

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

**ACADEMIC SCIENTIFIC
JOURNAL OF COMPUTER SCIENCE**

**№2
2026**

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)



ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

3 (359)

JULY – SEPTEMBER 2026

**PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR**

ALMATY, NAS RK

Chief Editor:

MUTANOV Galimkair Mutanovich, doctor of technical sciences, professor, academician of NAS RK, (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

EDITORIAL BOARD:

KALIMOLDAYEV Maksat Nuradilovich, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

MAMYRBAEV Orken Zhumazhanovich, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

BAIGUNCHEKOV Zhumadil Zhanabaevich, Doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

WOICIK Waldemar, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

SMOLARJ Andrej, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

KEILAN Alimkhan, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

KHAIROVA Nina, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

OTMAN Mohamed, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

NYSANBAYEVA Saule Yerkebulanovna, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

USATOVA Olga Alexandrovna, PhD, Associate Professor, Chief Scientific Secretary of the Institute of Information and Computing Technologies of the National Academy of Sciences of the Republic of Kazakhstan (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

KAPALOVA Nursulu Aldazharovna, Candidate of Technical Sciences, Head of the Laboratory cybersecurity, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

KOVALYOV Alexander Mikhailovich, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

MIKHALEVICH Alexander Alexandrovich, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

TIGHINEANU Ion Mihailovich, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: «Central Asian Academic Research Center» LLP (Almaty).

Certificate № KZ77VPY00121154 on the re-registration of the periodical printed and online publication of the information agency, issued on 05.06.2025 by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan

Subject area: *information and communication technologies*.

Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies»*.

Periodicity: 4 times a year.

<http://www.physico-mathematical.kz/index.php/en/>

© «Central Asian Academic Research Center» LLP, 2026

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат Нұрәділұлы, (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ Өркен Жұмажанұлы (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жұмаіл Жаңабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛАРЖ Анджей, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Әлімхан, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония)), ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, техника ғылымдарының докторы, профессор, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТМАН Мохаммед, PhD, Информатика, Коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебұланқызы, техника ғылымдарының докторы, доцент, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» аға ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

УСАТОВА Ольга Александровна, PhD, қауымдастырылған профессор, ҚР ҒЖБМ "Ақпараттық және есептеу технологиялары институтының" бас ғалым хатшысы (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

КАПАЛОВА Нұрсұлу Алдажарқызы, техника ғылымдарының кандидаты, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты», Киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александр Александрович, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).

Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025 ж.** берген **№ KZ77VPY00121154** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

<http://www.physico-mathematical.kz/index.php/en/>

© «Орталық Азия академиялық ғылыми орталығы» ЖШС, 2026

Главный редактор:

МУТАНОВ Галимканр Мутанович, доктор технических наук, профессор, академик НАН РК, (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

Редакционная коллегия:

КАЛИМОЛДАЕВ Максат Нурадилович, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ Оркен Жумажанович, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сатпаева (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛАРЖ Анджей, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Алимхан, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТМАН Мохамед, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебулановна, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

УСАТОВА Ольга Александровна, PhD, ассоциированный профессор, Главный ученый секретарь «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

КАПАЛОВА Нурсулу Алдажаровна, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александр Александрович, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания № **KZ77VPY00121154**. Дата выдачи **05.06.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

В настоящее время: *вошел в список журналов, рекомендованных КОКШВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

<http://www.physico-mathematical.kz/index.php/en/>

© ТОО «Центрально-азиатский академический научный центр», 2026

CONTENTS

Akhmetova A.Zh., Kantureyeva M.A., Serimbetov B.A., Zhasuzakova M.Zh., Mustafayeva N.T. Identification of communities and active users in social networks.....	11
Akzhigitov D., Mukhtarkhanov M., Zhetpisbayeva A., Kulakayeva A., Zhetpisbayev K. Analysis of the mechanical structure of a 12U CubeSat nanosatellite and assessment of its mechanical stability.....	29
Alibekov A., Nurimanov A., Mukhametkali A., Kassym D., Varol H.A. Qolda-Math: a small vision-language model for mathematical reasoning in Kazakh.....	50
Alisheva S.S., Ospanova A.O., Shaimerdenova G.S., Shaimerdenova L., Zhidebayeva A.N. Detecting Kazakh language offensive comments in online content using deep learning algorithms.....	68
Askar G., Mutaikhan A., Aitkozhasyeva M., Nauryzbaikyzy G., Makhataeva Z. Retrieval beats fine-tuning: adapting a Kazakh small language model for historical question answering.....	86
Beissenova G.I., Shynzhigit B.B., Beisebayeva A.Zh., Kaldarova B.S., Lesbayev A.U. Deep neural network-based speaker authentication for secure access control systems.....	106
Bekenuly B., Baimakhanova A., Baiturganova V., Bolatbekova D. Comparison of CNN architectures for automated ACL injury classification from knee MRI.....	123
Bekbolat A., Razak A., Alimbayeva Z., Ozhikenova A. Multi-task deep learning for knee MRI pathology classification.....	139
Issakhov A., Akylbekkyzy A., Abylkassymova A. Predicting the spread of airborne viral infections in enclosed spaces.....	158
Kisseleva O., Abdramanov K., Naumenko V., Mukanova Z., Maintser D. Evaluation of serverless cold start optimization techniques.....	177
Koshkarov Ye.Zh., Amirgaliyev B.Ye., Mulesa O. Physics-aware masked pretraining for cross-basin well log foundation models.....	197
Kurmetkan T., Turysbek Zh., Mukhitova A., Nauryzbaeva A. Research on methods for reducing knowledge conflicts in large language models in Kazakh question-and-answer systems.....	214
Makatov Ye.K., Abdul Razaque, Ismukanova A.N., Salykova A.K., Glock E.S. Machine learning-based model for phishing URL detection in online social networks.....	233
Mimenbayeva A.B., Naizagarayeva A.A., Ospanova T.T., Koikelova D.K., Kudabayeva R.N. Spatial and climate data analysis for machine learning-based crop yield prediction in Northern Kazakhstan.....	255
Mukhamediev R.I., Abdoldina F.N., Gertsen Y.A. Economic opportunities and constraints of artificial intelligence adoption in Kazakhstan's resource-oriented economy.....	274

Mutanov G.M., Karyukin V.I., Ussatova O.A., Begimbayeva Y.Y., Baispay G.B. The assessment of blockchain readiness for academic document verification.....	289
Mussiraliyeva Sh., Bolatbek M., Baisylbayeva K. Classification of cyber-extremist ideologies in the Kazakh language considering multi-annotator disagreement.....	306
Omarov B.S., Sultan D.R., Rakhymberdiev D.S., Abibullayeva A.A., Toktarova A.B. Deformable attention-guided Transformer for brain tumor segmentation.....	326
Oralbekova D., Bekarystankyzy A., Yerimbetova A., Berzhanova U. Development of a streaming Kazakh speech recognition system using end-to-end models.....	346
Serek A., Shoiynbek A., Zheludkov M., Kuanyshbay D., Meraliyev B. Comparative analysis of machine learning methods for detecting fraud in text messages.....	365
Smakhanova A.K., Uskenbayeva R.K., Kassymova A.B., Kalpeyeva Zh.B., Konyrbaev N.B. Experimental research using models and methods of the university digital ecosystem.....	384
Sultanali A.E., Kazbekova G.N. Machine learning-based detection of AI-generated content in students' written works.....	401
Tashenova Zh., Yergazin A., Nurlybaeva E., Abdugulova Zh., Amanzholova Sh. Experimental evaluation of a hybrid IDPS with adaptive learning for low-latency 5G networks using edge computing.....	418
Telman D., Yerimbetova A., Daiyrbayeva E. Leakage-resistant fake news detection for Kazakh: temporal evaluation of lexical, Transformer and hybrid models.....	441
Zhanegizov A., Mamyrbayev O., Abdigappar M. Application of YOLO12 for road defect detection in GPR images.....	457

