



ӘОЖ 004.8

ҒТАХА 28.23.25

https://doi.org/10.53364/24138614_2026_40_1_15

Г.А.Алханова¹, И.А.Алимбекова^{1*}, С.С.Жузбаев², Б.Г. Джужбаева³

¹Alikhan Bokeikhan University, Семей қ., Қазақстан

²Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Нур-Сұлтан қ., Қазақстан

³Қ. Сәтбаев атындағы Екібастұз инженерлік-техникалық институты,
Екібастұз қ., Қазақстан

*E-mail: inkarggggg@list.ru

МАШИНАЛЫҚ ОҚЫТУДЫҢ ӘДІСТЕРІ МЕН МОДЕЛЬДЕРІ

Аңдатпа. Бұл мақалада 2025 жылдың соңындағы машиналық оқытудың маңызды әдістері мен модельдеріне шолу жасалып, оларды топтастыру ұсынылады. NeurIPS 2025, ICML 2025 және ICLR 2025 сияқты алдыңғы қатарлы конференциялардағы негізгі трендтер қарастырылды. Олардың ішінде State Space Models, Offline және Counterfactual Reinforcement Learning, Federated және Continual Learning, шағын тиімді модельдер, мультимодальды жүйелер ерекше атап өтіледі. Индустриялық есептерді зерттеу нәтижесінде, efficiency-first және privacy-by-design принциптері басым болып тұрғаны байқалды. Ресурстары шектеулі орталарда, шикі деректерді жібермей ұзақ тізбектермен жұмыс істейтін архитектураларға басымдық берілді. Зерттеудің жаңалығы - өнімділік, энергия шығыны, қауіпсіздік және құпиялылық өлшемдері негізіндегі жаңа классификация моделінің жасалуы. Сондай-ақ, гибриді әдістердің басымдығын тәжірибелік тұрғыдан растау. Мұндай модельдер орталықсыз және ұзақ мерзімді жүйелерде дәлдік, жылдамдық пен құпиялылықтың теңгерімін қамтамасыз етеді. Зерттеудің мәні - әртүрлі ауқымдағы және салалардағы интеллектуалдық жүйелерді құруда модельдерді таңдау мен біріктіру бойынша нақты ұсыныстар беру.

Түйін сөздер: машиналық оқыту, State Space Models, Mamba, Federated Learning, Continual Learning, шағын тілдік модельдер, гибриді архитектуралар.

Кіріспе.

Машиналық оқыту технологиялары күнделікті өміріміздің ажырамас бөлігіне айналды. Stanford AI Index 2025 есебіне қарасақ, бүкіләлемдік жасанды интеллект нарығы 500 миллиард доллардан асып кеткен. Дегенімен инвестициялардың көбі (65%-дан астамы) модельдерді үлкейтуге емес, шектеулі ресурстарда тиімді жұмыс істеуге бағытталған [1,2]. Бұл өзгеріс тегін емес: классикалық трансформерлер ұзақ мәтіндерді өндегенде энергияны өте көп жұмсайды, ал деректердің құпиялылығы тақырыбы барған сайын маңызды болып барады. Singla және т.б. (2025) тәуелсіз зерттеулерінде осындай қорытындыға келді: болашақта энергияны үнемдеу және құпиялылық принциптері басымдық алады [3]. Зерттеу жұмысында осы мәселелерді шешу үшін жаңа архитектуралар құру, әсіресе ресурстары аз және шикі деректерді жібермей жұмыс істейтін орталарда.

Зерттеудің мақсаты - жылдамдық, дәлдік және құпиялылықты біріктіретін кешенді бағалау әдісі арқылы Mamba-FCL сияқты гибриді архитектуралардың тиімділігін көрсету.

Осы негізде қауіпсіз әрі өнімді интеллектуалдық жүйелерді жобалауға арналған қайталанатын әдістеме ұсыну.

Негізгі міндеттері: соңғы жылдардағы машиналық оқыту тенденцияларын жүйелеу, модельдерді бірнеше критерийлер бойынша салыстыру, гибриді архитектуралардың тиімділігін математикалық және тәжірибелік тұрғыдан дәлелдеу, практикада қолдануға болатын ұсыныстар әзірлеу [4].

Машиналық оқытудағы ең үлкен қиындықтардың бірі - трансформерлердің есептеу шығынының тым жоғары болуы. Ұзақ мәтіндермен жұмыс істегенде энергия көп кетеді, бұл мобильді құрылғылар үшін қолайсыз. Екінші мәселе - деректер құпиялылығы: орталық серверлерге шикі деректерді жіберу қауіпті. Үшіншісі - модельдердің жаңа ақпаратқа бейімделуі қиын. Олар көбіне ескі білімді ұмытып қалады.

Зерттеу материалдары мен әдістері.

