



THE ISSUE CONTAINS:

Proceedings of the 8th
International Scientific
and Practical Conference

**CONCEPTS FOR THE DEVELOPMENT OF
SOCIETY'S SCIENTIFIC POTENTIAL**

Prague, Czech Republic
19-20.10.2025

SCIENTIFIC COLLECTION
INTERCONF+

No 62 [268]
October, 2025



Scientific Collection «InterConf+ »

No 62(268)

October 2025

THE ISSUE CONTAINS:

Proceedings of the 8th International
Scientific and Practical Conference

CONCEPTS FOR THE
DEVELOPMENT OF SOCIETY'S
SCIENTIFIC POTENTIAL

PRAGUE, CZECH REPUBLIC

October 19–20, 2025



PRAGUE
2025

UDC 001.1

- S 40** *Scientific Collection «InterConf+»*, 62(268): with the Proceedings of the 8th International Scientific and Practical Conference «Concepts for the Development of Society's Scientific Potential» (October 19-20, 2025; Prague, Czech Republic) / comp. by LLC SPC «InterConf». Prague: Author-publishers miscellaneous, 2025. 270 p.

ISSN 2709-4685

DOI 10.51582/interconf.19-20.10.2025

EDITOR

Anna Svoboda

Doctoral student
University of Economics;
Czech Republic
annasvobodaprague@yahoo.com

COORDINATOR

Mariia Granko

Coordination Director
LLC Scientific Publishing Center
«InterConf»; Ukraine
info@interconf.center

EDITORIAL BOARD

Dmytro Marchenko (PhD in Engineering)
Mykolayiv National Agrarian University
(MNAU); Ukraine;

Mariana Vereskliia (PhD in Pedagogy)
Lviv State University of Internal Affairs;
Ukraine

Dan Goltsman (Doctoral student)
Riga Stradiņš University;
Republic of Latvia;
goltsman.dan@inbox.lv

Katherine Richard (DSc in Law),
Hasselt University; Kingdom of Belgium
katherine.richard@protonmail.com;

Bashirov Ansar (Doctor of Medicine),
EMIH of Almaty region, Republic of Kazakhstan

Stanyslav Novak (DSc in Engineering)
University of Warsaw; Poland
novaks657@gmail.com;

Kanako Tanaka (PhD in Engineering),
Japan Science and Technology Agency; Japan;

Vagif Sultanly (DSc in Philology)
Baku State University; Republic of Azerbaijan

Davit Tchiotashvili (Doctor of Economics),
Gori State University, Georgia;

Richard Brouillet (LL.B.),
University of Ottawa; Canada;

Kamilə Əliağa qızı Əliyeva (DSc in Biology)
Baku State University; Republic of Azerbaijan

Giuli Giguashvili (Doctor of Economics),
Gori State University, Georgia;

Tamar Makasarashvili (Doctor of Economics),
Gori State University, Georgia;

Khaliana Chitadze (Doctor of Economics),
Gori State University, Georgia;

Svitlana Lykholat (PhD in Economics),
Lviv Polytechnic National University; Ukraine

Viktor Yanchenko (PhD in Pharm. Sc.),
T.H. Shevchenko National University
«Chernihiv Colehium»; Ukraine

Rakhmonov Aziz Bositovich (PhD in Pedagogy)
Uzbek State University of World Languages;
Republic of Uzbekistan;

Asta Marija Inkėnienė (Doctor of Pharm. Sc.),
Lithuanian University of Health Sciences,
Republic of Lithuania;

Vera Gorak (PhD in Economics)
Karlovarská Krajská Nemocnice; Czech Republic
veragorak.assist@gmail.com;

Polina Vuitsik (PhD in Economics)
Jagiellonian University; Poland
p.vuitsik.prof@gmail.com;

Alexander Schieler (PhD in Sociology),
Transilvania University of Brasov; Romania
alexanrds.schieler@protonmail.ch

George McGrown (PhD in Finance)
University of Florida; USA
mcgown.geor@gmail.com;

Mark Alexandr Wagner (DSc. in Psychology)
University of Vienna; Austria
mw6002832@gmail.com;

Larysa Kupriianova (PhD in Medicine)
Humanitas University, Italy

Temur Narbaev (DSc in Medicine)
Tashkent Pediatric Medical Institute,
Republic of Uzbekistan;
temur1972@inbox.ru

Nataliia Mykhalitska (PhD
in Public Administration)
Lviv State University of
Internal Affairs; Ukraine

Please, cite as shown below:

1. Surname, N. & Surname, N. (2025). Title of an article. *Scientific Collection «InterConf+»*, 62(268), 21-27. <https://doi.org/10.51582/interconf...>

This issue of Scientific Collection «InterConf+» contains the materials of the International Scientific and Practical Conference. The conference provides an interdisciplinary forum for researchers, practitioners and scholars to present and discuss the most recent innovations and developments in modern science. The aim of conference is to enable academics, researchers, practitioners and college students to publish their research findings, ideas, developments, and innovations.