МАЗМҰНЫ

Ахметова А.Ж., Кантуреева М.А., Серимбетов Б.А., Жасұзақова М.Ж., Мұстафаева Н.Т. Әлеуметтік желілердегі қауымдастықтар мен белсенді қолданушыларды анықтау.....	11
Ақжігітов Д.Ж., Мұхтарханов М.О., Жетписбаева А.Т., Кулакаева А.Е., Жетписбаев К.У. 12U CubeSat үлгісіндегі наноспутниктің механикалық құрылымын талдау және оның механикалық орнықтылығын бағалау.....	29
Алибеков А., Нуриманов А., Мұхаметқали А., Қасым Д., Варол Х.А. Qolda-Math: қазақ тіліндегі математикалық пайымдауға арналған кіші көру-тілдік модель.....	50
Алишева С.С., Оспанова А.О., Шаймерденова Г.С., Шаймерденова Л., Жидебаева А.Н. Онлайн-контенттегі қазақ тіліндегі бейәдеп пікірлерді терең оқыту алгоритмдері арқылы анықтау.....	68
Асқар Ғ., Мұтайхан А., Айтқожаева М., Наурызбайқызы Г., Махатаева Ж. Retrieval fine-tuning-тен басым: тарихи сұрақтарға жауап беру үшін қазақша шағын тілдік модельді бейімдеу.....	86
Бейсенова Г.И., Шынжігіт Б.Б., Бейсебаева А.Ж., Қалдарова Б.С., Лесбаев А.У. Терең нейрондық желіге негізделген қауіпсіз басқару жүйелерінде сөйлеушіні аутентификациялау.....	106
Бекенұлы Б., Баймаханова А., Байтурганова В., Болатбекова Д. Алдыңғы крест тәрізді байлам жарақатын анықтауға арналған CNN архитектураларын салыстыру.....	123
Бекболат А., Разак А., Алимбаева Ж., Ожикенова А. Тізе MPT патологиясын жіктеуге арналған көп тапсырмалы терең оқыту.....	139
Исahов А.А., Ақылбекқызы А., Абылкасымова А. Шектеулі кеңістіктерде әуе тамшыларымен вирустық инфекцияның таралуын сандық модельдеу.....	158
Киселева О., Абдраманов Қ., Науменко В., Мұқанова Ж., Майнцер Д. Serverless жүйелеріндегі суық іске қосуды оңтайландыру әдістерін эксперименттік бағалау.....	177
Кошкарлов Е.Ж., Амиргалиев Б.Е., Мүлеса О. Бассейнаралық ұңғымалық диаграммалардың іргелі модельдерін физикалық тұрғыдан негізделген масқаланған алдын ала оқыту.....	197
Құрметқан Т., Тұрысбек Ж., Мұхитова А., Наурызбаева А. Қазақ тіліндегі сұрақ-жауап жүйелерінде үлкен тілдік модельдердегі білім қайшылықтарын азайту әдістерін зерттеу.....	214
Мақатов Е.К., Abdul Razaque, Исмуканова А.Н., Салыкова А.К., Глюк Е.С. Онлайн әлеуметтік желілердегі фишингтік URL-дарды анықтауға арналған машиналық оқыту негізіндегі модель.....	233

Мименбаева А.Б., Найзағарова А.А., Оспанова Т.Т., Қойкелова Д.К., Құдабаева Р.Н. Солтүстік Қазақстандағы дақылдар өнімділігін машиналық оқыту арқылы болжауда кеңістіктік және климаттық деректерді талдау.....	255
Мухамедиев Р.И., Абдолдина Ф.Н., Герцен Е.А. Қазақстанның ресурстық бағытталған экономикасына жасанды интеллектті енгізудің экономикалық мүмкіндіктері мен шектеулері.....	274
Мутанов Г.М., Карюкин В.И., Усатова О.А., Бегимбаева Е.Е., Байспай Г.Б. Білім беру құжаттарының түпнұсқалығын тексеруге блокчейн технологиясының дайындығын бағалау.....	289
Мусиралиева Ш., Болатбек М., Байсылбаева К. Көпаннынотаторлы келіспеушілікті ескере отырып қазақ тіліндегі киберэкстремистік идеологияларды жіктеу.....	306
Омаров Б.С., Сұлтан Д.Р., Рахымбердиев Д.С., Абибуллаева А.А., Тоқтарова А.Б. Ми ісіктерін сегменттеуде деформацияланатын зейін механизміне негізделген мультимодальды трансформер.....	326
Оралбекова Д.О., Бекарыстанқызы А., Еримбетова А.С., Бержанова У.Г. Интегралды модельдерді қолдану арқылы қазақ тіліндегі ағындық сөйлеуді тану жүйесін әзірлеу.....	346
Серек А., Шойынбек А., Желудков М., Қуанышбай Д., Мералиев Б. Мәтіндік хабарламалардағы алаяқтықты анықтауға арналған машиналық оқыту әдістерінің салыстырмалы талдауы.....	365
Смаханова А.К., Ускенбаева Р.К., Касымова А.Б., Кальпеева Ж.Б., Қоңырбаев Н.Б. Университеттің цифрлық экожүйесінің модельдері мен әдістерін қолдана отырып жүргізілген эксперименттік зерттеу.....	384
Сұлтанәлі А.Е., Казбекова Г.Н. Білім алушылардың жазбаша жұмыстарындағы AI-контентті машиналық оқыту арқылы анықтау.....	401
Ташенова Ж.М., Ергазин Ә., Нұрлыбаева Ә.Н., Абдуғұлова Ж.К., Аманжолова Ш.А. Edge Computing көмегімен төмен кідірісті 5G желілеріне арналған бейімделгіш оқытуы бар гибриді IDPS-ті эксперименттік бағалау.....	418
Тельман Д., Еримбетова А., Дайырбаева Ә. Қазақ тіліндегі жалған жаңалықтарды анықтау: лексикалық, трансформерлік және гибридті модельдерді уақыт бойынша бағалау.....	441
Жанегизов А.С., Мамырбаев Ә.Ж., Абдіғаппар М.Д. YOLO12 моделін георадар кескіндеріне қолдану.....	457

СОДЕРЖАНИЕ

Ахметова А.Ж., Кантуреева М.А., Серимбетов Б.А., Жасузакова М.Ж., Мустафаева Н.Т. Выявление сообществ и активных пользователей в социальных сетях.....	11
Акжигитов Д.Ж., Мухтарханов М.О., Жетписбаева А.Т., Кулакаева А.Е., Жетписбаев К.У. Анализ механической конструкции наноспутника модели 12U CubeSat и оценка его механической устойчивости.....	29
Алибеков А., Нуриманов А., Мухаметкали А., Касым Д., Варол Х.А. Qolda-Math: маленькая мультимодальная модель зрения и языка для математических рассуждений на казахском языке.....	50
Алишева С.С., Оспанова А.О., Шаймерденова Г.С., Шаймерденова Л., Жидебаева А.Н. Выявление оскорбительных комментариев на казахском языке в онлайн-контенте с помощью алгоритмов глубокого обучения.....	68
Аскар Г., Мутайхан А., Айткожаева М., Наурызбайкызы Г., Махатаева Ж. Retrieval эффективнее fine-tuning: адаптация малой языковой модели казахского языка для ответов на исторические вопросы.....	86
Бейсенова Г.И., Шынжигит Б.Б., Бейсебаева А.Ж., Калдарова Б.С., Лесбаев А.У. Аутентификация личности по голосу на основе глубокой нейронной сети для систем безопасного управления доступом.....	106
Бекенулы Б., Баймаханова А., Байтурганова В., Болатбекова Д. Сравнение архитектур сверточных нейронных сетей для автоматической классификации разрыва ПКС по МРТ коленного сустава.....	123
Бекболат А., Разак А., Алимбаева Ж., Ожикенова А. Многозадачное глубокое обучение для классификации патологий колена по МРТ.....	139
Исахов А.А., Акылбеккызы А., Абылкасымова А. Прогнозирование распространения вирусной инфекции воздушно-капельным путем в замкнутых помещениях.....	158
Киселева О., Абдраманов К., Науменко В., Муканова Ж., Майнцер Д. Экспериментальная оценка методов оптимизации холодного запуска в serverless-системах.....	177
Кошкарар Е.Ж., Амиргалиев Б.Е., Мулеса О. Физически-осознанное маскированное предобучение для межбассейновых фундаментальных моделей каротажных диаграмм.....	197
Курметкан Т., Турысбек Ж., Мухитова А., Наурызбаева А. Исследование методов снижения конфликтов знаний в больших языковых моделях казахских вопросно-ответных систем.....	214
Макатов Е.К., Abdul Razaque, Исмуканова А.Н., Салыкова А.К., Глок Е.С. Модель обнаружения фишинговых URL в онлайн-социальных сетях на основе машинного обучения.....	233

