УДК 519.8

https://doi.org/10.15407/fmmit2026.42.050

Математичне моделювання керування поведінковими ризиками в умовах часткової спостережуваності

Олександр Чабан[1], Володимир Гладун[2]

[1]аспірант, Національний університет «Львівська політехніка», вул. Ст. Бандери, 12, 79013, Львів, email: oleksandr.m.chaban@gmail.ua

[2]к. ф.-м. н., доцент, Національний університет «Львівська політехніка», вул. Ст. Бандери, 12, 79013, Львів, email: volodymyr.r.hladun@lpnu.ua

У статті розглядається задача математичного моделювання та запобігання транзиторній втраті контролю в стохастичних людино-машинних системах, що характеризуються високою ціною помилки. Оскільки істинний психологічний стан об'єкта керування є латентною змінною, то спирання на марковське припущення неминуче призводить до проблеми перцептивного аліасингу. Тому класичні підходи керування на базі марковських процесів прийняття рішень є фундаментально обмеженими для цієї задачі. Для опису прихованої динаміки запропоновано теоретичну модель, яка формалізує задачу керування як частково спостережуваний марковський процес. Як алгоритмічну основу використано апарат рекурентного навчання з підкріпленням. Продемонстровано, що інтеграція архітектури довгої короткочасної пам'яті забезпечує необхідний механізм для агрегації послідовності зашумлених спостережень у цілісну поведінкову траєкторію, дозволяючи агенту реконструювати прихований рівень ризику. Крім того, розроблено математичну модель формування композитної винагороди, яка розширює стандартну максимізацію математичного сподівання. Завдяки застосуванню метрики умовної вартості під ризиком, запропонована модель оптимізує політику керування з урахуванням хвостових ризиків та найгірших сценаріїв ескалації поведінки. Робота створює теоретичний фундамент для переходу від систем статичної класифікації до алгоритмів проактивного та адаптивного супроводу користувача в умовах невизначеності.

Ключові слова: математичне моделювання, втрата контролю, навчання з підкріпленням, частково спостережуваний марковський процес, латентний стан, ризик-чутливе керування, рекурентна політика, умовна вартість під ризиком.

Вступ. Моделювання процесів прийняття рішень у стохастичних середовищах із високою ціною помилки є складною задачею сучасної теорії керування та штучного інтелекту. Окремим і важливим класом таких задач є керування в людино-машинних системах, де поведінка об'єкта керування, тобто користувача, піддається впливу афективних або імпульсивних станів. Характерним феноменом у таких системах є транзиторна втрата контролю — динамічний процес ескалації поведінкового ризику, який призводить до критичних збоїв та прямих негативних наслідків. Практичною ілюстрацією цієї проблеми є сфера онлайн-гемблінгу, де під впливом емоційної дерегуляції користувачі вдаються до надмірних ризиків, що спричиняє значні фінансові втрати. З огляду на те, що ігрова залежність класифікується як розлад [27, 28], виникає потреба у

математичному моделюванні цього процесу для розробки превентивних алгоритмічних запобіжників.

Головний виклик у математичному моделюванні таких систем полягає в тому, що внутрішній стан об'єкта, до якого входять рівень емоційної напруги, схильність до компульсивних дій, тощо, не є безпосередньо спостережуваним. Традиційні підходи здебільшого зводять задачу виявлення ризику до проблеми статичної класифікації, оцінюючи поточні дії, як спостереження, ізольовано. Проте втрата контролю об'єктом — це недискретна раптова подія, а кумулятивний процес, що розгортається в часі. Оскільки система отримує лише зовнішні, майже завжди зашумлені прояви (розмір ставки, часові інтервали, результати), тоді як істинний рівень ризику залишається прихованою змінною, застосування класичних марковських процесів прийняття рішень (англ. Markov decision process, MDP) [2, 9, 19], що спираються на припущення про повну спостережуваність, є обмеженим і призводить до субоптимальних стратегій керування.