Scientific Collection «InterConf+» and its content are indexed in:

Index Copernicus; Google Scholar; WorldCat; OUCI (Open Ukrainian Citation Index); CrossRef;
Semantic Scholar; Mendeley; Scilit; OpenAIRE (pan-European research information system), etc.

© 2025 Authors

© 2025 Author-publishers miscellaneous

© 2025 LLC SPC «InterConf»

TABLE OF CONTENTS

MANAGEMENT

	Kurbaniyazov S.K.	ANALYSIS OF THE ACTIVITIES OF CORPORATE STRUCTURES IN THE DEVELOPMENT OF THE GLOBAL ECONOMY	6
	Pereshliuga K.	AI-DRIVEN CLIENT MANAGEMENT IN PERSONAL SERVICES: CASE STUDY FROM THE BEAUTY INDUSTRY	11
	Zharkov S.	DEVELOPMENT OF A UNIFIED PROJECT MANAGEMENT FRAMEWORK FOR CONSTRUCTION COMPANIES	21

PEDAGOGY AND EDUCATION

	Mammadova Z.N. Mammadov F.M.	WAYS OF FORMING ENVIRONMENTAL EDUCATION IN TEACHING THE TOPIC «CARBON AND ITS COMPOUNDS»	28
	Manapbayeva Z. Yelyukenova M. Alpysbayeva A.Y.	TEACHING ENGLISH AS A SECOND LANGUAGE TO SCHOOL-AGED STUDENTS WITH AUTISM SPECTRUM DISORDER (ASD)	35
	Orujova S.	THE RELEVANCE OF PREVENTING HARMFUL HABITS	53
	Айкадамова А.Ж. Абдрахманова Р.Б.	НАУЧНО-МЕТОДИЧЕСКИЕ ПОДХОДЫ К ОЦЕНКЕ ЭФФЕКТИВНОСТИ ДЕЯТЕЛЬНОСТИ ПЕДАГОГА-ПСИХОЛОГА В РАБОТЕ С ТРУДНЫМИ ПОДРОСТКАМИ	61
	Карапетян С.Г. Киракосян А.А.	РОЛЬ СЕМЬИ В ОРГАНИЗАЦИИ РАННЕГО ВМЕШАТЕЛЬСТВА ПРИ НАРУШЕНИЯХ РЕЧИ	67
	Кошель В.М. Кисла А.А.	ДО ПРОБЛЕМИ МЕТОДИЧНОГО СУПРОВОДУ ЗБЕРЕЖЕННЯ ЗДОРОВ'Я ДІТЕЙ В ЗАКЛАДІ ДОШКІЛЬНОЇ ОСВІТИ	72

PHILOSOPHY AND COGNITION

	Ибрагимова М.Э.	ПОНЯТИЕ «РАЦИОНАЛЬНОСТЬ» В КОНТЕКСТЕ МОДЕРНИСТСКОЙ ПАРАДИГМЫ СОЦИОКУЛЬТУРНОЙ САМОИДЕНТИФИКАЦИИ (КОММУНИКАТИВНЫХ СИСТЕМ)	86
---	-----------------	---	----

SOCIOLOGY AND SOCIETY

	Pliushch V.	SOCIAL INTEGRATION STRATEGIES OF UKRAINIAN SOCIETY DURING THE WARTIME	96
---	-------------	---	----

GEOGRAPHY AND LOCAL HISTORY

	Hummetova E.B.	APPLICATION OF GIS TECHNOLOGIES IN URBAN PLANNING AND SUSTAINABLE DEVELOPMENT	108
---	----------------	---	-----

