



FIȘA UNITĂȚII DE CURS/MODULULUI

1. Date despre program

1.1. Instituția de învățământ superior	Universitatea de Stat din Moldova
1.2. Facultatea	Matematică și Informatică
1.3. Departamentul	Informatică
1.4. Domeniul de formare profesională	0613 Dezvoltarea produselor program și a aplicațiilor
1.5. Ciclul de studii	Studii superioare de master, ciclul II
1.6. Programul de studii	Inteligența Artificială și Știința Datelor

2. Date despre unitatea de curs/modul

2.1. Denumirea disciplinei	Modele de limbaje de dimensiuni mari (LLM)						
2.2. Titularul activităților de curs	?						
2.3. Titularul activităților de laborator	?						
2.4. Anul de studii	2	2.5. Semestrul	1	2.6. Tipul de evaluare	E	2.7. Categoria formativă	SA

3. Timpul total estimat (ore pe semestru al activităților didactice)

3.1. Număr de ore pe săptămână	4	din care: 3.2 curs	2	3.3. laborator	2
3.4. Total ore în planul de învățământ	40	din care: 3.5 curs	20	3.6. laborator	20
Distribuția fondului de timp pentru studiul individual:					ore
Studiu după manual, suport de curs, bibliografie și notițe					40
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate					35
Pregătire laboratoare, teme, referate, portofolii					35
3.7. Total ore studiu individual	110				
3.8. Total ore pe semestru	150				
3.9. Număr de credite					

4. Precondiții de acces la unitatea de curs/modul

4.1. de curriculum	Programare, Învățarea automată, Rețele neuronale, Învățare profundă
4.2. de competențe	Cunoștințe de programare, algoritmi și structuri de date

5. Condiții de desfășurare

5.1. a cursului	Sală de curs cu tablă mare și proiector, laptop cu conexiune internet.
5.2. a laboratorului	Sală de laborator, conexiune internet, calculatoare care au instalate Python și TensorFlow.

6. Obiectivele disciplinei – rezultate așteptate ale învățării la formarea cărora contribuie parcurgerea și promovarea disciplinei

Cunoștințe	(C1) Înțelegerea fundamentelor LLM: Cunoștințe detaliate despre arhitectura, designul și funcționarea modelelor de limbaj mare (LLM), inclusiv mecanismele de bază ale transformatorilor, tokenizării, și modelelor de limbaj. (C2) Tipuri de LLM: Familiarizare cu diverse tipuri de LLM, cum ar fi BERT, GPT, T5, și modelele multimodale, inclusiv aplicațiile și limitările lor specifice. (C3) Evaluare și Benchmarking: Cunoștințe despre metodele și criteriile de evaluare a performanței LLM-urilor, precum și utilizarea benchmark-urilor standardizate.
------------	--



	<p>(C4) Tehnici de fine-tuning și optimizare: Înțelegerea proceselor de fine-tuning pentru personalizarea LLM pentru sarcini specifice și tehnici de optimizare a performanței.</p> <p>(C5) Aplicații practice ale LLM: Cunoștințe aplicate în domenii precum generarea de text, traducerea automată, aplicațiile de dialog (chatbots), și crearea de conținut multimodal.</p> <p>(C6) Etica și reglementarea LLM: Înțelegerea problemelor etice asociate cu utilizarea LLM și cunoștințe despre reglementările și politicile emergente în domeniu.</p>
Abilități	<p>(A1) Dezvoltarea și implementarea LLM: Capacitatea de a crea și implementa soluții bazate pe LLM, adaptând modelele la diverse aplicații industriale și comerciale.</p> <p>(A2) Evaluarea performanței LLM: Abilitatea de a aplica tehnici de benchmarking pentru a evalua și compara eficiența și precizia diferitelor LLM.</p> <p>(A3) Personalizarea și fine-tuning-ul LLM: Competența de a efectua fine-tuning pe modele existente pentru a le adapta la cerințe specifice de sarcini.</p> <p>(A4) Implementarea sistemelor de căutare și recuperare de informații: Abilitatea de a integra LLM în sisteme de căutare și RAG pentru îmbunătățirea recuperării și generării de conținut.</p> <p>(A5) Aplicații multi-modale: Capacitatea de a dezvolta aplicații LLM care integrează text, imagine, și alte forme de date.</p> <p>(A6) Utilizarea tehnicilor de prompt engineering: Abilitatea de a formula și optimiza prompturi pentru a obține rezultate dorite de la LLM.</p> <p>(A7) Abordarea problemelor etice: Abilitatea de a recunoaște și rezolva dilemele etice care pot apărea în dezvoltarea și utilizarea LLM.</p>
Responsabilitate și autonomie	<p>(R1) Responsabilitate în utilizarea LLM: Dezvoltarea unei conștiințe profesionale privind impactul utilizării LLM în diferite domenii, asumându-și responsabilitatea pentru deciziile etice și sociale.</p> <p>(R2) Autonomie în cercetare și dezvoltare: Capacitatea de a efectua cercetări independente și de a dezvolta noi aplicații bazate pe LLM, inovând în funcție de nevoile specifice ale proiectelor.</p> <p>(R3) Gândire critică și decizională: Abilitatea de a evalua critic utilizarea LLM, de a lua decizii informate și de a propune soluții la probleme complexe legate de LLM.</p> <p>(R4) Managementul proiectelor LLM: Capacitatea de a gestiona și coordona proiecte de mare amploare care implică dezvoltarea și implementarea LLM, asigurând respectarea termenelor și standardelor de calitate.</p> <p>(R5) Responsabilitate etică: Asumarea responsabilității pentru respectarea standardelor etice și legale în utilizarea LLM, inclusiv protejarea confidențialității și securității datelor.</p> <p>(R6) Capacitatea de a colabora în echipă: Dezvoltarea abilităților de a lucra eficient într-un mediu de echipă multidisciplinară, contribuind la succesul proiectelor bazate pe LLM.</p> <p>(R7) Învățare continuă și adaptabilitate: Capacitatea de a se adapta rapid la noile tendințe și tehnologii emergente în domeniul LLM, menținându-se la curent cu evoluțiile din domeniu.</p>