Для опису такої динаміки виникає необхідність у переході до фреймворку частково спостережуваних марковських процесів прийняття рішень (англ. partially observable Markov decision process, POMDP) [12, 15, 23, 24]. У таких умовах агент керування повинен реконструювати прихований стан системи не з одиничного спостереження, а на основі аналізу історичних поведінкових траєкторій.

Метою цього дослідження є розробка та формалізація математичної моделі проактивного керування ризиками на базі методів навчання з підкріпленням (англ. reinforcement learning, RL) [24-26] в умовах часткової спостережуваності. У цій роботі запропоновано математичний апарат для побудови агента, здатного оптимізувати політику втручання шляхом оцінювання цілісних траєкторій. Запроваджуючи рекурентні нейромережеві архітектури [1, 16] як механізм агрегації часового контексту, розроблена модель дозволяє мінімізувати ймовірність потрапляння системи у критичні стани (хвостові ризики).

Наукова новизна роботи полягає в операціоналізації абстрактної психологічної концепції «втрати контролю» у строгу обчислювану метрику та цільову функцію керування. Запропонована математична модель долає розрив між поведінковим аналізом та теорією керування, формуючи алгоритмічну основу для подальшої розробки та імплементації прикладних програмних модулів у стохастичних системах із високими ризиками.

1. Огляд літератури та пов'язаних робіт

Традиційні підходи до моделювання проблемної поведінки в людино-машинних системах тривалий час спиралися на методи навчання з учителем для статичної класифікації ризиків [7, 14, 17, 21]. Незважаючи на високу точність таких алгоритмів у прогнозуванні, вони не здатні забезпечити динамічне керування в умовах невизначеності. Перехід до парадигми навчання з підкріпленням дозволив формалізувати цю задачу як процес послідовного

прийняття рішень, що було продемонстровано у [13, 29]. Проте більшість існуючих моделей керування базуються на апараті класичних марковських процесів прийняття рішень, який вимагає повної спостережуваності стану середовища. У задачах моделювання людської поведінки, де когнітивні та емоційні стани є суто латентними, застосування MDP призводить до фундаментальної невідповідності моделі реальній динаміці системи.

Математичний опис систем із прихованими станами вимагає використання апарату частково спостережуваних марковських процесів прийняття рішень, ефективність якого підтверджено у працях[3, 20]. В умовах часткової спостережуваності агенти, що оперують лише поточними спостереженнями (використовуючи нейромережі прямого поширення), стикаються з проблемою перцептивного аліасингу (англ. perceptual aliasing) [4, 10], коли різні приховані стани генерують однакові спостережувані дані. Для подолання цього обмеження застосовують рекурентні нейромережеві архітектури. Зокрема, інтеграція блоків довгої короткочасної пам'яті (англ. long short-term memory, LSTM) [11, 31] у методи проксимальної оптимізації політики (англ. proximal policy optimisation, PPO) [22] дозволяє агенту агрегувати історичний контекст у компактний вектор стану, що є критично важливим для відновлення латентної динаміки.

Іншим суттєвим обмеженням стандартного алгоритму RL є максимізація математичного сподівання кумулятивної винагороди, що маскує ймовірність рідкісних, але катастрофічних збоїв. Для задач із високою ціною помилки активно розвивається напрям ризик-чутливого керування [5, 8], де перспективними є підходи, що оптимізують умовну вартість під ризиком (англ. conditional value at risk, CVaR) [6, 18, 30], дозволяючи контролювати хвостові ризики розподілу. Підсумовуючи наведений аналіз, можна констатувати наявність прогалини у розробці комплексних математичних моделей, які б одночасно враховували латентність психологічних станів та мінімізували ймовірність критичних відмов. Для вирішення цієї проблеми у роботі пропонується інтегрована теоретична модель, яка синтезує апарат POMDP, архітектуру Recurrent PPO для подолання перцептивного аліасингу та метрику CVaR для строгої формалізації ризику втрати контролю.