HISTORY AND ARCHEOLOGY, ARCHIVAL STUDIES

	Олешко Н.П. Шевченко М.П.	УКРАЇНСЬКА КУЛЬТУРА НА МІЖНАРОДНІЙ АРЕНІ ЗА ЧАСИ ПОВНОМАСШТАБНОЇ РОСІЙСЬКО-УКРАЇНСЬКОЇ ВІЙНИ (2022– 2025 РР.)	114
---	------------------------------	--	-----

MEDICINE AND PHARMACY

	Baigenzheeva R.K. Ganikyzy A. Yeszhanova Z.A.	IMPACT OF LIFESTYLE ON STROKE RISK AMONG YOUNG PEOPLE IN KAZAKHSTAN	123
	Salivon H.	PATHOLOGICAL MECHANISMS OF GESTATIONAL DIABETES MELLITUS DEVELOPMENT	130
	Shagazatova B.X. Vafojev S.F.	QANDLI DIABET 2-TURI BILAN KASALLANGAN BEMORLARDA METFORMINNING ICHAK MIKROBIOTASI HOLATIGA TA'SIRINI BAHOLASH	143

ENERGETICS

	Білюк І.С. Савченко О.В. Масленников Д.В. Данилко С.А. Бессоновський Ю.С. Пономаренко І.В.	АВТОМАТИЗОВАНИЙ ЕЛЕКТРОПРИВОД СИСТЕМИ ОЧИЩЕННЯ ТРАНСФОРМАТОРНОГО МАСТИЛА	146
---	---	---	-----

AGROTECHNOLOGIES AND AGRICULTURAL INDUSTRY

	Dovbysh L. Kravchuk M. Skolub A. Sokhatska A. Pleshivenko O.	THE INFLUENCE OF COMPLEX FERTILIZER NORMS ON GROWTH DYNAMICS AND BIOMASS ACCUMULATION OF MAIZE UNDER PRECISION FARMING IN THE CONDITIONS OF THE RIGHT-BANK FOREST-STEPPE OF UKRAINE	155
	Hrytsun A. Bilous V.	OPTIMIZATION OF AUTONOMOUS ENERGY SUPPLY FOR LIVESTOCK ENTERPRISES THROUGH THE USE OF BIOGAS TECHNOLOGIES	168

RADIO ENGINEERING, ELECTRONICS AND ELECTRICAL ENGINEERING

	Barylo H. Kobyk Y. Pavlenko M. Hrynchyshyn A. Shliusar Y.	MODELING AND IMPLEMENTATION OF AN ANALOG FRONT-END FOR PIEZOELECTRIC ULTRASONIC SENSORS WITH BUILT-IN SELF-DIAGNOSTICS	177
	Barylo H. Shkribynets V. Savytskyi A. Soroka V. Nykon O.	DESIGN AND IMPLEMENTATION OF AN OPTICAL OXYGEN SENSOR BASED ON THERMALLY ACTIVATED DELAYED FLUORESCENCE ORGANIC MATERIALS	183

INFORMATION AND WEB TECHNOLOGIES

	Baigenzheeva R.K. Kendirbek A. Kabduakhit A.	HOW ARTIFICIAL INTELLIGENCE CAN HELP IN COMBATING INFECTIOUS DISEASES	190
	Nuriyeva V.A. Şükürova L.N. Nəsirova Ş.S. Əhmədova N.K.	RƏQƏMSALLAŞDIRILMIŞ CƏMIYYƏT: «AĞILLI EV» KONTEKSTINDƏ	206
	Мінухін С.В. Рудой В.В.	ДОСЛІДЖЕННЯ АРХІТЕКТУР НЕЙРОННИХ МЕРЕЖ ДЛЯ ОБРОБКИ МУЛЬТИМОДАЛЬНИХ ДАНИХ	214

PHYSICAL EDUCATION AND SPORTS

	Базилевич Н.О. Безкоровайний С.Б. Середній Г.М. Деак З.-Д.З.	ТЕОРЕТИКО-МЕТОДИЧНІ АСПЕКТИ БАЗОВОЇ ТЕХНІЧНОЇ ПІДГОТОВКИ БОРЦІВ	223
---	---	--	-----

MILITARY AFFAIRS AND NATIONAL SECURITY

	Filiunkin Y.V. Skorodid S.P. Kutsenko V.V.	SPECIFIC FEATURES OF BUILDING AND CONDUCTING DEFENSE BY RUSSIAN FEDERATION UNITS AND SUBUNITS DURING THE WAR IN UKRAINIAN	241
	Камалтинов Г.Г. Тютюнник В.О. Колеснік О.М. Нелінь Д.П.	РАДІОЛОКАЦІЙНІ ЗАСОБИ ВИЯВЛЕННЯ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ, КЛАСИФІКАЦІЯ, НАПРЯМКИ РОЗВИТКУ, ТЕХНОЛОГІЇ ПОБУДОВИ	251

