02/2025 Università della Svizzera italiana. Mendrisio, Switzerland

[Graphs & Networks Conference 2025](#)

04/12/2024 – 06/12/2024 Aarhus University, Denmark.
Computational Humanities Research 2024

07/10/2024 – 09/10/2024 Institute of Russian Literature. Saint Petersburg, Russia.
Graphomania. Russian Literature and Its Boundaries / Графомания. Русская литература и ее границы

26/09/2024 – 27/09/2024 Istituto Italiano di Studi Germanici. Roma, Italy.
Aligned Translations for Digital Scholarly Editions: Methodologies, Methods, and Workflows.

06/08/2024 – 09/08/2024 George Mason University, Washington, DC, USA.
ADHO Digital Humanities Conference 2024: Reinvention & Responsibility.

● **FELLOWSHIPS**

07/2024 – 08/2024
FAIR Data Fellowships NFDI4Memory / Herder Institute for Historical Research on East Central Europe

01/08/2024 – 31/01/2025
Sylla Internship

● **CREATIVE WORKS**

Как живут Digital Humanities в Китае: рассказ очевидца

Link <https://sysblok.ru/philology/kak-zhivut-digital-humanities-v-kitae-rasskaz-ochevidca/>

Как можно улучшить ответы языковых моделей? Гайд по промптам

Link <https://sysblok.ru/knowhow/kak-mozhno-uluchshit-otvety-jazykovyh-modelej-gajd-po-promtam/>

● **RESEARCH INTERESTS**

Research Interests

- Application of Large Language Models (LLMs) to humanities research
- Digital Scholarly Editions
- Historical document processing and OCR/HTR systems
- Open science approaches in computational humanities
- Digital methods in literary studies
- Development of automated research tools for manuscript analysis

Research Interests

● **EDUCATION AND TRAINING**

18/09/2022 – 28/03/2025 Bologna, Italy
MA IN DIGITAL HUMANITIES Bologna University

Website <https://www.unibo.it> | Final grade 110 cum laude |

Thesis LLM-based Scaffolded Learning Framework: A Case Study In Digital Humanities Programming Education

01/09/1996 – 30/06/1998 Saint Petersburg, Russia
MA IN EDUCATION Herzen State Pedagogical University

Website <https://www.herzen.spb.ru/>

● **LANGUAGE SKILLS**

Mother tongue(s): **RUSSIAN**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C1	C2	B2	B2	C2
ITALIAN	B1	B1	A2	A2	A2

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