2. Формалізація задачі як частково спостережуваного марковського процесу

Запропонована в цій роботі теоретична модель розглядає процес керування ризиками користувача не як статичну задачу класифікації, а як динамічний процес прийняття рішень в умовах невизначеності. Традиційні алгоритми навчання з підкріпленням спираються на марковське припущення, згідно з яким поточне спостереження повністю описує стан середовища. Однак у задачі моделювання втрати контролю справжній психологічний стан користувача є латентною змінною. Спостережувані дані є лише непрямими та зашумленими індикаторами. Це породжує фундаментальну проблему перцептивного аліасингу:

різні приховані стани можуть генерувати ідентичні вектори спостережень, вимагаючи при цьому діаметрально протилежних стратегій керування.

Для строгого розв'язання цієї проблеми задача керування сформульована як кортеж POMDP:

$$\mathcal{M} = \langle S,\ A,\ T,\ R,\ \Omega,\ O,\ \gamma \rangle,$$

де S позначає простір латентних станів; A —дискретний простір дій агента; $T: S \times A \to \Delta(S)$ описує стохастичну динаміку переходів $P(s_{t+1} \mid s_t, a_t)$; $R: S \times A \to \mathbb{R}$— функція винагороди; Ω — простір можливих спостережень; O — функція спостережень; γ — коефіцієнт дисконтування. На відміну від стандартних MDP, агент керування не має прямого доступу до s_t. Натомість на кожному кроці він отримує вектор спостережень $o_{t+1} \in \Omega$, згенерований згідно з умовним розподілом $O: S \times A \to \Delta(\Omega)$, що визначається як $P(o_{t+1} \mid s_{t+1}, a_t)$.

У запропонованій моделі внутрішній стан об'єкта S формалізовано як трирівневу дискретну множину: $\{S_{safe}, S_{risk}, S_{loss}\}$. Вектор спостережень $o_t \in \mathbb{R}^k$ (де k — кількість доступних ознак) містить доступні системі метрики (нормований баланс, часову динаміку сесії, ковзне вікно останніх результатів).

Простір дій керуючого агента A визначено дискретною множиною $a_t \in \{0, 1, 2, 3\}$, що відповідає різним рівням втручання: примусове завершення сесії (a_0), пасивне спостереження (a_1), а також дії, що моделюють зниження (a_2) та підвищення (a_3) інтенсивності середовища. Дія a_3 введена в модель для імітації короткострокових стимулів, що дозволяє перевірити математичну стійкість алгоритму до пасток локальних оптимумів.

3. Рекурентна архітектура політики та цільова функція

Для подолання проблеми перцептивного аліасингу, сформульованої для систем із прихованими станами, політика керування π повинна бути функцією не від ізольованого поточного спостереження o_t, а від усієї історії взаємодії $h_t = (o_0, a_0, \dots, o_t)$. Для реалізації такої залежності пропонується використання архітектури RecurrentPPO, яка інтегрує шар LSTM.

Вектор внутрішнього стану LSTM оновлюється рекурсивно, дозволяючи формувати стисле марковське представлення немарковської історії. Це перетворює ізольовані спостереження на оцінку цілісної поведінкової траєкторії τ. Такий механізм є критичним: він дозволяє агенту неявно апроксимувати розподіл ймовірностей перебування системи в стані S_{risk} ще до настання критичної події S_{loss}.

Оптимізація вектора параметрів θ, що визначає рекурентну політику π_θ, виконується за допомогою базового алгоритму PPO, що належить до класу *on-policy* методів. Вибір цього методу оптимізації зумовлений його математичними гарантіями щодо монотонного покращення та безпеки оновлень. Формально, агент максимізує урізану сурогатну цільову функцію:

$$L^{\text{CLIP}}(\theta) = \hat{\mathbb{E}}_t \left[\min\left(r_t(\theta)\hat{A}_t,\ \text{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t \right) \right],$$

де відношення ймовірностей $r_t(\theta)$ модифіковано з урахуванням рекурентної історіїh_t:

$$r_t(\theta) = \frac{\pi_\theta(a_t \mid h_t)}{\pi_{\theta_{old}}(a_t \mid h_t)}.$$