INFORMATION AND WEB TECHNOLOGIES

 DOI 10.51582/interconf.19-20.10.2025.024

Дослідження архітектур нейронних мереж для обробки мультимодальних даних

**Мінухін Сергій Володимирович¹,
Рудой Валерій Валерійович²**

¹ доктор технічних наук, професор, професор кафедри системотехніки;
Харківський національний університет радіоелектроніки; Україна

² аспірант;
Харківський національний університет радіоелектроніки; Україна

Анотація.

У роботі проведено дослідження сучасних архітектур нейронних мереж, призначених для оброблення мультимодальних даних, що включають текстові та візуальні компоненти. Розглянуто основні підходи до інтеграції модальностей, зокрема раннє, пізнє та гібридне злиття, а також контрастивне навчання та крос-модальну увагу. Виконано порівняльний аналіз моделей CLIP, ViLT, LXMERT та Perceiver IO за критеріями ефективністю, швидкістю навчання, масштабованістю та інтерпретованістю результатів. Результати дослідження демонструють переваги гібридних та контрастивних підходів у задачах мультимодальної класифікації та прогнозування, а також перспективи застосування таких моделей у медичній діагностиці, автономних системах та інтелектуальних системах рекомендацій.

Ключові слова:

мультимодальні дані
нейронні мережі
архітектури глибинного навчання
контрастивне навчання
крос-модальна увага
оптимізація архітектури

INFORMATION AND WEB TECHNOLOGIES

Вступ. Розвиток штучного інтелекту в останнє десятиліття ознаменувався значним прогресом у напрямі побудови нейронних мереж, здатних працювати з різномірними потоками даних. Сучасне інформаційне середовище характеризується мультисенсорністю, тобто поєднує тексти, зображення, аудіо, відео, сигнали з фізичних датчиків, медичні зображення та параметри тощо [1, 2]. Такі дані, що мають різну природу, називають мультимодальними. Їх ефективна обробка вимагає моделей, які здатні узгоджувати семантику між різними модальностями, забезпечуючи спільне розуміння контексту.

Традиційні алгоритми машинного навчання, орієнтовані на оброблення однорідних наборів ознак, виявляються недостатньо ефективними для розв'язання подібних завдань. Вони не можуть одночасно враховувати структурні відмінності між модальностями, зокрема просторову організацію зображень та послідовну природу тексту. Саме тому в останні роки активно розвиваються архітектури нейронних мереж, орієнтовані на мультимодальне навчання. Їхня ключова ідея полягає у створенні спільного латентного простору, у якому різні види даних представлені у єдиній семантичній формі.

Актуальність теми дослідження визначається потребою у створенні систем, здатних комплексно сприймати навколишній світ: розпізнавати сцени, аналізувати контексти, інтерпретувати мову та зображення одночасно. Такі системи використовуються у сфері комп'ютерного зору, медичної діагностики, автономного транспорту, інтелектуальних рекомендаційних платформ, а також у мультимодальних чат-ботах. У зв'язку з цим аналіз архітектур, що забезпечують ефективне узгодження модальностей, має важливе як теоретичне, так і прикладне значення.

Підходи до оброблення мультимодальних даних. У сучасній науковій літературі виділяють три базові стратегії об'єднання мультимодальних ознак: раннє, пізнє та гібридне злиття.

Раннє злиття (Early Fusion) полягає у конкатенації або статистичному узгодженні вхідних ознак до подачі в нейронну мережу [2]. Його перевага – спрощена реалізація та можливість одночасного навчання однієї моделі. Однак цей підхід може призводити до втрати специфічних характеристик модальності, оскільки дані різної природи зливаються без урахування їхніх особливостей.

Пізнє злиття (Late Fusion) передбачає незалежну обробку

INFORMATION AND WEB TECHNOLOGIES

кожної модальності окремою моделлю, результати якої потім поєднуються на рівні класифікатора або ансамблю [3]. Такий підхід забезпечує більшу гнучкість і дозволяє оптимізувати параметри під конкретну модальність, але ускладнює побудову спільного латентного простору.

