

Rezumatul activității și a rezultatelor obținute în proiect în anul 2025

Cifra proiectului **25.80013.5007.57ROMD**

Denumirea Proiectului **Automatic detection of AI-generated/deep-fake texts from Moldova and Romania (ADAMo)**

Rezumat în limba română pentru anul 2025

Scopul principal a primei etape a proiectului este îmbogățirea corpusului CoRoLa cu texte originare din Republica Moldova prin încorporarea textelor dintr-o gamă diversă de surse din Republica Moldova.

Pentru primele patru luni a proiectului în anul 2025 este planificat de a colecta texte în volum de cel puțin 3 milioane de tokeni (cuvinte) și de a semna cel puțin un acord de colaborare care va permite utilizarea legitimată a textelor colectate în corpusul CoRoLa în conformitate cu legile în vigoare.

Acțiunile întreprinse pentru a atinge scopul dat:

A fost explorat spațiul web moldovenesc și depistate cele mai populare ziare și portale de noutăți online. Au fost contactate editurile Universității Tehnice a Moldovei, Universității de Stat a Moldovei, Universitatea de Stat Medicină și Farmacie „Nicolae Testemițanu” și altele cu scopul negocierilor. A fost explorat site <https://www.legis.md/> ca sursa de documente legislative care nu au protecția drepturilor de autor. A fost explorat site YouTube ca sursa de texte vorbite de cetățenii Moldovei.

Au fost create scripturi în limbajul de programare Python utilizând biblioteca BeautifulSoup (<https://beautiful-soup-4.readthedocs.io/en/latest/>) pentru unele site-uri cu noutăți precum ca SP (<https://esp.md/ro>), Diez (<https://diez.md/>), Ziarul de Gardă (<https://www.zdg.md/>), NewsMaker (<https://newsmaker.md/>) și altele. La fel, fiecare script conține și o parte care extrage și formează metadatele ce corespund textelor extrase.

A fost organizată o serie de instruiți a echipei din Moldova din partea colegilor din România în privința creării metadatelor. Activitatea aceasta vizează crearea metadatelor identice celor din CoRoLa pentru textele din Moldova ce vor fi adăugate în corpusul proiectului.

Modelele acordurilor furnizate de colegii din România au fost editate și aprobate de jurist universității noastre Corina Mardari. Se duc tratativele despre semnarea acordului. Au fost editate și pregătite două tipuri de acorduri: pentru edituri și pentru ziare online. În prezent, sunt semnate acorduri cu o editură (Totex-Lux) și un ziar (DIEZ.md). Drept urmare, editura Totex-Lux a furnizat cărți în format doc în volum total de 905.483 de cuvinte, iar de pe site-ul ziarului esp.md au fost extrase textele în volum de 4.252.341 de cuvinte. Toate aceste texte au metadate bine formate, în conformitate cu regulile corpusului CoRoLa.

Din cauza reînnoirii complete a site-ului facultății CIM, site-ul grupului nostru de cercetare a fost anulat și urmează să fie rescris și uploadat din nou. Pe parcurs, a fost selectată platforma web adecvată pentru crearea versiunii noi a site-ului – Wordpress; a fost finalizat și designul site-ului proiectului ADAMo. Site-ul este accesibil prin link [ADAMo](#).

Rezumat în limba engleză pentru anul 2025 1 pagină

The main goal of the first stage of the project is to enrich the CoRoLa corpus with texts originating from the Republic of Moldova by incorporating texts from a diverse range of sources from the Republic of Moldova.

For the first four months of the project in 2025, it is planned to collect texts in a volume of at least 3 million tokens (words) and to sign at least one collaboration agreement that will allow the legitimate use of the texts collected in the CoRoLa corpus in accordance with the laws in force.

Actions taken to achieve this goal:

The Moldovan web space was explored and the most popular newspapers and online news portals were identified. The publishing houses of the Technical University of Moldova, the State University of Moldova, the State University of Medicine and Pharmacy "Nicolae Testemițanu" and others were contacted for the purpose of negotiations. The website <https://www.legis.md/> was explored as a source of legislative documents that do not have copyright protection. The YouTube website was explored as a source of texts spoken by Moldovan citizens.

Scripts were created in the Python programming language using the BeautifulSoup library (<https://beautiful-soup-4.readthedocs.io/en/latest/>) for some news sites such as SP (<https://esp.md/ro>), Diez (<https://diez.md/>), Ziarul de Garda (<https://www.zdg.md/>), NewsMaker (<https://newsmaker.md/>) and others. Each script also contains a part that extracts and forms the metadata corresponding to the extracted texts.

Our team members contacted the publishing houses in Moldova with information about our project, its goals and proposals to conclude collaboration agreements. The agreement models provided by our Romanian colleagues were reviewed and approved by our university's lawyer Corina Mardari. Negotiations are underway to sign the agreement.

A series of trainings for the Moldovan team from our Romanian colleagues on metadata creation were organized online. This activity aims to create metadata identical to those in CoRoLa for Moldovan texts that will be added to the corpus within the project.

Two types of agreements were edited and prepared: for publishing houses and for online newspapers. At the moment, agreements are signed with a publishing house (Totex-Lux) and a newspaper (DIEZ.md). As a result, the publishing house Totex-Lux provided books in doc format in the total volume of 905,483 words, and from the newspaper esp.md website, the texts in the volume of 4,252,341 words have been extracted. All these texts have well-formed metadata in accordance with the rules of the CoRoLa corpus.

Due to the complete renewal of our faculty website, the website of our research group has been canceled and is to be rewritten and uploaded again. Along the way, the appropriate web platform for creating the new version of the website was selected – Wordpress; the design of the ADAMo project website was also finalized. The website is accessible via the [ADAMo](#) link.

Conducătorul de proiect Bobicev Victoria / (numele, prenumele, semnătura)

Data: 03.12.2025

LŞ