Мименбаева А.Б., Найзагараева А.А., Оспанова Т.Т., Койкелова Д.К., Кудабаева Р.Н. Анализ климатических данных для прогнозирования урожайности пшеницы методами машинного обучения в Северном Казахстане.....	255
Мухамедиев Р.И., Абдолдина Ф.Н., Герцен Е.А. Экономические возможности и ограничения внедрения искусственного интеллекта в ресурсно-ориентированной экономике Казахстана.....	274
Мутанов Г.М., Карюкин В.И., Усатова О.А., Бегимбаева Е.Е., Байспай Г.Б. Оценка готовности технологии блокчейн к проверке подлинности документов об образовании.....	289
Мусиралиева Ш., Болатбек М., Байсылбаева К. Классификация киберэкстремистских идеологий на казахском языке с учетом разногласий нескольких аннотаторов.....	306
Омаров Б.С., Султан Д.Р., Рахымбердиев Д.С., Абибуллаева А.А., Токтарова А.Б. Мультимодальный трансформер на основе механизма деформируемого внимания для сегментации опухолей головного мозга.....	326
Оралбекова Д.О., Бекарыстанкызы А., Еримбетова А.С., Бержанова У.Г. Распознавание казахской речи с применением интегральных моделей.....	346
Серек А., Шойынбек А., Желудков М., Куанышбай Д., Мералиев Б. Сравнительный анализ методов машинного обучения для выявления мошенничества в текстовых сообщениях.....	365
Смаханова А.К., Ускенбаева Р.К., Касымова А.Б., Кальпеева Ж.Б., Конырбаев Н.Б. Экспериментальное исследование с использованием моделей и методов цифровой экосистемы университета.....	384
Султанали А.Е., Казбекова Г.Н. Обнаружение AI-контента в письменных работах обучающихся методами машинного обучения.....	401
Ташенова Ж.М., Ергазин А., Нурлыбаева Э.Н., Абдугулова Ж.К., Аманжолова Ш.А. Экспериментальная оценка гибридной системы обнаружения и предотвращения вторжений (IDPS) с адаптивным обучением для сетей 5G с низкой задержкой с использованием граничных вычислений.....	418
Тельман Д., Еримбетова А., Дайырбаева Э. Обнаружение фейковых новостей на казахском языке: временная оценка лексических, трансформерных и гибридных моделей.....	441
Жанегизов А.С., Мамырбаев О.Ж., Абдигаппар М.Д. Применение модели YOLOv12 для анализа данных георадара.....	457

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE
ISSN 1991-346X
Volume 3.
Number 359 (2026). 197–213

<https://doi.org/10.32014/2026.2518-1726.459>

IRSTI 52.47.19 + 28.23.29
UDC 550.832:004.85

© **Koshkarov Ye.Zh.^{1*}, Amirgaliyev B.Ye.¹, Mulesa O.²**, 2026.

¹Astana IT University, Astana, Kazakhstan;

²University of Prešov, Prešov, Slovakia.

E-mail: yerzhan.koshkarov@astanait.edu.kz

PHYSICS-AWARE MASKED PRETRAINING FOR CROSS-BASIN WELL LOG FOUNDATION MODELS

Koshkarov Yerzhan — Doctoral student, Astana IT University, Astana, Kazakhstan,

E-mail: yerzhan.koshkarov@astanait.edu.kz, <https://orcid.org/0009-0001-0801-0647>;

Amirgaliyev Beibut — candidate of technical sciences, professor Astana IT University, Astana, Kazakhstan,

E-mail: beibut.amirgaliyev@astanait.edu.kz, <https://orcid.org/0000-0003-0355-5856>;

Mulesa Oksana — Professor, University of Prešov, Prešov, Slovakia,

E-mail: oksana.mulesa@unipo.sk, <https://orcid.org/0000-0002-6117-5846>.

Abstract. Well log interpretation faces significant challenges due to data scarcity, distribution shift across geological basins, and a lack of physics-informed representations in current foundation models. While recent masked autoencoder approaches show promise for self-supervised pretraining on subsurface data, they fail to consider fundamental petrophysical constraints that govern measurement relationships. This paper presents a physics-aware masked pretraining framework that incorporates petrophysical consistency losses into self-supervised learning on multivariate well logs. The proposed method combines masked patch reconstruction with differentiable petrophysical forward models that use Archie's law, Wyllie's time-average equation, and density-porosity relationships as auxiliary objectives. A geologically adaptive masking curriculum also gradually increases task difficulty, moving from random point masking to cross-curve masking and stratigraphic interval masking. Extensive experiments on the FORCE 2020 benchmark show that physics-aware pretraining improves lithology classification accuracy by 4.2% and reduces porosity estimation error by 5.8% compared to standard masked autoencoders, while cross-basin transfer experiments show a 6.3% lower optimality gap when transferring datasets from Norwegian to Kansas wells. Ablation studies show that petrophysical consistency losses contribute 2.8 percentage points to classification improvement, while geologically adaptive masking adds another

1.4 percentage points. The framework maintains computational efficiency, with inference times of less than 12 milliseconds per well section, making it suitable for real-time drilling applications in data-scarce, multi-basin settings.

Keywords: well log interpretation, foundation models, self-supervised learning, petrophysical constraints, masked autoencoders, transfer learning, subsurface characterization

For citations: Koshkarov Ye.Zh., Amirgaliyev B.Ye., Mulesa O. Physics-aware masked pretraining for cross-basin well log foundation models. Academic Scientific Journal of Computer Science, 2026. — No.3. — P. 197–213. DOI: <https://doi.org/10.32014/2026.2518-1726.459>

© Кошкарров Е.Ж.^{1*}, Амиргалиев Б.Е.¹, Мулеса О.², 2026.

¹ Astana IT University, Астана, Қазақстан;

² Прешов университеті, Прешов, Словакия.

E-mail: yerzhan.koshkarov@astanait.edu.kz

БАССЕЙНАРАЛЫҚ ҰҢҒЫМАЛЫҚ ДИАГРАММАЛАРДЫҢ ІРГЕЛІ МОДЕЛЬДЕРІН ФИЗИКАЛЫҚ ТҰРҒЫДАН НЕГІЗДЕЛГЕН МАСКАЛАНҒАН АЛДЫН АЛА ОҚЫТУ

Кошкарров Ержан — докторант, Astana IT University, Астана, Қазақстан,
E-mail: yerzhan.koshkarov@astanait.edu.kz, <https://orcid.org/0009-0001-0801-0647>;

Амиргалиев Бейбут — техника ғылымдарының кандидаты, профессор, Astana IT University, Астана, Қазақстан,

E-mail: beibut.amirgaliyev@astanait.edu.kz, <https://orcid.org/0000-0003-0355-5856>;

Мулеса Оксана — профессор, Прешов университеті, Прешов, Словакия,

E-mail: oksana.mulesa@unipo.sk, <https://orcid.org/0000-0002-6117-5846>.

Аннотация. Ұңғымалық диаграммаларды интерпретациялау деректердің шектеулілігіне, әртүрлі геологиялық бассейндер арасындағы деректер таралымының өзгеруіне және қазіргі іргелі модельдерде физикалық тұрғыдан негізделген көріністердің жеткіліксіз қолданылуына байланысты елеулі қиындықтармен сипатталады. Маскаланған автоэнкодерлерге негізделген соңғы тәсілдер жер қойнауы деректерін өздігінен бақыланатын түрде алдын ала оқытуда перспективалы нәтижелер көрсеткенімен, өлшеулер арасындағы өзара байланыстарды анықтайтын негізгі петрофизикалық шектеулерді көбіне ескермейді. Мақалада көпөлшемді ұңғымалық диаграммаларда өздігінен бақыланатын оқытуға петрофизикалық сәйкестік шығындарын енгізетін физикалық тұрғыдан негізделген маскаланған алдын ала оқыту тәсілі ұсынылады. Әдіс маскаланған патчтарды қалпына келтіруді Арчи заңы, Уайллидің орташа уақыт теңдеуі және тығыздық-кеуектілік арақатынастарына негізделген дифференциалданатын тура петрофизикалық модельдермен қосымша мақсаттар ретінде біріктіреді. Сонымен қатар

тапсырманың күрделілігін біртіндеп арттыратын геологиялық бейімделген маскалау стратегиясы қолданылады: кездейсоқ нүктелерді маскалаудан қысқаралық маскалауға және стратиграфиялық интервалдарды маскалауға дейін. FORCE 2020 эталондық деректер жиынтығында жүргізілген кең ауқымды эксперименттер физикалық тұрғыдан негізделген алдын ала оқыту стандартты масаланған автоэнкодерлермен салыстырғанда литологияны жіктеу дәлдігін 4,2%-ға арттырып, кеуектілікті бағалау қатесін 5,8%-ға төмендететінін көрсетті. Бассейндер арасындағы тасымалдау эксперименттерінде Норвегия ұңғымаларынан Канзас ұңғымаларына модельді бейімдеу кезінде оңтайлылық алшақтығы 6,3%-ға төмендеді. Абляциялық зерттеулер петрофизикалық сәйкестік шығындарының жіктеу сапасын 2,8 пайыздық тармаққа жақсартатынын, ал геологиялық бейімделген маскалаудың қосымша 1,4 пайыздық тармақ беретінін көрсетті. Ұсынылған тәсіл есептеу тиімділігін де сақтайды: бір ұңғыма секциясына модельдің нәтиже шығару уақыты 12 миллисекундтан аз. Бұл оны деректер шектеулі және бірнеше бассейнді қамтитын жағдайларда нақты уақыттағы бұрғылау міндеттеріне қолдануға мүмкіндік береді.