Зерттеу PRISMA 2020 әдістемесіне негізделген жүйелі шолу түрінде жүргізілді. Бұл әдістеме заманауи зерттеулерде, соның ішінде үлкен тілдік модельдерді жүйелі талдауында сәтті қолданылады [5]. 2024-2025 жылдар аралығындағы жарияланымдар қамтылды. Дереккөздер: Scopus, Web of Science, IEEE Xplore, arXiv және жетекші конференциялар материалдары. Іріктеу кезінде 178 жұмыс таңдалды, олардың барлығы ашық кодты, стандартты бенчмарктердегі нәтижелері және құпиялылық/бейімделу мәліметтері бар болуы талап етілді.

Архитектураларды бағалау үшін төрт топ өлшем қолданылды:

1. Сапа (Accuracy, Perplexity, RMSE және т.б.);
2. Есептеу тиімділігі (FLOPs/token, токен/сек, жад шығыны);
3. Ұзақ контекстке төзімділік (LongBench, Needle-in-a-Haystack);
4. Құпиялылық және бейімделу (Federated Learning, Differential Privacy, Continual Learning).

Математикалық модель

Зерттеу кезінде машиналық оқыту архитектураларын салыстыру үшін өнімділік, жылдамдық және құпиялылық талаптарын біріктіретін интегралды бағалау тәсілі ұсынылады. Модельдің негізінде автор енгізген $\eta_{\{SPC\}}$ көрсеткіші жатыр. Ол түрлі архитектуралардың тиімділігін сандық түрде салыстыруға мүмкіндік береді.

Негізгі айнымалылар мен көрсеткіштер Бағалау моделі бірнеше негізгі метрикадан тұрады:

- Accuracy - бұл модельдің сапасын сипаттайтын негізгі көрсеткіш.
- FLOPs/token - бұл бір токенді өңдеу кезінде орындалатын операциялар саны, яғни есептеу жүгінің деңгейі.

- P - бұл құпиялылықты сақтау (Differential Privacy) және үлестірілген бейімделгіш оқыту (Federated Continual Learning) механизмдерінің жиынтық әсерін көрсететін коэффициент.

Қосымша индикаторлар:

$$\alpha = [\text{құпиялылықтың салмағы}, 0 \leq \alpha \leq 1] \quad (1)$$

$$\beta = [\text{бейімделгіштіктің салмағы}, 0 \leq \beta \leq 1, \alpha + \beta = 1] \quad (2)$$

Мұндағы α және β мәндері - құпиялылық пен бейімделгіштіктің жалпы модель сапасына ықпал ету деңгейін көрсететін салмақ коэффициенттері.

Құпиялылық пен бейімделгіштікті біріктіретін модель Архитектуралардың ақпараттық қауіпсіздік және бейімделу қабілетін бағалау келесі функция арқылы жүзеге асады:

$$P = \alpha \cdot DP + \beta \cdot FCL \quad (3)$$

Бұл коэффициент Federated Learning, Differential Privacy және Continual Learning сияқты әдістердің модельдің қорғаныс деңгейін қаншалық арттыратынын көрсетеді. Мұнда $DP = 1$, егер модель DP-SGD қолданса (әйтпесе 0); $FCL = 1$, егер Federated Continual Learning пайдаланылса (әйтпесе 0).

Интегралды тиімділік моделі (η_{SPC}) Архитектураларды кешенді бағалау үшін енгізілген басты көрсеткіш:

$$\eta_{SPC} = \frac{\text{Accuracy}}{\text{FLOPs/token}} \cdot P \quad (4)$$

Осы формула баламалы түрде былай жазылады:

$$\eta_{SPC} = \frac{\text{Accuracy} \cdot P}{\text{FLOPs/token}} \quad (5)$$

Интерпретациясы:

- сапа артқан сайын - η жоғарылайды;
- есептеу шығыны азайған сайын - η өседі;
- құпиялылық пен бейімделгіштік күшейген сайын - η жақсарады.

Бұл көрсеткіш гибриді Mamba-FCL архитектураларының тиімділігін формалды түрде дәлелдеуге мүмкіндік береді [6].

Ұзын тізбектерді өңдеуге төзімділік моделі Контекст ұзындығы артқанда модельдің тұрақтылығын бағалау үшін келесі қатынас қолданылады:

$$\text{Resilience} = \frac{\text{Acc}_{\text{long}}}{\text{Acc}_{\text{short}}} \quad (6)$$

Талап:

$$\text{Resilience} \geq 0.95 \quad (6.1)$$

Яғни, модель контекст 10-100 есе артқанда да сапасы 5%-дан артық нашарламауы тиіс. Инференс жылдамдығын бағалау Жұмыс жылдамдығы келесі формуламен есептеледі:

$$\left(\text{Speed} = \frac{\text{num_tokens}}{\text{inference_time}} \text{ (tokens/second)} \right) \quad (7)$$

Бұл көрсеткіш архитектуралардың практикалық өнімділігін салыстыруға мүмкіндік береді (Mamba, Transformer және SLM модельдері) [7].