Гібридне (проміжне) злиття (Intermediate Fusion) об’єднує переваги зазначених підходів, поєднуючи модальності вже на рівні латентних уявлень [4]. Це дозволяє зберегти специфіку кожного типу даних і водночас створити єдине узгоджене представлення. Гібридні архітектури покладено в основу більшості сучасних мультимодальних моделей, зокрема CLIP, LXMERT та ViLT.

Сучасні архітектури мультимодальних нейронних мереж. Серед значної кількості архітектур, розроблених у сфері мультимодального навчання, виділяють наступні ключові моделі, які стали базовими орієнтирами для напрямів подальших досліджень.

Модель CLIP поєднує текстовий і візуальний енкодери, що навчаються контрастивно. Візуальний енкодер реалізований на базі ResNet або Vision Transformer, а текстовий – на базі трансформера [5]. Обидва енкодери перетворюють свої модальності у вектори, після чого система навчається таким чином, щоб пари «зображення-опис» мали максимальну подібність у спільному просторі. Це дозволяє моделі розпізнавати об’єкти на зображеннях за текстовим запитом без додаткового навчання (рис. 1).

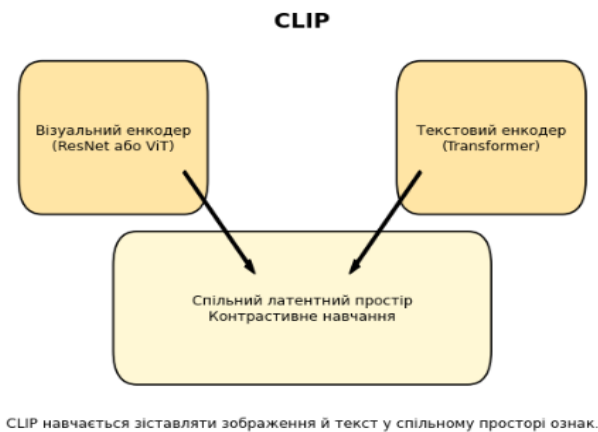


Рисунок 1
Модель CLIP

INFORMATION AND WEB TECHNOLOGIES

ViLT об'єднує зображення і текст у межах одного трансформера. Зображення попередньо розбивається на патчі, що трактуються як токени, і подається на вхід разом із токенами тексту [6]. Така інтеграція забезпечує суттєве зменшення кількості параметрів та високу швидкість навчання. Модель ефективна у завданнях мультимодальної класифікації та пошуку семантичних відповідностей (рис. 2).