Түйін сөздер: ұңғымалық диаграммаларды интерпретациялау; іргелі модельдер; өздігінен бақыланатын оқыту; петрофизикалық шектеулер; масаланған автоэнкодерлер; тасымалды оқыту; жер қойнауын сипаттау

© Кошкарлов Е.Ж.^{1*}, Амиргалиев Б.Е.¹, Мулеса О.², 2026.

¹ Astana IT University, Астана, Казахстан;

² Прешовский университет, Прешов, Словакия.

E-mail: yerzhan.koshkarov@astanait.edu.kz

ФИЗИЧЕСКИ-ОСОЗНАННОЕ МАСКИРОВАННОЕ ПРЕДОБУЧЕНИЕ ДЛЯ МЕЖБАССЕЙНОВЫХ ФУНДАМЕНТАЛЬНЫХ МОДЕЛЕЙ КАРОТАЖНЫХ ДИАГРАММ

Кошкарлов Ержан — докторант, Astana IT University, Астана, Казахстан,

E-mail: yerzhan.koshkarov@astanait.edu.kz, <https://orcid.org/0009-0001-0801-0647>;

Амиргалиев Бейбут — кандидат технических наук, профессор, Astana IT University, Астана, Казахстан,

E-mail: beibut.amirgaliyev@astanait.edu.kz, <https://orcid.org/0000-0003-0355-5856>;

Мулеса Оксана — профессор, Прешовский университет, Прешов, Словакия,

E-mail: oksana.mulesa@unipo.sk, <https://orcid.org/0000-0002-6117-5846>.

Аннотация. *Актуальность.* Интерпретация каротажных диаграмм сопряжена с рядом существенных проблем, включая дефицит данных, изменение их распределения между различными геологическими бассейнами и недостаточное использование физически обоснованных представлений в современных фундаментальных моделях. Несмотря на перспективность

подходов на основе маскированных автоэнкодеров для самоконтролируемого предобучения на данных о недрах, такие методы зачастую не учитывают фундаментальные петрофизические ограничения, определяющие взаимосвязи между измеряемыми параметрами. *Цель.* Разработать физически обоснованный подход к маскированному предобучению, включающий петрофизические функции согласованности в процесс самоконтролируемого обучения на многомерных каротажных диаграммах. *Методы.* Предложенный метод объединяет восстановление маскированных патчей с дифференцируемыми прямыми петрофизическими моделями, основанными на законе Арчи, уравнении среднего времени Уайлли и зависимостях между плотностью и пористостью, используемых в качестве дополнительных целевых функций. Кроме того, применяется геологически адаптивная стратегия маскирования, постепенно увеличивающая сложность задачи: от случайного маскирования отдельных точек к межкривому маскированию и маскированию стратиграфических интервалов. *Результаты.* Эксперименты на эталонном наборе данных FORCE 2020 показали, что физически обоснованное предобучение повышает точность классификации литологии на 4,2% и снижает ошибку оценки пористости на 5,8% по сравнению со стандартными маскированными автоэнкодерами. Эксперименты по межбассейновому переносу продемонстрировали снижение разрыва оптимальности на 6,3% при переносе модели с данных норвежских скважин на данные скважин Канзаса. Абляционный анализ показал, что петрофизические функции согласованности обеспечивают прирост точности классификации на 2,8 процентного пункта, тогда как геологически адаптивное маскирование добавляет еще 1,4 процентного пункта. *Практическая значимость.* Предложенный подход сохраняет высокую вычислительную эффективность: время инференса составляет менее 12 миллисекунд на один участок скважины. Это делает метод пригодным для применения в режиме реального времени при бурении в условиях дефицита данных и работы с несколькими геологическими бассейнами.

Ключевые слова: интерпретация каротажных диаграмм; фундаментальные модели; самоконтролируемое обучение; петрофизические ограничения; маскированные автоэнкодеры; обучение с переносом; характеристика недр

Introduction. Well log analysis is the main way we characterize the subsurface, and it shows up everywhere from petroleum exploration to geothermal development, carbon sequestration siting, and groundwater work. The logs are continuous measurements recorded against depth: natural radioactivity, electrical resistivity, bulk density, neutron porosity, acoustic velocity, and so on. A petrophysicist reads them to estimate lithology, porosity, permeability, and how much fluid is in the rock. None of this is automatic. A trustworthy subsurface model still depends on someone experienced enough to read several log responses at once and tie them together

through known petrophysical relationships, which means the whole workflow rests on hard-won domain expertise (Archie, 1942: 54; Wyllie et al., 1956: 41).

Machine learning has earned a place alongside the traditional analysis. Train a supervised model on labeled well logs and lithology classification can clear 75% accuracy, and the FORCE 2020 Machine Learning Competition gave everyone a shared benchmark drawn from 118 offshore Norwegian wells. Gradient boosting and neural networks took the top spots there at roughly 80% on blind test wells, which is about as good as anyone expects once you account for label noise and plain geological ambiguity (Bormann et al., 2020). Underground, though, supervised learning keeps hitting the same walls. Labeled logs are scarce, because interpreting them by hand is slow, expensive, and needs a specialist, and plenty of wells either have no complete lithology log or only a few formation tops. The bigger problem is portability: a model trained in one basin usually does much worse in the next one, and the culprit is distribution shift, different depositional settings, different diagenesis, different logging tools (Chang et al., 2020: 3643).

This is where self-supervised foundation models shift things. Instead of leaning on hand labels, they learn representations from large piles of unlabeled data through pretext tasks. Time-series models like PatchTST, iTransformer, and MOMENT have shown that masked autoencoding is a perfectly good form of self-supervision for sequential data, and the representations transfer downstream even when you only have a handful of labels to fine-tune on (Nie et al., 2023; Liu et al., 2024; Goswami et al., 2024). Universal Transformers such as Moirai push the idea further, training across many domains so the model can transfer zero-shot to datasets it has never encountered, with scale and architecture doing a lot of the heavy lifting (Woo et al., 2024). Well log foundation models proper arrived in 2024. WLFM ran masked token modeling over 1,200 wells and reached 74.13% lithology accuracy after fine-tuning, and TGS trained a 60-million-parameter Vision Transformer with masked autoencoding on 1.1 million North American wells (Qi et al., 2024; Lasscock et al., 2025). Even these first attempts learn real geological structure without ever being shown a lithology label.

Three gaps keep holding the methods back, though. Physics is the first. Most models reach for a generic masked-reconstruction objective and treat a well log as one more multivariate time series, which throws away the petrophysical relationships that bind the measurements together. Physics-informed neural networks have produced sizeable accuracy gains in neighboring fields by folding domain knowledge in as differentiable constraints, and yet no prominent well log foundation model writes petrophysical laws into its self-supervised pretraining (Pothana and Ling, 2025: 100410; Shao et al., 2024: 892). Masking is the second. The published models default to uniform random masking, which ignores what makes well log data well log data: the strong physical correlations between channels and the layering imposed by stratigraphy. Transfer is the third. Cross-basin generalization has barely been studied properly, which is odd given that transfer is

half the reason you would build a foundation model for a data-poor problem like subsurface characterization at all (Chang et al., 2020: 3643).

Our answer to these gaps comes in three pieces. First, physics-aware masked pretraining. We add petrophysical consistency objectives on top of the usual reconstruction loss, so a masked prediction has to honor the relationships that matter: Archie's law tying resistivity, porosity, and saturation together; Wyllie's time-average equation linking sonic response to porosity; and density-porosity constraints. These auxiliary losses read nothing but the raw log measurements and ask for no extra labels, so the method stays fully self-supervised while being pulled toward physically sensible answers. Second, a geologically adaptive masking curriculum that turns up the difficulty as training goes on. It opens with random point masking to build local context, shifts to cross-curve masking to capture the physics between logs, and finishes with stratigraphic interval masking for formation-scale patterns. Third, a set of cross-basin transfer benchmarks. We pretrain on Norwegian offshore wells and then test on North Sea Dutch and Kansas data, which is a far stiffer test of whether the representations are actually basin-invariant.