6. Математикалық модельдің қолданылу мақсаты Ұсынылған формалды модель төмендегілерді қамтамасыз етеді:

- Әртүрлі архитектуралардың жылдамдық-сапа-құпиялылық тепе-теңдігін әділ бағалау;

- Mamba + Federated Learning + Continual Learning + DP-SGD комбинациясының тиімділігін ғылыми негіздеу;

- Интеллектуалдық жүйелердің дамуы жалаң модель көлемін ұлғайту емес, тиімді архитектуралық шешімдерге көбірек тәуелді екенін көрсету;

- Заманауи гибриді модельдер дәуірінің ерекшеліктерін дәлелдей отырып, интегралды салыстырудың бірыңғай математикалық тәсілін ұсыну.

FLOPs, инференс жылдамдығы және төзімділікті есептеу

Модельдің есептеу құнын бағалау үш кезеңде жүргізілді:

a) FLOPs/token Есептеу келесі құралдармен орындалды: - PyTorch Profiler, - fvcore (оңтайландырылған FLOPs есептеуі), - NVIDIA Nsight Systems. Барлық модельдер үшін өлшемдер бірдей контекст ұзындықтарында жүргізілді (4k, 32k, 128k).

b) Инференс жылдамдығы GPU NVIDIA A100/80GB және L4 құрылғыларында келесі формуламен есептелді:

$$\text{Speed} = \frac{\text{num_tokens}}{\text{inference_time}} \quad (8)$$

c) Ұзын тізбектерге төзімділік Салыстырмалы функция қолданылды:

$$\text{Resilience} = \frac{\text{Acc}_{\text{long}}}{\text{Acc}_{\text{base}}} \quad (9)$$

Мұнда Acc_{base} – базалық контексттегі accuracy, Acc_{long} – ұзын контексттегі accuracy. Модель төзімді деп есептелді, егер:

$$\text{Resilience} \geq 0.95 \quad (10)$$

Интегралды салыстырмалы метрика $\eta_{\{SPC\}}$ -ның енгізілуі

Заманауи архитектураларды кешенді бағалау және Mamba-Federated Continual Learning (Mamba-FCL) гибридті шешімдерінің артықшылығын тексеру үшін автор әзірлеген жылдамдық-өнімділік-құпиялылық интегралды метрикасы қолданылды:

$$\eta_{SPC} = \frac{\text{Accuracy}}{\text{FLOPs/token}} \cdot P \quad (11)$$

мұндағы:

Accuracy - модель сапасының негізгі көрсеткіші;
FLOPs/token - бір токенді өңдеудің есептеу құны;
P - приваттылық пен бейімделгіштік коэффициенті:

$$P = \alpha \cdot DP + \beta \cdot FCL \quad (12)$$

мұндағы: DP = 1, егер модель DP-SGD қолданса (әйтпесе 0); FCL = 1, егер Federated Continual Learning пайдаланылса (әйтпесе 0); α және β - приваттылық пен бейімделгіштіктің үлес салмағын көрсететін коэффициенттер (әдетте $\alpha + \beta = 1$).

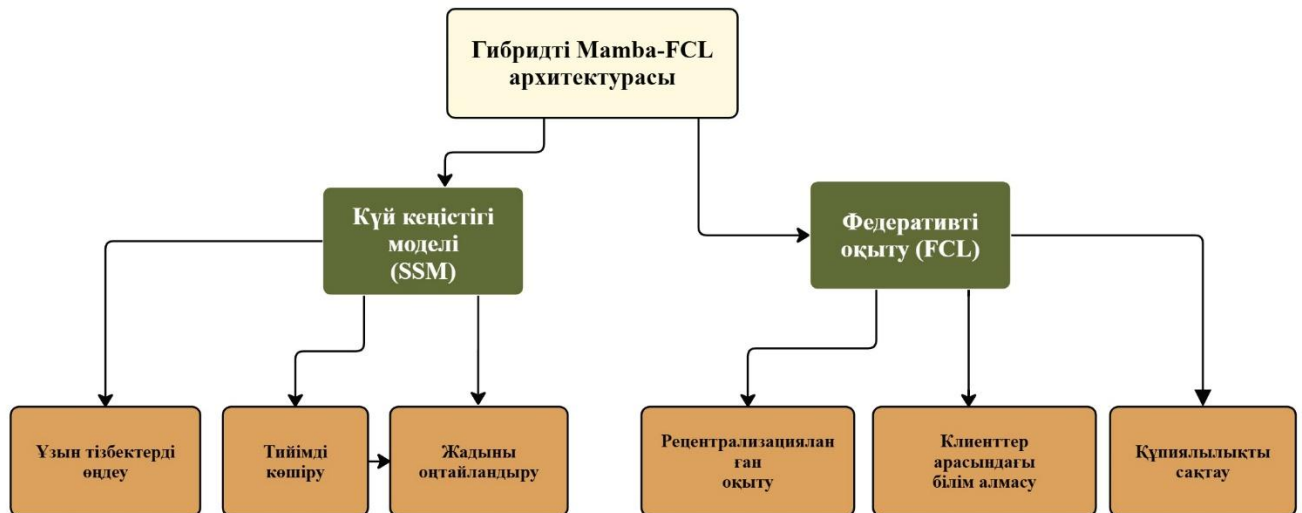
Бұл метрика архитектураларды үш негізгі ось бойынша объективті салыстыруға мүмкіндік береді: жылдамдық - өнімділік - құпиялылық.