7. Conținutul unității de curs/modulului

7.1. Curs	Numărul de ore
<i>C1. Introducere în Modele de limbaje de dimensiuni mari (LLM).</i> Model de limbaje de dimensiuni mari (LLM) ca modele de Inteligență Artificială Generativă. Aplicarea rețelelor neuronale profunde în LLM. Principii de bază ale LLM. Exemple de aplicații LLM în diferite domenii. Impactul social și economic al LLM	1
<i>C2. Arhitectura și designul modelelor LLM.</i> Componentele cheie ale arhitecturii LLM. Definiția și importanța modelelor fundamentale. Metode de optimizare pentru LLM. Prezentarea arhitecturii BERT și GPT. Parametri de performanță pentru LLM. Studii de caz și aplicații pentru fiecare model. Avantaje și limitări în utilizarea LLM.	1
<i>C3. GPT - modele generative și modele multimodale.</i> Prezentarea arhitecturii GPT. Aplicații practice ale GPT. Evaluarea avantajelor și limitărilor GPT. Exemple de utilizare a datelor multimodale. Avantajele și provocările modelelor multimodale. Studii de caz în implementarea modelelor multimodale. Exemple de implementare a GPT în aplicații reale.	2
<i>C4. Date pentru modelele LLM.</i> Categoriile de date folosite în LLM. Baze de date vectoriale și funcționalitatea lor. Avantajele bazelor de date vectoriale. Tehnici de prelucrare a datelor. Metode de curățare a datelor. Evaluarea surselor de date. Problematika bias-ului și diversității în date.	1
<i>C5. Antrenarea și evaluarea modelelor LLM.</i> Algoritmi cheie pentru antrenarea LLM. Ajustarea hiperparametrilor. Probleme comune și soluții în antrenarea LLM. Metode de evaluare utilizate în LLM. Stabilirea standardelor de performanță. Evaluarea capacității de generalizare a LLM. Validarea și testarea în medii variate.	3
<i>C6. Prompt Engineering.</i> Noțiunea de Prompt Engineering. Clasificarea prompturilor și exemple practice. Metode de optimizare a prompturilor. Evaluarea răspunsurilor și ajustarea prompturilor. Exemple de aplicații de succes ale Prompt Engineering	2
<i>C7. Retrieval Augmented Generation (RAG).</i> Definierea și importanța RAG. Avantajele RAG pentru LLM. Studii de caz în implementarea RAG. Exemple de fluxuri de lucru îmbunătățite cu RAG.	1
<i>C8. Aplicații ale modelelor LLM.</i> Generarea de text cu LLM. Utilizarea LLM pentru rezumate și descrieri. Tehnici de traducere automată și localizare. Analiza sentimentelor și exemple de utilizare. Generarea de conținut pentru campanii de marketing. Sistemele de răspuns la întrebări și asistență virtuală. Personalizarea și optimizarea interacțiunilor. Integrarea în platforme și sisteme existente. Evaluarea performanței chatbot-urilor.	3
<i>C9. Aplicații specializate ale LLM.</i> Tehnici de generare de imagini. Exemple de aplicații în artă și design. Aplicații LLM în domeniul medical. Utilizarea LLM în educație și formare. LLM în științele umaniste și cercetare. Utilizarea LLM analiza financiară și planificare financiară. Optimizarea vânzărilor cu LLM. Marketing personalizat și gestionarea relațiilor cu clienții. Utilizarea LLM pentru îmbunătățirea serviciilor publice. Studii de caz.	3
<i>C10. Modele Open Source și Hugging Face.</i> Definiția și importanța modelelor open source. Prezentarea platformei Hugging Face. Avantajele utilizării modelelor open source. Exemple de implementare din Hugging Face. Studiu de caz asupra aplicațiilor open source.	2
<i>C10. Etica și reglementarea LLM.</i> Etica în utilizarea LLM. Reglementări și standarde pentru LLM. Provocări etice și soluții în dezvoltarea LLM. Studii de caz în aplicarea eticii și reglementării. Viitorul reglementărilor în domeniul LLM.	1
<i>Total prelegeri:</i>	20