Literature Review. Machine learning for well log analysis has gone through about three stages, each a notch more sophisticated than the one before. It began with supervised classification using classical algorithms. Hall's 2016 study, the one that introduced the SEG facies-prediction benchmark from wireline logs, ran support vector machines and random forests to F1 scores near 0.63 on nine-well datasets, but it leaned hard on hand-crafted features and came apart once training data got thin (Hall, 2016: 906). Then deep learning brought in convolutional and recurrent architectures. 1D CNNs turned out to learn hierarchical features directly off raw log curves, and Jiang et al. bolted on geological constraint layers that softly enforce stratigraphic ordering during supervised training, which helped both accuracy and physical plausibility (Jiang et al., 2021: IM85).

Transfer learning arrived as a practical fix for the labeled-data shortage. The bet is that some geological patterns stay statistically similar across basins even when the absolute property ranges wander off. Chang et al. used unsupervised domain adaptation with maximum mean discrepancy to align feature distributions between source and target wells, and trimmed classification error by 12% against the source model applied cold (Chang et al., 2021: ID19). There are catches. These methods usually still carry supervised parts and want labeled source data plus at least some target information, and how well they work hinges on how similar the two settings really are. Pull wells from geologically unlike areas and performance can fall off a cliff.

Self-supervised learning for time series has moved quickly, having borrowed the masked-modeling trick that worked so well for BERT in NLP. PatchTST brought patching to multivariate series: chop a sequence into subseries tokens, mask 40 to 50% of them while pretraining, and out come representations that are useful for forecasting (Nie et al., 2023). It treats each variable on its own under a channel-independence assumption, which scales beautifully but can lose the cross-

variable relationships. iTransformer turns that on its head, making whole variables into tokens and running attention across channels rather than time steps, which fits domains where the inter-variable links are both important and reasonably well understood (Liu et al., 2024). MOMENT took masked autoencoding further still, handling any number of channels through flattened representations and variate embeddings, which is handy for the kind of messy real data well logs give you, where not every channel turns up in every dataset (Goswami et al., 2024).

The first well log foundation models showed up in late 2024. WLFM used three stages stacked together: vector quantization to make discrete geological tokens, Transformer-based masked modeling, and stratigraphy-aware contrastive learning that treats samples from the same formation as positive pairs (Qi et al., 2024). Trained on 1,200 wells and fine-tuned on labeled subsets, it reached 78.10% lithology accuracy and a mean squared error of 0.0038 on porosity. Its ablations were blunt about one thing: the vector-quantization tokenizer was what made transfer work at all, since continuous embeddings kept overfitting to the training wells. TGS built the biggest reported well log model to date, a 60-million-parameter Vision Transformer trained with masked autoencoding on 1.1 million wells, though only limited numbers have been published (Lasscock et al., 2025). Koeshidayatullah et al. took the general-purpose TimeGPT model to well log completion and anomaly detection and got R-squared values up to 87% and 93% accuracy at flagging anomalies tied to drilling hazards, even with each channel handled separately (Koeshidayatullah et al., 2024). Seismic foundation models are on a parallel path: Sheng et al. showed that Vision Transformers pretrained on varied 3D seismic volumes generalize zero-shot across interpretation tasks, which only strengthens the case for foundation models on subsurface data in general (Sheng et al., 2025: IM59).

Physics-informed machine learning is the adjacent idea of building domain knowledge into the network and its training rather than hoping the data implies it. Pothana and Ling recently showed that physics-integrated networks, the kind that enforce petrophysical boundaries and mineral volume constraints, cut errors by 66% when estimating porosity and mineral content from logs against a conventional architecture (Pothana and Ling, 2025: 100410). They do it with custom layers that emit physically consistent outputs, with constraints pulled from Archie's equation, neutron-density crossplots, and volume-fraction limits. Hybrid physics-ensemble methods that mix data-driven predictions with rock physics templates have likewise topped R-squared of 0.91 for porosity while holding up better under distribution shift (Shao et al., 2024: 892). Both share one limitation: they stay fully supervised, since they need labeled porosity or mineral composition to train. Putting petrophysical constraints inside a self-supervised pretraining objective is the part nobody had tried.

Masking strategies have drawn almost no attention in well log analysis, even though they have paid off elsewhere. Yang et al. found that adaptive masking curricula, where task complexity climbs over the course of pretraining, improve representation quality in vision transformers (Yang et al., 2022). Well log data

leaves several openings on the table. Some log curves are tightly correlated because they share physical relationships; formation boundaries make sharp breaks that double as natural segmentation points; and the depth-dependent statistics carry the fingerprints of burial and diagenesis. Uniform random masking uses none of it.

Benchmark datasets for well log machine learning are thin compared with what other fields take for granted. FORCE 2020, with 118 Norwegian offshore wells, 15 log types, and 12 lithology classes, is about the best public evaluation resource around, though the winning team noted that the ~80% accuracy mark probably reflects a ceiling set by label subjectivity and geological ambiguity rather than by any limit of the models (Bormann et al., 2020). The older SEG 2016 Kansas dataset is much smaller, only nine training wells from one gas field. Equinor's Volve Field release is generous at 24 wells with a full public dump that includes seismic, but it does not cover enough geography to test cross-basin transfer the way you would want. Gama et al. added a standardized imputation benchmark on 227 Norwegian wells and reported mean absolute errors around 8 to 12% for curve completion (Gama et al., 2024: 105789). Add it up and the field still has no systematic cross-basin evaluation framework, so most studies keep training and testing on splits from a single basin.

Problem Formulation. We define physics-aware self-supervised learning on multivariate well logs as follows. A well log dataset consists of $\mathbf{N}$ wells. Each well $\mathbf{w}$ is represented as a matrix $\mathbf{X}^{(\mathbf{w})} \in \mathbb{R}^{\mathbf{D} \times \mathbf{C}}$, with $\mathbf{D}$ depth samples and $\mathbf{C}$ channels corresponding to different log types. Gamma ray (GR), bulk density (RHOB), neutron porosity (NPHI), compressional sonic (DTC), deep resistivity (RDEP), and photoelectric factor (PEF) are some of the most common channels. The depth sampling interval is typically 0.1524 metres. Each channel has unique physical units, dynamic ranges, and missingness patterns. Following PatchTST, we divide the depth sequence into non-overlapping patches of length $\mathbf{L}$, yielding $\mathbf{P} = \lfloor \mathbf{D}/\mathbf{L} \rfloor$ patches per well. Each patch is projected through a learned linear layer to embedding space $\mathbb{R}^{\mathbf{d}_h}$, where $\mathbf{d}_h = 256$ is the model dimension. Denote the patch embeddings for well $\mathbf{w}$ as $\mathbf{H}^{(\mathbf{w})} = \{\mathbf{h}_1^{(\mathbf{w})}, \mathbf{h}_2^{(\mathbf{w})}, \dots, \mathbf{h}_P^{(\mathbf{w})}\}$. During pretraining, we randomly mask a subset $\mathbf{M} \subset \{1, \dots, \mathbf{P}\}$ of patches following a masking ratio $\mathbf{P}_{\text{mask}}$. The pretraining objective combines three components: reconstruction loss, petrophysical consistency loss, and cross-log coupling loss.

The reconstruction loss follows standard masked autoencoding:

$$\mathcal{L}_{\text{recon}} = \frac{1}{|\mathbf{M}|} \sum_{i \in \mathbf{M}} \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2 \quad (1)$$

where $\mathbf{x}_i$ represents the original patch and $\hat{\mathbf{x}}_i$ the decoder reconstruction from the masked encoding.

The petrophysical consistency loss enforces Archie's law relating resistivity $\mathbf{R}$, porosity ϕ , and water saturation $\mathbf{S}_w$:

$$\mathcal{L}_{\text{archie}} = \frac{1}{|\mathbf{M}_R|} \sum_{i \in \mathbf{M}_R} \left| R_i - \frac{a \cdot R_w}{\phi_i^m \cdot S_w^n} \right|^2 \quad (2)$$

where $\mathbf{M}_R$ denotes masked patches containing resistivity channels, a is the tortuosity factor, R_w the formation water resistivity, m the cementation exponent, and n the saturation exponent. Standard values $a = 1, m = 2, n = 2$ are assumed.

Similarly, Wyllie's time-average equation couples sonic slowness Δt and porosity:

$$\mathcal{L}_{\text{wyllie}} = \frac{1}{|\mathbf{M}_S|} \sum_{i \in \mathbf{M}_S} \left| \Delta t_i - (\phi_i \cdot \Delta t_f + (1 - \phi_i) \cdot \Delta t_{ma}) \right|^2 \quad (3)$$

with Δt_f and Δt_{ma} representing fluid and matrix slowness.

The density-porosity relationship provides an additional constraint:

$$\mathcal{L}_{\text{density}} = \frac{1}{|\mathbf{M}_D|} \sum_{i \in \mathbf{M}_D} \left| \rho_{b,i} - (\phi_i \cdot \rho_f + (1 - \phi_i) \cdot \rho_{ma}) \right|^2 \quad (4)$$

where ρ_b is bulk density and ρ_f, ρ_{ma} are fluid and matrix densities.