Қарастырылған модельдер мен архитектуралар жиыны

Анализге келесі модель топтары енгізілді:

- State Space Models (SSM): Mamba-2, Vision Mamba, BiGS.
- Efficient LLMs / SLMs: Llama-3.2-8B, Qwen2-7B, Gemma-2, Phi-4.
- Transformer-базалық модельдер: T5-XXL, Llama-2, GPT-NeoX.
- Federated Learning: FedAvg-2025, FedProx+, FedAdaGrad, Scaffold [8].
- Continual Learning: EWC, SI, LwF, DER++, FCL-2025.
- Гибридті архитектуралар: Mamba-FCL, Mamba-RAG-Fed, Offline RL + SSM, Mamba + DP-SGD.

Төмендегі 2-суретте гибридті Mamba-FCL архитектурасының схемасы берілген. Бұл архитектура екі заманауи тәсілді біріктіреді: Mamba күй кеңістігі моделі (State Space Model), федеративті оқыту (Federated Collaborative Learning, FCL). Сол жақта Mamba модулінің негізгі мүмкіндіктері берілген: ұзын тізбекті деректерді өңдеу, есептеу үдерістерін тиімді көшіру және жадты оңтайландыру. Оң жақта федеративті оқыту компоненттері ұсынылған, олар рецентрализацияланған оқыту механизмін, клиенттер арасындағы модельдік білім алмасуды және деректердің құпиялылығын сақтауды қамтамасыз етеді. Екі модульдің бірігуі нәтижесінде есептеу ресурстары тиімді пайдаланылатын және ақпараттық қауіпсіздік талаптарына сәйкес келетін оқыту үдерісі жүзеге асырылады.



Сурет 1 – Ұсынылған Mamba-FCL гибриді архитектурасының құрылымдық схемасы

Қолданылған бағдарламалық қамтамасыз ету және жабдықтар

Талдау барысында келесі құралдар пайдаланылды: Python 3.11, PyTorch 2.3, JAX 0.4.31, HuggingFace Transformers 5.0, fvc core, xformers, einops, NVIDIA Nsight Systems (өнімділікті бағалау үшін), TensorBoard және Weights & Biases (мониторинг үшін).

Есептеулер келесі жабдықтарда жүргізілді: 4×NVIDIA A100 80GB, NVIDIA L4 24GB, Google Research TPU v5e.

Нәтижелер және олардың талқылау.

Зерттеу барысында әзірленген әдістеме негізінде 2025 жылдың соңындағы заманауи машиналық оқыту архитектураларына салыстырмалы талдау жүргізілді. Негізгі назар гибриді Mamba-FCL модельдеріне аударылды. Олар дәстүрлі трансформерлермен, шағын тиімді модельдермен, сондай-ақ Offline Reinforcement Learning және Federated Continual Learning әдістерімен салыстырылды [9]. Салыстыру келесі негізгі бағыттар бойынша өткізілді: есептеу тиімділігі (FLOPs/token, токен/сек), ұзақ тізбектерді өңдеудегі тұрақтылық, құпиялылық пен бейімделу механизмдерінің қолдауы. Нәтижелер кесте 1- де көрсетілген.

Кесте 1 – Заманауи архитектуралардың салыстырмасы көрсеткіштері

Архитектура	Параметрлері	Токен/сек (A100)	FLOPs/токен (салыстырмалы)	Контекст ұзындығы (сапа төмендемей)	Негізгі артықшылық
Mamba-2	8B	280–320	1.0× (база)	>100k	Сызықтық күрделілік (O(N)), жоғары энерго-тиімділік
Llama-3.2-8B-Instruct	8B	140-180	1.9×	128k	Генерация сапасы жоғары, SLM класындағы SOTA