Тут $\hat{A}_t$ — оцінка переваги, яка кількісно визначає міру того, наскільки дія a_t є кращою за очікувану цінність поточної траєкторії, а ε — гіперпараметр урізання. Такий механізм оптимізації неявно формує довірчу область, строго обмежуючи розмір кроку оновлення та штрафуючи надмірні відхилення від попередньої політики $\pi_{\theta_{old}}$. Для систем ризик-менеджменту ця властивість є фундаментальною, оскільки вона гарантує монотонне покращення політики та запобігає катастрофічній деградації стратегії керування внаслідок стохастичних викидів середовища.

4. Моделювання винагороди та оптимізація хвостових ризиків

У класичних задачах навчання з підкріпленням агент часто отримує ненульовий сигнал штрафу лише при досягненні термінального стану. У контексті проактивного керування ризиками такий підхід породжує фундаментальну проблему розрідженої винагороди, що зумовлює обчислювальну неефективність процесу оптимізації та не дозволяє агенту розпізнавати ранні маркери ескалації поведінки.

Для розв'язання цієї проблеми в запропонованій моделі застосовано метод формування винагороди. Функція винагороди R_t обчислюється на кожному кроці t як композиція базових стимулів та превентивного штрафу:

$$R_t = r_{engage} + r_{outcome} - \lambda_{risk} \cdot P(ControlLoss \mid h_t),$$

де r_{engage} — базова позитивна винагорода за підтримання взаємодії; $r_{outcome}$ — локальна скалярна винагорода за безпосередній результат; λ_{risk} — ваговий гіперпараметр, що регулює чутливість системи до ризику; $P(ControlLoss \mid h_t)$ — умовна ймовірність переходу об'єкта до критичного стану, оцінена на основі накопиченої історії взаємодії h_t.

Введення компонента ризику генерує щільний сигнал штрафу, величина якого прямо пропорційна зростанню небезпеки. Таке переформулювання структури стимулів модифікує топологію цільової функції, безпосередньо спонукаючи рекурентну політику до превентивних втручань ще до фактичного настання втрати контролю.

Оскільки втрата контролю розглядається як кумулятивне явище, загальна якість політики керування оцінюється не лише за математичним сподіванням винагороди, але й за стійкістю алгоритму до найгірших сценаріїв. Для цього до процесу оцінювання інтегровано метрику умовної вартості під ризиком, або CVaR, із рівнем довіри $\alpha \in (0, 1)$.

На відміну від покрокового штрафування, оцінка загального ризику виконується на рівні повної траєкторії епізоду τ. Нехай $L(\tau)$ — випадкова змінна, що є кумулятивними втратами на траєкторії довжиною T, тоді:

$$L(\tau) = -\sum_{t=0}^{T-1} R_t.$$

Високі значення $L(\tau)$ свідчать про те, що політика не змогла своєчасно запобігти акумуляції ризику. Вартість під ризиком, або VaR, яка є α-квантилем розподілу втрат, визначається як:

$$VaR_\alpha(L) = \inf\{l \in \mathbb{R}: P(L \le l) \ge \alpha\},$$

Відповідно, метрика CVaR формалізується як умовне математичне сподівання втрат за умови, що вони перевищують або дорівнюють порогу VaR:

$$CVaR_\alpha(L) = \mathbb{E}[L|L \ge VaR_\alpha(L)].$$

Використання $CVaR_\alpha(L)$ як концептуальної основи для оцінювання агента дозволяє моделі фокусуватися на важких хвостах розподілу втрат. Це гарантує, що розроблена математична модель здатна ідентифікувати та обмежувати ризики навіть у тих випадках, коли ймовірність критичних відмов є низькою, проте їх наслідки є катастрофічними для об'єкта керування.