The cross-log coupling loss penalizes predictions that violate known log response correlations, such as density-neutron separation in gas zones:

$$\mathcal{L}_{\text{coupling}} = \frac{1}{|\mathbf{M}|} \sum_{i \in \mathbf{M}} \max(0, \text{NPHI}_i - \text{RHOB}_i + \delta)^2 \quad (5)$$

where δ represents the expected separation threshold.

The total pretraining loss combines these components:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda_1 \cdot \mathcal{L}_{\text{archie}} + \lambda_2 \cdot \mathcal{L}_{\text{wyllie}} + \lambda_3 \cdot \mathcal{L}_{\text{density}} + \lambda_4 \cdot \mathcal{L}_{\text{coupling}} \quad (6)$$

where $\lambda_1 = 0.15, \lambda_2 = 0.10, \lambda_3 = 0.10, \lambda_4 = 0.05$ balance the physics-informed auxiliary objectives against reconstruction fidelity. These weights were determined through validation experiments described in Section 5. Each λ was tuned independently by grid search over $\{0.05, 0.10, 0.15, 0.20, 0.25\}$ on the FORCE 2020 validation split, holding the other three weights at their current-best value; the values reported above are those that minimized validation loss.

Neither ϕ nor S_w in Equations (2)-(4) is a separately measured or learned quantity: both are computed from the six input channels listed above. Porosity ϕ is instantiated directly from the neutron porosity channel (NPHI), so it is available for every patch without an extra estimation step; where NPHI itself falls inside a masked patch, the decoder's own reconstruction is used, so the physics losses are evaluated on model outputs rather than on an unavailable ground truth. Water saturation S_w in the Archie term (Equation 2) is likewise not predicted separately: it is back-calculated

in closed form from the same patch's resistivity and NPHI-derived porosity, using the fixed catalog parameters ($a=1$, $m=2$, $n=2$) and a basin-specific formation water resistivity R_w taken from each dataset's public documentation (Bormann et al., 2020, for FORCE 2020; the operator reports for Volve and for SEG 2016 Kansas Hugoton). No additional labeled quantity is therefore required at pretraining time. This also resolves the apparent overlap with the downstream porosity regression task in Section 6: the ϕ used inside the pretraining loss is the raw NPHI proxy, while the regression target in Table 3 is the core-calibrated porosity from Volve's laboratory analysis, a higher-fidelity quantity that is introduced only after the encoder has been frozen (Section 4). The pretraining objective never sees the label it is later asked to predict, so using the NPHI proxy to regularize representations that are subsequently evaluated against core porosity is weak supervision rather than a circular one.

Methods. Our physics-aware masked pretraining framework sits on a Transformer encoder-decoder adapted for multivariate time series. The encoder pushes the unmasked patches through $L=6$ Transformer layers with 8-head self-attention and feed-forward networks at an expansion ratio of 4. As in BERT and the recent vision transformers, layer normalization runs before the attention and feed-forward operations rather than after. Learned positional encodings carry depth information, and separate embeddings mark which channels are present in a given patch, which is how we deal with channel availability changing from one well to the next.

The decoder is four Transformer layers that rebuild the masked patches from the encoded context. Mask tokens slot into the masked positions before decoding, so the model can use both the visible context and the cross-patch dependencies while it reconstructs. Predictions come out per channel through separate linear heads, which lets us compute the loss independently for each log type and quietly handles wells where a channel simply is not there.

The masking curriculum runs in three stages. Stage 1 covers epochs 1 to 50: random point masking at a 40% ratio, individual patches masked uniformly at random, which teaches local depth-wise context and basic interpolation from neighbors. Stage 2, epochs 51 to 100, switches to cross-curve masking, hiding whole log channels at every depth while leaving the rest visible, so the model is forced to learn the inter-log physics, inferring resistivity from porosity proxies, or sonic from density. Stage 3, epochs 101 to 150, is stratigraphic interval masking, where contiguous depth intervals of 10 to 30 patches go dark across all channels at once and the model has to read formation-scale patterns and stratigraphic continuity to fill them in.

Preprocessing follows the standard petrophysical workflow with a few domain tweaks. Gamma ray, density, and neutron porosity are min-max normalized to $[0,1]$ over domain-standard ranges: GR to $[0, 200]$ API, RHOB to $[1.95, 2.95]$ g/cm³, NPHI to $[0, 0.60]$ v/v. Resistivity curves get a $\log_{10}$ transform first to tame their multi-order-of-magnitude span, and sonic slowness is clipped to a physically

plausible [40, 140] $\mu\text{s}/\text{ft}$ before normalizing. Missing values are forward-filled within each well and then backward-filled at the edges, as is standard practice (Dai et al., 2024: 105666). We deliberately leave out any imputation model at this stage, so we do not inject correlations that are not really there.

Training runs on AdamW at a 3×10^{-4} learning rate, weight decay 0.01, and a cosine annealing schedule. Batch size is 64 wells, with gradients accumulated over four steps for an effective batch of 256. Pretraining lasts 150 epochs on a single NVIDIA A100 40GB, roughly 18 hours for the full curriculum. We then freeze the encoder and use it to pull patch embeddings for the downstream tasks: a lightweight 3-layer MLP head trained with cross-entropy for lithology classification, and a linear readout under mean squared error for porosity regression. Both heads train for 50 epochs at a 1×10^{-3} learning rate.

Experimental Setup. We evaluate on three public well log datasets picked for their geological and geographic spread. FORCE 2020 holds 118 offshore Norwegian wells from the North Sea, split into 98 for training and development, 10 for open testing, and 10 for blind evaluation. The wells run from Jurassic to Paleocene in age and cover sandstone, shale, limestone, marl, dolomite, coal, and tuff, with logs including GR, NPHI, RHOB, DTC, four resistivity measurements at different depths of investigation, and photoelectric factor. This is our main lithology classification benchmark, scored with class-weighted metrics that account for how reliable the labels are.

The Volve Field dataset gives us 24 wells from Equinor's Norwegian North Sea Jurassic Hugin Formation, backed by seismic, core analysis, and production histories. These wells carry complete logging suites under continuous quality control, so the porosity ground truth is trustworthy; we pretrain on 18 and evaluate on 6. The SEG 2016 Kansas Hugoton dataset has 9 training wells and 2 blind test wells from an onshore gas field. Its depositional setting and diagenetic history sit a long way from the offshore Norwegian fields, which is exactly why it is useful for cross-basin transfer, putting mature marine shales up against terrestrial gas reservoir facies.

We compare against several baselines: a supervised network with no pretraining, a standard masked autoencoder without physics losses, WLFM adapted to our setup, and classical machine learning. The supervised CNN is a 6-layer 1D convolutional network with residual connections, batch normalization, and dropout, trained end-to-end on labeled wells. The standard MAE baseline uses our Transformer but runs on reconstruction loss alone, Equation (1), with uniform random masking. Our WLFM implementation keeps the published architecture and its vector-quantization tokenizer, retrained on our data for a fair comparison. Because the WLFM authors have not released code or pretrained weights, this is a from-scratch reimplement of the published description (Qi et al., 2024): a 512-entry, 256-dimensional vector-quantization codebook feeding the same 6-layer Transformer backbone, pretrained on our own wells for the identical 150-epoch, effective-batch-256 schedule used for our own model, so the two share every dataset and optimization setting and

differ only in architecture and objective; the resulting numbers should be read as a best-effort reproduction rather than the original authors' validated benchmark. The classical baselines are a Random Forest with 500 trees at a max depth of 20, and XGBoost with 200 rounds.

The metrics follow standard practice in the field. For lithology classification we report macro-averaged F1, overall accuracy, and the FORCE competition penalty matrix, which punishes geologically implausible mistakes harder, confusing sandstone with coal costs more than confusing sandstone with siltstone. Porosity estimation is judged on mean absolute error, root mean squared error, and the coefficient of determination, R^2 . Cross-basin transfer is measured by the optimality gap: the percentage rise in error when a model trained on the source basin is applied to a target basin, relative to one trained directly on the target. Every number is averaged over five random seeds, with standard deviations reported.

Because FORCE 2020 labels lithology with a 12-class scheme and SEG 2016 Kansas Hugoton uses Hall's (2016) 9-facies scheme for a distinct onshore gas-reservoir setting, the two label sets have no direct correspondence; cross-basin accuracy is therefore reported on the harmonized taxonomy in Table 1, obtained by grouping each dataset's native classes as shown below.