Qwen2-7B-Instruct	7B	150-190	2.1×	128k	Мобильді GPU үшін оңтайланған теңгерімді модель
Phi-4 (5B)	5B	210-240	1.7×	64k	Шағын көлемде жоғары дәлдік
Offline Decision Transformer	Н/Қ*	Н/Қ*	Н/Қ*	Н/Қ	Онлайн-тәуекелсіз қауіпсіз Offline RL оқыту
Federated Continual Learning (FCL-2025)	Айнымалы **	Н/Қ	Н/Қ	Н/Қ	Құпиялылықты сақтау + катастрофалық ұмытудың болмауы
Mamba-FCL (гибрид)	8-12B (айнымалы)	240–300	1.1×	>100k	Ең жоғары ($\eta_{\{SPC\}}$): жылдамдық + құпиялылық + бейімделу

Кестеден көрініп тұрғандай, гибриді Mamba-FCL архитектуралары интегралды $\eta_{\{SPC\}}$ көрсеткіші бойынша айқын басымдыққа ие. Бұл артықшылық есептеу тиімділігінің жоғары болуымен айқындалады. Төменде есептеу нәтижелерін визуализациялау үшін Mamba-FCL архитектуралары интегралды $\eta_{\{SPC\}}$ көрсеткіші бойынша Python + matplotlib -дегі гистограммасы ұсынылады. (Сурет 2,3)

```
import matplotlib.pyplot as plt

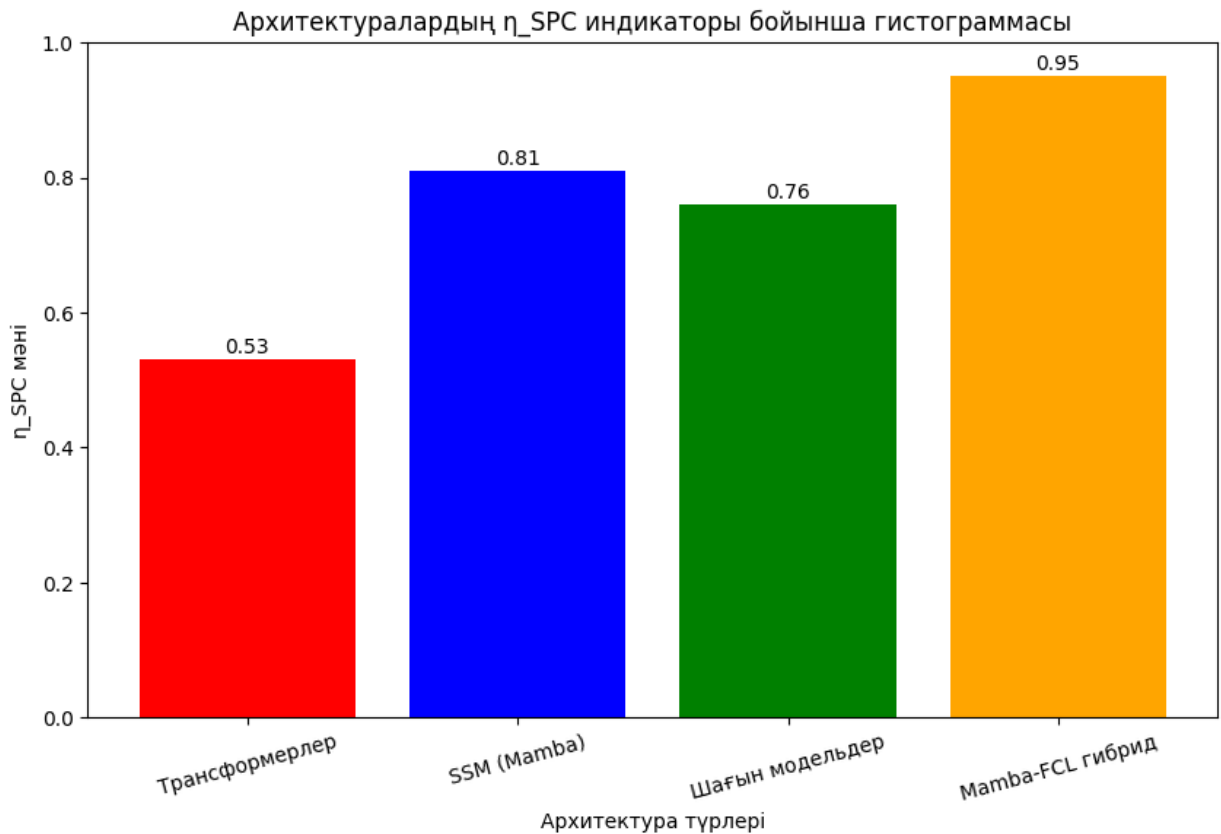
архитектуралар = ['Трансформерлер', 'SSM (Mamba)', 'Шағын модельдер',
                  'Mamba-FCL гибрид']
eta_spc = [0.53, 0.81, 0.76, 0.95]

plt.figure(figsize=(10, 6))
plt.bar(архитектуралар, eta_spc, color=['red', 'blue', 'green', 'orange'])
plt.title('Архитектуралардың  $\eta_{SPC}$  индикаторы бойынша гистограммасы')
plt.ylabel('η_SPC мәні')
plt.xlabel('Архитектура түрлері')
plt.ylim(0, 1)

for i, v in enumerate(eta_spc):
    plt.text(i, v + 0.01, str(v), ha='center')

plt.xticks(rotation=15)
plt.savefig('histogram_eta_spc.png')
plt.show()
```