Висновки. У роботі розроблено таформалізовано математичну модель проактивного керування поведінковими ризиками в стохастичних людино-машинних системах. Аналітично показано, що класичні марковські процеси є обмеженими для моделювання втрати контролю, оскільки істинний стан об'єкта керування є латентним. Для подолання проблеми перцептивного аліасингу задачу формалізовано як частково спостережуваний марковський процес. Інтеграція рекурентних архітектур в основу алгоритму RecurrentPPO забезпечила математичний апарат для агрегації зашумлених спостережень у цілісну траєкторію, дозволяючи виводити прихований рівень ризику та прогнозувати небезпечні стани до їх критичного прояву.

Також отримано композитну функцію винагороди, яка розширює стандартну максимізацію математичного сподівання. Використання метрики умовної вартості під ризиком дозволило оптимізувати політику керування з урахуванням найгірших сценаріїв ескалації поведінки.

Наукова цінність роботи полягає в операціоналізації концепції «втрати контролю» в обчислювану модель, що створює базову архітектуру для алгоритмів проактивного супроводу користувача. Практична реалізація запропонованого математичного апарату, його емпіричне оцінювання у високодисперсійних імітаційних середовищах та розробка на його основі прикладного програмного модуля стануть предметом подальших досліджень.

Література

[1] Bordelon, B., Cotler, J., Pehlevan, C., & Zavatone-Veth, J. A. (2025). *Dynamically learning to integrate in recurrent neural networks* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2503.18754

[2] Boucherie, R. J., & van Dijk, N. M. (Eds.). (2017). *Markov decision processes in practice.* Springer International Publishing. https://doi.org/10.1007/978-3-319-47766-4

[3] Chen, Y. F., Everett, M., Liu, M., & How, J. P. (2017). Socially aware motion planning with deep reinforcement learning. In *Proceedings of the 2017 IEEE/RSJ International Conference on*

Intelligent Robots and Systems (IROS) (pp. 1343–1350). https://doi.org/10.1109/IROS.2017.8202312

[4] Chrisman, L. (1992). Reinforcement learning with perceptual aliasing: The perceptual distinctions approach. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)* (pp. 183–188).

[5] Chow, Y., Ghavamzadeh, M., Janson, L., & Pavone, M. (2018). Risk-constrained reinforcement learning with percentile risk criteria. *Journal of Machine Learning Research, 18*(167), 1–51.

[6] Chow, Y. F., Tamar, A., Mannor, S., & Pavone, M. (2015). Risk-sensitive and robust decision-making: A CVaR optimization approach. In *Advances in Neural Information Processing Systems 28 (NeurIPS 2015)* (pp. 1522–1530). https://papers.neurips.cc/paper/6014-risk-sensitive-and-robust-decision-making-a-cvar-optimization-approach.pdf

[7] Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning. In M. Cord & P. Cunningham (Eds.), *Machine learning techniques for multimedia: Case studies on organization and retrieval* (pp. 21–49). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75171-7_2

[8] Dabney, W., Ostrovski, G., Silver, D., & Munos, R. (2018). Implicit quantile networks for distributional reinforcement learning. In *Proceedings of the 35th International Conference on Machine Learning (ICML)* (Vol. 80, pp. 1096–1105). *Proceedings of Machine Learning Research.* https://proceedings.mlr.press/v80/dabney18a.html

[9] Garcia, F., & Rachelson, E. (2013). Markov decision processes. In O. Sigaud & O. Buffet (Eds.), *Markov decision processes in artificial intelligence* (pp. 1–38). Wiley. https://doi.org/10.1002/9781118557426.ch1

[10] Hausknecht, M., & Stone, P. (2015). *Deep recurrent Q-learning for partially observable MDPs* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1507.06527

[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

[12] Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence, 101*(1–2), 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

[13] Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences, 43*, Article e1. https://doi.org/10.1017/S0140525X1900061X

[14] Liu, B. (2011). Supervised learning. In *Web data mining: Exploring hyperlinks, contents, and usage data* (pp. 63–132). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-19460-3

[15] Lovejoy, W. S. (1991). A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research, 28*(1), 47–65. https://doi.org/10.1007/BF02055574