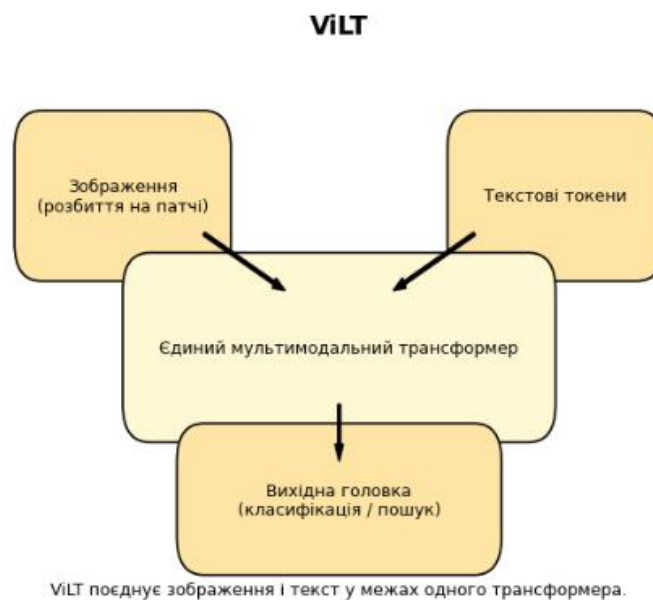


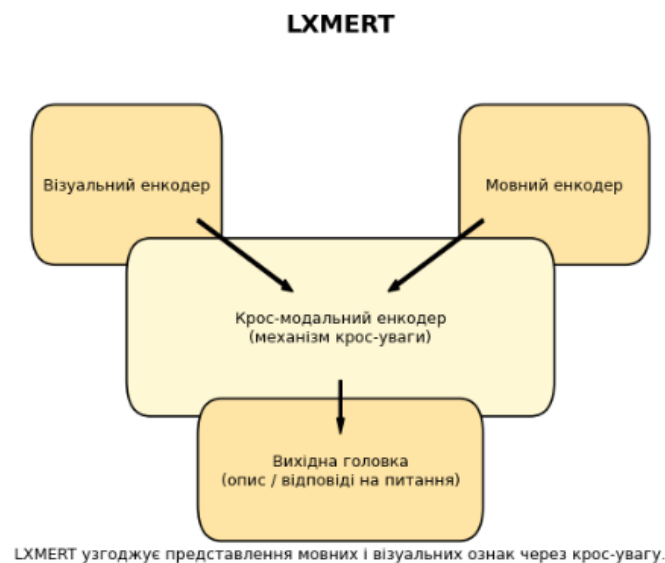
Рисунок 2
Модель ViLT

LXMERT складається з двох окремих енкодерів, мовного та візуального, об'єднаних за допомогою механізму крос-модальної уваги. Архітектура передбачає поетапне навчання: спочатку енкодери тренуються окремо, потім виконується спільне узгодження у латентному просторі [7]. Це дозволяє моделі детальніше враховувати міжмодальні зв'язки, що є критично важливим для складних завдань, таких як генерація описів або відповіді на запитання за зображеннями (рис. 3).

Perceiver IO реалізує універсальну схему обробки, здатну приймати на вхід будь-яку комбінацію модальностей (рис. 4). Модель використовує латентні вектори фіксованої довжини, на які через крос-увагу проєктуються всі вхідні сигнали [8]. Це

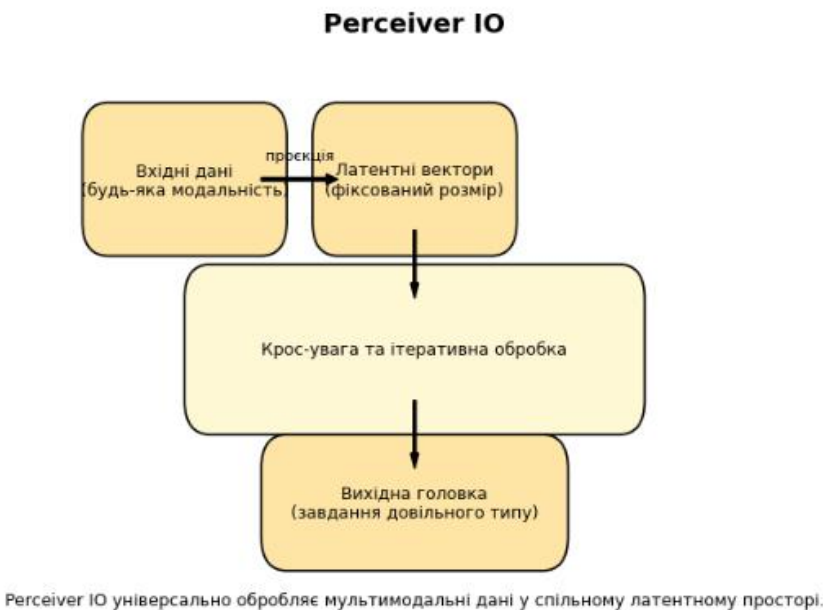
INFORMATION AND WEB TECHNOLOGIES

забезпечує масштабованість і стабільність навчання навіть при великій кількості даних різної природи.



LXMERT узгоджує представлення мовних і візуальних ознак через крос-увагу.

Рисунок 3
Модель LXMERT



Perceiver IO універсально обробляє мультимодальні дані у спільному латентному просторі.

Рисунок 4
Модель Perceiver IO

INFORMATION AND WEB TECHNOLOGIES

У табл. 1 наведено порівняльну характеристику найвідоміших мультимодальних архітектур за ключовими параметрами.

Таблиця 1

Порівняльний аналіз мультимодальних архітектур				
Модель	Тип злиття	Основна структура	Ключова перевага	Основне застосування
CLIP [9]	Пізнє/Контрастивне	Два енкодери (ResNet/Transformer)	Семантичне узгодження	Текстово-візуальний пошук
ViLT [10]	Раннє	Єдиний Transformer	Висока швидкість, менше параметрів	Мультимодальна класифікація
LXMERT [11]	Гібридне	Подвійний енкодер + Cross-Attention	Глибоке моделювання зв'язків	Відповіді на запитання, captioning
Perceiver IO [12]	Гнучке	Універсальний латентний простір	Масштабованість	Обробка гетерогенних даних

Щоб поглибити аналіз, доцільно розглянути архітектури з погляду їхньої структурної складності та ефективності узгодження модальностей (табл. 2).