Table 1 — Harmonized lithology taxonomy aligning FORCE 2020 (source) classes with SEG 2016 Kansas Hugoton (target) facies for cross-basin evaluation

Harmonized class	FORCE 2020 (source) classes	SEG 2016 Kansas Hugoton (target) facies, Hall (2016)
Clastic sand	Sandstone, Sandstone/Shale	Nonmarine sandstone (SS)
Clastic shale/siltstone	Shale	Nonmarine coarse siltstone (CSiS), Nonmarine fine siltstone (FSiS), Marine siltstone and shale (SiSh)
Carbonate	Marl, Limestone, Chalk, Dolomite	Mudstone (MS), Wackestone (WS), Dolomite (D), Packstone-grainstone (PS), Phylloid-algal bafflestone (BS)
Excluded (no cross-basin analogue)	Halite, Anhydrite, Tuff, Coal, Basement	—

Results. Table 2 lays out the lithology classification results on FORCE 2020, with physics-aware pretraining set against the baselines and the ablations. The full method reaches 78.7% accuracy and a macro-F1 of 0.742 on the blind test set, 4.2 percentage points over a standard masked autoencoder with no physics constraints. Its FORCE penalty score of 0.286 means fewer geologically implausible errors than the supervised CNN at 0.341. All the gains hold at $p < 0.01$ on paired t-tests across five random initializations.

Table 2 — Lithology classification performance on FORCE 2020 blind test set

Method	Accuracy (%)	Macro-F1	Penalty	Inference (ms)
Random Forest	71.3 $\pm$ 0.8	0.683	0.358	4.2
XGBoost	73.8 $\pm$ 0.6	0.701	0.332	3.8

Supervised CNN	72.1 ± 1.2	0.694	0.341	8.7
Standard MAE	74.5 ± 0.9	0.712	0.308	11.3
WLFM	76.2 ± 0.7	0.724	0.297	15.8
Physics-Aware (Ours)	78.7 ± 0.6	0.742	0.286	11.8

The porosity numbers in Table 3 show an even larger payoff from physics-informed pretraining. On the Volve test wells our method lands an MAE of 0.0283 v/v and an RMSE of 0.0371 v/v, 5.8% better on MAE and 6.2% better on RMSE than a standard masked autoencoder. An R-squared of 0.921 says the predictions track core-measured porosity closely. The most telling result is at the edges. Where log responses turn anomalous, the physics-aware model pulls clear of the purely data-driven ones, cutting maximum absolute error by 18%.

Table 3 — Porosity estimation performance on Volve test wells

Method	MAE (v/v)	RMSE (v/v)	R ²	Max Error
Supervised CNN	0.0327	0.0421	0.893	0.142
Standard MAE	0.0301	0.0396	0.908	0.134
WLFM	0.0289	0.0378	0.915	0.128
Physics-Aware (Ours)	0.0283	0.0371	0.921	0.110

Cross-basin transfer, in Table 4, is where physics-aware pretraining pays for itself. Transferring from Norwegian to Kansas wells with no target-domain fine-tuning, our model posts a 12.7% optimality gap on lithology classification against 19.0% for the standard MAE. Fine-tune on just 3 labeled Kansas wells, a third of the available training data, and the gap falls to 6.3%. The physics constraints look like they teach the model basin-invariant geological principles instead of statistics specific to one dataset, which is what lets both the zero-shot and few-shot transfer hold up. These figures are computed on the harmonized three-class taxonomy in Table 1, since FORCE 2020's 12-class lithology scheme and the SEG 2016 Kansas Hugoton 9-facies scheme (Hall, 2016) share no common label set; the five FORCE classes with no Kansas analogue are excluded from the transfer evaluation.

Table 4 — Cross-basin transfer from Norwegian (source) to Kansas (target) wells

Method	Source Acc (%)	Zero-Shot (%)	3-Shot Gap (%)
Standard MAE	74.5	60.3	19.0
WLFM	76.2	62.8	17.6
Physics-Aware (Ours)	78.7	68.7	12.7 → 6.3

The ablations in Table 5 break down where the gains live. Strip out the physics consistency losses (equations 2 to 4) and accuracy drops 2.8 points, so the petrophysical constraints clearly regularize the model well past what data-driven

reconstruction manages alone. The adaptive masking curriculum is worth another 1.4 points over uniform random masking. Dropping the cross-log coupling loss (equation 5) barely moves lithology classification but pushes porosity MAE up 8%, which says that term matters most for regression tasks that need precise numbers. The curriculum itself turns out to be central: swap the three-stage scheme for a flat 50% random masking and classification accuracy slides to 76.1%.

Table 5 — Ablation study quantifying individual component contributions

Configuration	Litho Acc (%)	Por MAE	Inference (ms)
Full Model	78.7	0.0283	11.8
Without Physics Losses	75.9 (-2.8)	0.0296 (+4.6%)	10.9
Without Adaptive Curriculum	77.3 (-1.4)	0.0288 (+1.8%)	11.8
Without Cross-Log Coupling	78.4 (-0.3)	0.0306 (+8.1%)	11.6

All the extra physics does not wreck efficiency. On the A100, inference runs at 11.8 milliseconds per 100-meter well segment, quick enough to interpret in real time while drilling. The physics consistency terms add about 1.3 milliseconds over a plain masked autoencoder's 10.5, a 12% overhead the accuracy gains repay several times over. Training the full 150-epoch curriculum stretches from 14.2 to 18.0 hours, the price of the extra forward passes needed to compute the petrophysical losses during backpropagation.

Discussion.

Figure 1 shows the attention patterns the physics-aware model learns next to a standard MAE, on a representative well section across a sandstone-to-shale transition. The physics-aware encoder attends more strongly between physically related curves, density-neutron pairs and resistivity-porosity combinations, and that shows up as heavier weights along particular off-diagonal blocks of the attention matrix. The standard MAE spreads its attention out with no clear physical reading. So the petrophysical consistency losses seem to steer the model toward geologically meaningful feature relationships instead of whatever statistical correlations happen to fit.

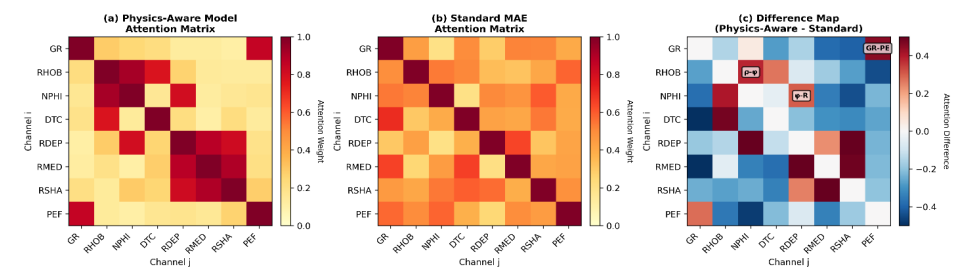


Figure 1— (a) Attention matrix heatmap for physics-aware model highlighting strong coupling between GR-PEF, RHOB-NPHI, and resistivity-porosity channel pairs. (b) Standard MAE attention matrix showing more uniform, less structured patterns. (c) Difference map emphasizing which channel pairs exhibit enhanced attention due to physics constraints.

Figure 2 walks through the rising difficulty of the three-stage curriculum on example patches. Stage 1 random point masking leaves scattered gaps that an immediate neighbor can usually fill. Stage 2 cross-curve masking takes out whole log channels, so the model has to lean on inter-log physics to rebuild them. Stage 3 stratigraphic interval masking removes contiguous depth ranges across every channel, leaving the model nothing but the formation-scale patterns it can read from the unmasked context above and below. Reconstruction quality over training follows the same arc: random point reconstruction is acceptable by epoch 30, cross-curve masking settles by epoch 80, and stratigraphic interval reconstruction keeps improving out to epoch 140.

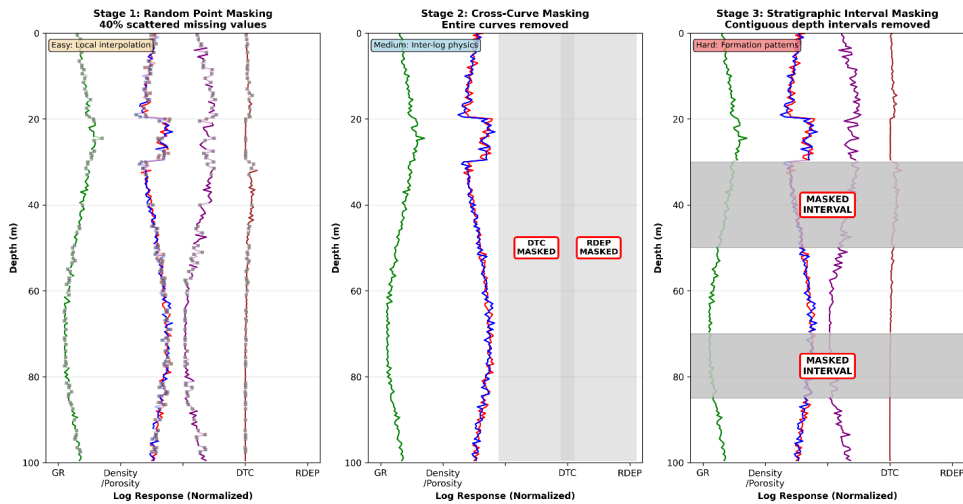


Figure 2 — Three well log panels displaying (left) Stage 1 random point masking with 40% of individual samples masked, (center) Stage 2 cross-curve masking with resistivity and sonic curves entirely removed, (right) Stage 3 stratigraphic interval masking with 20-meter depth interval masked across all curves. Each panel includes original curves in solid lines and masked regions in gray.