Bibliografie:

1. Mohamed Karim (2024). *The Complete LLM Engineering Handbook*.
2. Et Tu Code (2024). *Building, Training and Hardware for LLM AI: A Comprehensive Guide to Large Language Model Development Training*.
3. Valentina Alto (2024). *Modern Generative AI with ChatGPT and OpenAI Models: Leverage the capabilities of OpenAI's LLM for productivity and innovation with GPT3 and GPT4*. Packt Publishing
4. Valentina Alto (2024). *Building LLM Apps: Create Intelligent Apps and Agents with Large Language Models*. Packt Publishing
5. O. Campesato (2024). *Transformer, BERT, and GPT: Including ChatGPT and Prompt Engineering*. Mercury Learning and Information
6. Pradeep Savadi Archit, Goyal, Virender Savadi. *LLM Essentials A Busy Professional's Guide to Large Language Models*.
7. Denis Rothman (2022). *Transformers for Natural Language Processing*. Packt Publishing
8. Jay Alammar, Maarten Grootendorst (2025). *Hands-On Large Language Models*. O'Reilly Media
9. Akshay Kulkarni, Adarsha Shivananda, Anoosh Kulkarni, Dilip Gudivada (2023). *Applied Generative AI for Beginners Practical Knowledge on Diffusion Models, ChatGPT, and Other LLMs*. Apress

7.2. Laborator	Numărul de ore
L1. Construirea unui model simplu LLM folosind BERT	4
L2. Generarea automată de rezumate folosind LLM	4
L3. Implementarea unui sistem de traducere automată folosind LLM	4
L4. Dezvoltarea unei aplicații de generare de imagini folosind modele multimodale	4
L5. Construirea unui chatbot cu abilități multimodale	4
<i>Total laborator:</i>	20

Bibliografie:

1. Valentina Alto (2024). *Building LLM Powered Applications*. Packt Publishing
2. Henry Habib (2024). *OpenAI API Cookbook*. Packt Publishing
3. Aymen El Amri (2023). *OpenAI GPT For Python Developers*. Leanpub
4. Raj Arun R (2024). *Mastering Large Language Models with Python*. Orange Education
5. Louis-François Bouchard, Louie Peters (2024). *Building LLMs for Production*. Towards AI
6. Sandra Kublik and Shubham Saboo (2023). *GPT-3: Building Innovative NLP Products Using Large Language Models*. O'Reilly Media
7. Steve Tingiris (2021). *Exploring GPT-3*. Packt Publishing

8. Evaluare

Nota semestrială			
Atestarea 1	Atestarea 2	Evaluare curentă	Lucrul individual
25%	25%	25%	25%

Nota finală	
Nota semestrială	Nota la examen
60%	40%

Standard minim de performanță
Standard minim (cunoștințe și aptitudini necesare pentru nota 5)
<ul style="list-style-type: none"> • prezența la cursuri și seminarii conform cerințelor generale ale facultății



Universitatea de Stat din Moldova



- cunoașterea arhitecturii LLM
- abilitatea de a implementa un model simplu LLM
- capacitatea de a implementa un model de traducere automată în baza LLM

Evaluarea în cadrul disciplinei „Modele de limbaje de dimensiuni mari (LLM)” se realizează formativ - activitatea la orele de laborator, verificarea orală sau scrisă curentă, două lucrări de atestare (la mijloc și la sfârșit de semestru), și sumativ (final) – 4 lucrări independente, realizate pe parcursul semestrului și examenul final la disciplină. Nota generală la disciplină însumează nota de la examen și cea semestrială, în proporție de 40 și, respectiv, 60 la sută, și apreciază gradul de corespundere cu finalitățile scontate: cunoștințele și competențele acumulate, abilitatea de a aplica cunoștințele, gradul de integrare a cunoștințelor de către studenți etc.

Data completării

25.08.2024

Titular de disciplină

Dr., conf. univ. ?

Data avizării în departament

Șef departament