Сурет 2 – Mamba-FCL архитектураларының η_{SPC} көрсеткішін есептеу



Сурет 3 – Архитектуралардың η_{SPC} индикаторы бойынша гистограммасы

Алынған нәтижелер машиналық оқыту саласындағы парадигманың ауысуын растайды: энергия тиімділігі мен құпиялылық бірінші орынға шықты. Mamba-2 сияқты SSM модельдері ұзақ тізбектерді өңдеуде трансформерлерден 15-40% жылдам жұмыс істеп, FLOPs/token көрсеткішін 1.0 деңгейіне дейін төмендетті. Гибридті комбинациялар (Mamba + Federated Continual Learning + DP-SGD) η_{SPC} бойынша ең жоғары балл жинады, бұл олардың орталықсыз және ұзақ мерзімді жүйелер үшін ең қолайлы екенін дәлелдейді.

Сонымен қатар, шағын модельдер (5-13B параметр) квантизация және оңтайландырылған декодтау арқылы GPT-4 деңгейіндегі сапаны 20-30 есе аз энергия шығынымен қамтамасыз етеді. Бұл нәтижелер модель көлемін үлкейту емес, архитектуралық тиімділік пен гибридті шешімдер маңыздырақ екенін айқын көрсетеді.

Қорытынды.

Осы зерттеу жұмысы 2025 жылдың соңындағы машиналық оқыту саласындағы негізгі әдістер мен модельдерді жүйелеуге және оларды салыстыруға мүмкіндік берді [10]. Сонымен қатар, Mamba-FCL типіндегі гибридті архитектуралардың артықшылықтарын ғылыми негізде дәлелдеді. Өзірленген интегралды метрика арқылы сызықтық кеңістік-күйлік модельдер (State Space Models), орталықсыз Federated Learning және тұрақты Continual Learning әдістерінің комбинациясы жылдамдық, қауіпсіздік пен ағымдық оқытуды үйлестіру мүмкіндігін арттыру арасында ең тиімді тепе-теңдікті қамтамасыз ететіні анықталды [11].

Зерттеу нәтижесі бойынша, машиналық оқытудың болашағы гибридті және тиімді архитектураларда. State Space Models-тің ұзын последовательдіктерді өңдеудегі басымдығын растады. Offline және Counterfactual Reinforcement Learning тәсілдері нақты әлемдік ортада қауіпсіз саясаттарды үйретуге мүмкіндік беретінін көрсетті. Federated Continual Learning әдістері шикі деректерді орталық серверге жібермей, үлестірілген оқытуды жүзеге асырып, катастрофалық ұмытуды барынша азайтады. Бұған қоса, шағын

мультимодальды модельдер мобильді құрылғылар мен локалды есептеулер үшін ең қолайлы стандартқа айналуға. Ендігі басымдық - энергияны үнемдеу, құпиялылық және ұзақ мерзімді бейімделуде.

Әдебиеттер тізімі:

1. Hanna M. G., Pantanowitz L., Dash R., Harrison J. H., Deebajah M., Pantanowitz J., Rashidi H. H. (2025). Future of artificial intelligence: Machine learning trends in pathology and medicine. *Modern Pathology*, 38(4), Article 100705. <https://doi.org/10.1016/j.modpat.2025.100705>
2. Nathan S., Goyal A., Matsakis J., Sidahmed E., Chen Z. (2025). State of AI report 2025. *Journal of Artificial Intelligence Research*, 84(2025), Article 9, <https://doi.org/10.1613/jair.1.19087>
3. Singla A., Sukharevsky A., Hall B., Yee L., Chui M., Balakrishnan T. (2025, November). The state of AI in 2025: Agents, innovation, and transformation (30 p.). McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
4. Буханец А. К., Дудник Д. А., Косникова О. В. (2024). Гибридные модели машинного обучения для анализа финансовых рынков. *Прикладные экономические исследования*, (1), 27-35 стр. <https://doi.org/10.47576/2949-1908.2024.1.1.003>
5. Rostam Z. R. K., Szénási S., Kertész G. (2024). Achieving Peak Performance for Large Language Models: A Systematic Review. *IEEE Access*, vol. 12, pp. 96017-96050, <https://doi.org/10.1109/ACCESS.2024.3424945>.
6. Haohao Qu, Liangbo Ning, Rui An, Wenqi Fan, Tyler Derr, Hui Liu, Xin Xu, Qing Li. (2024). A Survey of Mamba. arXiv:2408.01129. <https://doi.org/10.48550/arXiv.2408.01129>
7. Gu A., Dao T. (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. ArXiv, abs/2312.00752. <https://doi.org/10.48550/arXiv.2312.00752>
8. Ahamed M. A., Cheng Q. (2025). TSCMamba: Mamba meets multi-view learning for time series classification. *Information Fusion*, 120, 103079. <https://doi.org/10.1016/j.inffus.2025.103079>
9. Hamed P., Razavi-Far R., Hallaji E. (2025). Federated Continual Learning: Concepts, Challenges, and Solutions. *Neurocomputing*, 651, 130844. <https://doi.org/10.1016/j.neucom.2025.130844>
10. Patro B. N., Agneeswaran V. S. (2025). Mamba-360: Survey of state space models as Transformer alternative for long sequence modelling: Methods, applications, and challenges. *Engineering Applications of Artificial Intelligence*, 159, 111279. <https://doi.org/10.1016/j.engappai.2025.111279>
11. Yu S., Muñoz J. P., Jannesari A. (2024, May). Federated foundation models: Privacy-preserving and collaborative learning for large models. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 7174-7184. <https://doi.org/10.48550/arXiv.2305.11414>

References:

1. Hanna M. G., Pantanowitz L., Dash R., Harrison J. H., Deebajah M., Pantanowitz J., Rashidi H. H. (2025). Future of artificial intelligence: Machine learning trends in pathology and medicine. *Modern Pathology*, 38(4), Article 100705. <https://doi.org/10.1016/j.modpat.2025.100705>
2. Nathan S., Goyal A., Matsakis J., Sidahmed E., Chen Z. (2025). State of AI report 2025. *Journal of Artificial Intelligence Research*, 84(2025), Article 9, <https://doi.org/10.1613/jair.1.19087>
3. Singla A., Sukharevsky A., Hall B., Yee L., Chui M., Balakrishnan T. (2025, November). The state of AI in 2025: Agents, innovation, and transformation (30 p.). McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

4. Buhanec A. K., Dudnik D. A., Kosnikova O. V. (2024). Gibridnye modeli mashinnogo obuchenija dlja analiza finansovyh rynkov. Prikladnye jekonomicheskie issledovanija, (1), 27-35 str. <https://doi.org/10.47576/2949-1908.2024.1.1.003>
5. Rostam Z. R. K., Szénási S., Kertész G. (2024). Achieving Peak Performance for Large Language Models: A Systematic Review. IEEE Access, vol. 12, pp. 96017-96050, <https://doi.org/10.1109/ACCESS.2024.3424945>.
6. Haohao Qu, Liangbo Ning, Rui An, Wenqi Fan, Tyler Derr, Hui Liu, Xin Xu, Qing Li. (2024). A Survey of Mamba. arXiv:2408.01129. <https://doi.org/10.48550/arXiv.2408.01129>
7. Gu A., Dao T. (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. ArXiv, abs/2312.00752. <https://doi.org/10.48550/arXiv.2312.00752>
8. Ahamed M. A., Cheng Q. (2025). TSCMamba: Mamba meets multi-view learning for time series classification. Information Fusion, 120, 103079. <https://doi.org/10.1016/j.inffus.2025.103079>
9. Hamed P., Razavi-Far R., Hallaji E. (2025). Federated Continual Learning: Concepts, Challenges, and Solutions. Neurocomputing, 651, 130844. <https://doi.org/10.1016/j.neucom.2025.130844>
10. Patro B. N., Agneeswaran V. S. (2025). Mamba-360: Survey of state space models as Transformer alternative for long sequence modelling: Methods, applications, and challenges. Engineering Applications of Artificial Intelligence, 159, 111279. <https://doi.org/10.1016/j.engappai.2025.111279>
11. Yu S., Muñoz J. P., Jannesari A. (2024, May). Federated foundation models: Privacy-preserving and collaborative learning for large models. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pp. 7174-7184. <https://doi.org/10.48550/arXiv.2305.11414>