Таблиця 2

Порівняння структурної складності та ефективності узгодження модальностей				
Модель	Кількість параметрів (млн)	Глибина (кількість шарів)	Рівень узгодження (0-1)	Придатність до нових модальностей
CLIP [9]	150-400	12-24	0.87	Висока
ViLT [10]	110	12	0.79	Середня
LXMERT [11]	180	24	0.85	Середня
Perceiver IO [12]	160	28	0.91	Дуже висока

Після аналізу таблиці можна виділити кілька ключових аспектів, які визначають ефективність мультимодальних архітектур:

1. Стратегія злиття модальностей безпосередньо впливає на якість отриманого спільного представлення. У CLIP контрастивне навчання дозволяє кожному вектору тексту максимально точно відповідати відповідному зображенню, що

INFORMATION AND WEB TECHNOLOGIES

формує глибокий семантичний зв'язок і забезпечує здатність моделі працювати з новими наборами даних без додаткового тренування. ViLT, навпаки, застосовує раннє злиття, що значно спрощує архітектуру та пришвидшує навчання, але деякі специфічні характеристики модальностей можуть частково втрачатися через недостатнє відокремлення інформації на початковому рівні.

2. Структурна складність архітектур визначає обчислювальні ресурси, необхідні для навчання та інференсу. LXMERT демонструє значну ефективність у завданнях, де потрібна глибока взаємодія між модальностями, завдяки крос-модальній увазі, проте вимагає більше обчислювальної потужності і великих датасетів для досягнення стабільних результатів. Perceiver IO показує, що навіть при складних і великомасштабних наборах даних можливо забезпечити ефективну обробку через універсальні латентні простори, що робить її перспективною для мультисенсорних застосувань.

3. Масштабованість і адаптивність моделей є критично важливими при роботі з реальними даними. Моделі з окремими енкодерами (CLIP, LXMERT) дозволяють гнучко додавати нові модальності без повного перенавчання, тоді як ViLT більш обмежена в цьому аспекті. Perceiver IO демонструє максимальну гнучкість, здатну інтегрувати одночасно будь-яку комбінацію модальностей, що відкриває перспективи для універсальних мультимодальних систем.

4. Інтерпретованість результатів залежить від механізмів уваги і контрастивного навчання. У LXMERT і Perceiver IO крос-модальна увага дозволяє відстежувати, які елементи тексту чи зображення найбільше вплинули на прогноз моделі. У CLIP контрастивне навчання формує зрозумілий семантичний простір, де відстань між векторами прямо корелює зі змістовною близькістю. ViLT, завдяки єдиному трансформеру, менш інтерпретована, проте компенсує це швидкістю навчання та компактністю моделі.

Додатково ще варто відзначити практичні аспекти використання цих архітектур. CLIP та ViLT показали високий потенціал у задачах пошуку та класифікації, LXMERT – у генерації описів та відповіді на питання, а Perceiver IO – у мультисенсорних завданнях, наприклад, для автономних систем або інтеграції відео, аудіо та сенсорних даних. Ефективність

INFORMATION AND WEB TECHNOLOGIES

моделей також визначається оптимізацією гіперпараметрів, балансом класів у навчальних датасетах та використанням технік регуляризації, таких як Dropout або Layer Normalization.

Отже, аналіз показує, що гібридні та контрастивні підходи забезпечують найкраще узгодження модальностей і високий рівень узагальнення, тоді як архітектури з раннім злиттям підходять для сценаріїв із обмеженими обчислювальними ресурсами [13]. Масштабованість і гнучкість моделей є важливими для практичного використання у реальних системах, де дані часто неповні або шумні.

Висновки. Дослідження сучасних архітектур нейронних мереж для обробки мультимодальних даних показало, що ефективність моделей визначається стратегією інтеграції модальностей та здатністю формувати спільні латентні простори. Архітектури на основі контрастивного навчання та крос-модальної уваги, такі як CLIP, LXMERT і Perceiver IO, демонструють високу точність і здатність узгоджувати різноманітні дані, тоді як ViLT відзначається швидкістю та компактністю моделі, що робить її придатною для систем з обмеженими ресурсами.