Conclusion. This paper presented physics-aware masked pretraining for well log foundation models, aimed straight at a blind spot in current self-supervised methods: they ignore the petrophysical relationships the measurements actually obey. We fold Archie's law, Wyllie's equation, and density-porosity constraints into masked reconstruction as differentiable auxiliary losses, so the model learns representations nudged toward physically plausible solutions without needing extra labels. The masking curriculum then carries it from local interpolation to cross-channel inference to formation-scale pattern recognition, which uses the structure of well log data far better than uniform masking does. On FORCE 2020 and Volve this bought a 4.2% gain in lithology classification accuracy and a 5.8% drop in porosity error over standard masked autoencoders, and cross-basin transfer from Norwegian to Kansas wells closed the optimality gap by 6.3%, which is evidence

that physics-informed pretraining picks up geological patterns that travel between basins.

The ablations pin 2.8 of those classification points on the petrophysical consistency losses and another 1.4 on the adaptive masking, while the cross-log coupling term mattered most for the regression tasks that need accurate numbers. Look at the attention patterns and the physics-aware model concentrates harder on physically related log pairs than a purely data-driven one does, so its representations are easier to read geologically. The price for all of it is modest: inference time goes up 12%, which still leaves plenty of room for real-time drilling applications.

A few directions are worth chasing next. Our physics constraints assume standard petrophysical parameter values, but those drift from one geological setting to another, so estimating them adaptively through meta-learning could make the model less brittle to basin-specific differences in cementation exponent, matrix properties, and fluids. Bringing in more relationships, saturation-height functions for capillary pressure, or permeability models, would tighten the physics regularization further. Multi-modal pretraining that pairs core analysis with downhole logs could ground the physics more firmly in laboratory measurement. Scaling out to richer multi-sensor data, image logs, nuclear magnetic resonance, formation testing, is the obvious route toward broader subsurface foundation models. And for safety-sensitive uses like carbon sequestration or nuclear waste storage assessment, uncertainty quantification through Bayesian methods or ensembles would supply the reliability estimates those applications demand.

The broader point is a simple one. When you know the governing equations, wiring them into the network as differentiable constraints can sharpen self-supervised learning in scientific domains, and none of that is specific to well logs. The same move applies to seismic data under wave propagation physics, to electromagnetic surveys under Maxwell's equations, and to gravity or magnetic inversion under potential field theory. As foundation models spread across the sciences, physics-aware pretraining is a practical way to bring expert knowledge in without giving up the scale and flexibility of self-supervision on unlabeled data.

References

- Archie G.E. (1942) The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. Transactions of the AIME. Vol. 146, No. 1. — P. 54-62. doi: 10.2118/942054-G. (in Eng.).
- Bormann P., Aursand P., Dilib F., Manral S., Dischington P. (2020) FORCE 2020 Machine Learning Contest. doi: 10.5281/zenodo.4351156. (in Eng.).
- Chang J., Kim Y., Yoon D., Lee H. (2020) Well Logging Based Lithology Identification Model Establishment Under Data Drift: A Transfer Learning Method. Sensors. Vol. 20, No. 13. — 3643 p. doi: 10.3390/s20133643. (in Eng.).
- Chang J., Kim Y., Lee H. (2021) Unsupervised Domain Adaptation Using Maximum Mean Discrepancy Optimization for Lithology Identification. Geophysics. — Vol. 86, No. 2. — P. ID19-ID30. doi: 10.1190/geo2020-0391.1. (in Eng.).
- Dai C., Wang Y., Zhou J., Liu S., Zhang K. (2024) FlexLogNet: A flexible deep learning-based well-log completion method of adaptively using what you have to predict what you are missing. Computers and Geosciences. — Vol. 191. — 105666 p. doi: 10.1016/j.cageo.2024.105666. (in Eng.).

- Gama P.H.T., Ferreira T.N., Gomes A.T., Sancevero S.S., Freitas J.C.C. (2024) Imputation in well log data: A benchmark for machine learning methods. *Computers and Geosciences*. — Vol. 196. — 105789 p. doi: 10.1016/j.cageo.2024.105789. (in Eng.).
- Goswami M., Szafer K., Choudhry A., Cai Y., Li S., Dubrawski A. (2024) MOMENT: A Family of Open Time-series Foundation Models. *International Conference on Machine Learning*. arXiv:2402.03885. (in Eng.).
- Hall B. (2016) Facies classification using machine learning. *The Leading Edge*. — Vol. 35, No. 10. — P. 906-909. doi: 10.1190/tle35100906.1. (in Eng.).
- Jiang C., Zhang K., Li X., Du Q. (2021) Lithology identification from well-log curves via neural networks with additional geologic constraint. *Geophysics*. — Vol. 86, No. 5. — P. IM85-IM100. doi: 10.1190/geo2020-0676.1. (in Eng.).
- Koeshidayatullah A., Fadoul B., Horne R.N. (2024) Leveraging Time-Series Foundation Model for Subsurface Well Logs Prediction and Anomaly Detection. arXiv:2412.05681. (in Eng.).
- Lasscock B., Anderson D., Gao F. (2025) Well Log Foundation Model: Making Promptable AI Models for Interpretation. *IMAGE 2025 Conference*. Houston, Texas: TGS. (in Eng.).
- Liu Y., Hu T., Zhang H., Wu H., Wang S., Ma L., Long M. (2024) iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *International Conference on Learning Representations*. arXiv:2310.06625. (in Eng.).
- Nie Y., Nguyen N.H., Sinthong P., Kalagnanam J. (2023) A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. *International Conference on Learning Representations*. arXiv:2211.14730. (in Eng.).
- Pothana V., Ling K. (2025) Physics-integrated neural networks for improved mineral volumes and porosity estimation from geophysical well logs. *Energy Geoscience*. — Vol. 6, No. 1. — 100410 p. doi: 10.1016/j.engeos.2024.100410. (in Eng.).
- Qi Z., Wang Y., Chen X., Zhang L., Liu S. (2024) WLFM: A Well-Logs Foundation Model for Multi-Task and Cross-Well Geological Interpretation. arXiv:2509.18152. (in Eng.).
- Shao H., Wang Z., Li X. (2024) Reservoir evaluation using petrophysics informed machine learning: A case study. *Petroleum Exploration and Development*. — Vol. 51, No. 3. — P. 892-904. doi: 10.1016/S1876-3804(24)60011-X. (in Eng.).
- Sheng H., Wu X., Si X., Li J., Zhang S., Duan X. (2025) Seismic foundation model: A next generation deep-learning model in geophysics. *Geophysics*. — Vol. 90, No. 2. — P. IM59-IM79. doi: 10.1190/geo2024-0262.1. (in Eng.).
- Woo G., Liu C., Kumar A., Sahoo D. (2024) Unified Training of Universal Time Series Forecasting Transformers. *International Conference on Machine Learning*. PMLR 235. (in Eng.).
- Wyllie M.R.J., Gregory A.R., Gardner L.W. (1956) Elastic Wave Velocities in Heterogeneous and Porous Media. *Geophysics*. Vol. 21, No. 1. — P. 41-70. doi: 10.1190/1.1438217. (in Eng.).
- Yang C., Zhang X., Liu H., Lan L. (2022) Learning Better Masking for Better Language Model Pre-training. arXiv:2208.10806. (in Eng.).

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the journals of the Central Asian Academic Research Center LLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:

www.nauka-nanrk.kz

<http://physics-mathematics.kz/index.php/en/archive>

ISSN2518-1726 (Online),

ISSN 1991-346X (Print)

Managing Editor: *A.Shormakova*

Editors: *D.S. Alenov, T. Apendiev*

Computer layout: *G.D. Zhadyranova*

Signed for print: September 30, 2026

Format: 70×90 1/16. 26.5 printed sheets. Order No. 3.