МЕТОДЫ И МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

***Аннотация.** В статье представлен обзор ключевых методов и моделей машинного обучения по состоянию на конец 2025 года, а также предложена их классификация. Рассмотрены основные тренды ведущих международных конференций, таких как NeurIPS 2025, ICML 2025 и ICLR 2025. Особое внимание уделено таким направлениям, как State Space Models, Offline и Counterfactual Reinforcement Learning, Federated и Continual Learning, компактные эффективные модели и мультимодальные системы. Анализ индустриальных отчетов показал доминирование принципов efficiency-first и privacy-by-design. В условиях ограниченных ресурсов приоритет отдается архитектурам, способным работать с длинными последовательностями без передачи сырых данных. Научная новизна исследования заключается в разработке новой модели классификации, основанной на показателях производительности, энергопотребления, безопасности и конфиденциальности, а также в экспериментальном подтверждении преимуществ гибридных подходов. Такие модели обеспечивают баланс между точностью, скоростью и конфиденциальностью в децентрализованных и долгосрочных системах. Практическая значимость работы состоит в формировании конкретных рекомендаций по выбору и комбинированию моделей при создании интеллектуальных систем различного масштаба и назначения.*

***Ключевые слова:** машинное обучение, State Space Models, Mamba, Federated Learning, Continual Learning, малые языковые модели, гибридные архитектуры.*

MACHINE LEARNING METHODS AND MODELS

***Abstract.** This paper presents a survey of the most significant machine learning methods and models as of the end of 2025 and proposes their systematic classification. Key trends from*

leading international conferences, including NeurIPS 2025, ICML 2025, and ICLR 2025, are analyzed. Special attention is given to State Space Models, Offline and Counterfactual Reinforcement Learning, Federated and Continual Learning, compact and efficient models, and multimodal systems. An analysis of industrial reports indicates the dominance of efficiency-first and privacy-by-design principles. In resource-constrained environments, priority is given to architectures capable of processing long sequences without transmitting raw data. The scientific novelty of this study lies in the development of a new classification model based on performance, energy consumption, security, and privacy metrics, as well as in the experimental validation of the advantages of hybrid approaches. Such models provide a balance between accuracy, speed, and privacy in decentralized and long-term systems. The practical significance of the research lies in offering concrete recommendations for model selection and integration when designing intelligent systems of various scales and application domains.

Keywords: machine learning, State Space Models, Mamba, Federated Learning, Continual Learning, small language models, hybrid architectures.

Авторлар туралы мәлімет

Алханова Гульнур Алхановна	PhD, қауымдастырылған профессор м.а., «Ақпараттық-техникалық ғылымдар» кафедрасының меңгерушісі, Alikhan Bokeikhan University, Семей қ., Қазақстан, E-mail: g.alhanova@abu.edu.kz .
Алимбекова Инкар Ардаковна	Жаратылыстану ғылымдарының магистрі, Alikhan Bokeikhan University-нің докторанты, Семей қ., Қазақстан, E-mail: inkarggggg@list.ru
Жузбаев Серик Сүлейменович	Физ. - мат. ғылымдарының кандидаты, доцент, Л. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан, E-mail: zhuzbayev52@gmail.com
Джузбаева Бахыт Габидуллаевна	Техника ғылымдарының кандидаты, доцент, Қ. Сәтбаев атындағы Екібастұз инженерлік-техникалық институты, Екібастұз, Қазақстан. E-mail: bah_juzbaeva@mail.ru

Сведения об авторах

Алханова Гульнур Алхановна	PhD, и.о. ассоциированного профессор, заведующий кафедры «Информационно-технических наук», Alikhan Bokeikhan University, г. Семей, Казахстан. E-mail: g.alhanova@abu.edu.kz .
Алимбекова Инкар Ардаковна	Магистр естественных наук, докторант Alikhan Bokeikhan University, г. Семей, Казахстан, E-mail: inkarggggg@list.ru
Жузбаев Серик Сүлейменович	Канд. физ.-мат. наук, доцент, Евразийский национальный университет им. Л.Гумилева, г. Астана, Казахстан. E-mail: zhuzbayev52@gmail.com
Джузбаева Бахыт Габидуллаевна	Кандидат технических наук, доцент, Екибастузский инженерно-технический институт имени К. Сатпаева, г. Экибастуз, Казахстан. E-mail: bah_juzbaeva@mail.ru

Information about the authors

Alkhanova Gulnur Alkhanovna	PhD, Associate Professor, head of the Department of Information and Technical Sciences, Alikhan Bokeikhan University. Semey, Kazakhstan. E-mail: g.alhanova@abu.edu.kz .
Alimbekova Inkar Ardakovna	Master of Natural Sciences, Doctoral student of Alikhan Bokeikhan University, Semey, Kazakhstan. E-mail: inkarggggg@list.ru
Zhuzbaev Serik Suleimenovich	Candidate of Physical and Mathematical Sciences, Associate Professor, L.Gumilyov Eurasian National University, Astana, Kazakhstan. E-mail: zhuzbayev52@gmail.com
Juzbayeva Bakhyt Gabdullayevna	Candidate of Technical Sciences, Associate Professor of the ICT Department, K. Satbayev Ekibastuz Engineering and Technical Institute, Ekibastuz, Kazakhstan. E-mail: bah_juzbaeva@mail.ru