Модульність, масштабованість і адаптивність архітектур, а також застосування оптимізаційних методів підвищують точність і стабільність моделей навіть при неповних або шумних даних. Практична значимість мультимодальних нейронних мереж проявляється у медичній діагностиці, автономному транспорті та задачах генерації текстових описів для зображень [14].

References:

- [1] Wang, Z. SimVLM: Simple Visual Language Model Pretraining with Weak Supervision / Z. Wang, J. Yu, A. W. Yu [et al.] // arXiv. – 2021. <https://doi.org/10.48550/arXiv.2108.10904>.
- [2] Мінухін С.В. Дослідження моделі сегментації зображень з використанням розподілених режимів TensorFlow та згорткової нейронної мережі U-Net / С. В. Мінухін. // Системи обробки інформації. – 2020. – № 1(160). – С. 115–122. <https://doi.org/10.30748/soi.2020.160.15>.
- [3] Bai, Y. Sequential Modeling Enables Scalable Learning for Large Vision Models / Y. Bai, X. Geng, K. Mangalam [et al.] // arXiv. – 2022. <https://doi.org/10.48550/arXiv.2206.11389>.
- [4] Dosovitskiy, A. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale / A. Dosovitskiy, L. Beyer, A. Kolesnikov [et al.] // International Conference on Learning Representations. – 2021. <https://doi.org/10.48550/arXiv.2010.11929>.

INFORMATION AND WEB TECHNOLOGIES

- [5] Garcia, N. A Survey on Vision-Language Pre-training: Evolution, Foundations and Applications / N. Garcia, C. Ye, Z. Liu [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2023. - Vol. 45, No. 9. - P. 10781-10800. <https://doi.org/10.1109/TPAMI.2023.3261989>.
- [6] Kim, W. ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision / W. Kim, B. Son, I. Kim // Proceedings of the 38th International Conference on Machine Learning. - 2021. - Vol. 139. - P. 5583-5594. <https://doi.org/10.48550/arXiv.2102.03334>.
- [7] Li, J. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation / J. Li, D. Li, C. Xiong, S. Hoi // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. - 2022. - P. 12808-12817. <https://doi.org/10.48550/arXiv.2201.12086>.
- [8] Liu, Y. A Closer Look at Invariances in Self-Supervised Pre-training for Vision-Language Models / Y. Liu, Z. Wu, Z. Kira, Y. Wang // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. - 2022. - P. 15378-15388. <https://doi.org/10.48550/arXiv.2205.13135>.
- [9] Zeng, Z. Pixel-BERT: Aligning Image Pixels with Text by Deep Multi-Modal Transformers / Z. Zeng, Z. Huang, B. Liu [et al.] // arXiv. - 2020. <https://doi.org/10.48550/arXiv.2004.00849>.
- [10] Radford, A. Learning Transferable Visual Models from Natural Language Supervision / A. Radford, J. W. Kim, C. Hallacy [et al.] // Proceedings of the 38th International Conference on Machine Learning. - 2021. - Vol. 139. - P. 8748-8763. <https://doi.org/10.48550/arXiv.2103.00020>.
- [11] Alayrac, J.-B. Flamingo: a Visual Language Model for Few-Shot Learning / J.-B. Alayrac, J. Donahue, P. Luc [et al.] // Advances in Neural Information Processing Systems. - 2022. - Vol. 35. - P. 23716-23736. <https://doi.org/10.48550/arXiv.2204.14198>.
- [12] Yu, J. CoCa: Contrastive Captioners are Image-Text Foundation Models / J. Yu, Z. Wang, V. Vasudevan [et al.] // Transactions on Machine Learning Research. - 2022. <https://doi.org/10.48550/arXiv.2205.01917>.
- [13] Zhang, Y. Contrastive Learning of Medical Visual Representations from Paired Images and Text / Y. Zhang, H. Jiang, Y. Miura [et al.] // Proceedings of the Conference on Health, Inference, and Learning. - 2022. - P. 2-25. <https://doi.org/10.48550/arXiv.2110.07931>.
- [14] Zhou, L. Unified Vision-Language Pre-Training for Image Captioning and VQA / L. Zhou, H. Palangi, L. Zhang [et al.] // Proceedings of the AAAI Conference on Artificial Intelligence. - 2020. - Vol. 34, No. 07. - P. 13041-13049. <https://doi.org/10.48550/arXiv.1909.11059>.