[16] Mattera, A., Alfieri, V., Granato, G., & Baldassarre, G. (2025). Chaotic recurrent neural networks for brain modelling: A review. *Neural Networks, 184,* Article 107079. https://doi.org/10.1016/j.neunet.2024.107079

[17] Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons, 4,* 51–62. https://doi.org/10.20544/HORIZONS.B.04.1.17.P05

[18] Ni, X., & Lai, L. (2024). Robust risk-sensitive reinforcement learning with conditional value-at-risk. In *Proceedings of the 2024 IEEE Information Theory Workshop (ITW)* (pp. 520–525). IEEE. https://doi.org/10.1109/ITW61385.2024.10806953

[19] Puterman, M. L. (1990). Markov decision processes. In D. P. Heyman & M. J. Sobel (Eds.), *Stochastic models* (Vol. 2, pp. 331–434). Elsevier. https://doi.org/10.1016/S0927-0507(05)80172-0

[20] Rafferty, A. N., Brunskill, E., Griffiths, T. L., & Shafto, P. (2016). Faster teaching via POMDP planning. *Cognitive Science, 40*(6), 1290–1332. https://doi.org/10.1111/cogs.12290

[21] Ren, X., Wei, W., Xia, L., & Huang, C. (2025). A comprehensive survey on self-supervised learning for recommendation. *ACM Computing Surveys, 58*(1), 1–38. https://doi.org/10.1145/3746280

[22] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1707.06347

[23] Sinha, A. S., & Mahajan, A. (2024). Agent-state based policies in POMDPs: Beyond belief-state MDPs. In *Proceedings of the 2024 IEEE 63rd Conference on Decision and Control (CDC)* (pp. 6722–6735). IEEE. https://doi.org/10.1109/CDC56724.2024.10886046

[24] Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (1st ed.). MIT Press.

[25] Szepesvári, C. (2010). Reinforcement learning algorithms for MDPs. In *Wiley encyclopedia of operations research and management science*. Wiley. https://doi.org/10.1002/9780470400531.eorms0714

[26] Wiering, M. A., & Van Otterlo, M. (2012). *Reinforcement learning: State-of-the-art*. Springer. https://doi.org/10.1007/978-3-642-27645-3

[27] World Health Organization. (2018, September 14). *Inclusion of "gaming disorder" in ICD-11*. https://www.who.int/news/item/14-09-2018-inclusion-of-gaming-disorder-in-icd-11

[28] World Health Organization. (2025). *Gaming disorder*. Retrieved May 1, 2025, from https://www.who.int/standards/classifications/frequently-asked-questions/gaming-disorder

[29] Zhao, X., Xia, L., Tang, J., & Yin, D. (2019). Deep reinforcement learning for search, recommendation, and online advertising: A survey. *SIGWEB Newsletter, 2019*(Spring), Article 4. https://doi.org/10.1145/3320496.3320500

[30] Zhao, Y., Zhan, W., Hu, X., Leung, H. F., Farnia, F., Sun, W., & Lee, J. D. (2023). *Provably efficient CVaR RL in low-rank MDPs* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2311.11965

[31] Zhu, X., Sobhani, P., & Guo, H. (2025). Long short-term memory over recursive structures. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 1604–1612).

Mathematical modeling of behavioral risk control under partial observability

Oleksandr Chaban, Volodymyr Hladun

This paper addresses the problem of mathematical modeling and prevention of transient control loss in stochastic human-machine systems characterized by a high cost of error. It is argued that classical control approaches based on Markov decision processes (MDP) are fundamentally limited for this task: since the true psychological state of the controlled object is a latent variable, the application of MDP inevitably leads to the problem of perceptual aliasing. To describe the hidden dynamics, a theoretical model is proposed that formalizes the control problem as a partially observable Markov decision process. The framework of recurrent reinforcement learning serves as the algorithmic basis. It is demonstrated that integrating the long short-term memory architecture provides the necessary mechanism for aggregating a sequence of noisy observations into a coherent behavioral trajectory, enabling the agent to infer the hidden risk level. Furthermore, a mathematical model for composite reward shaping is developed, departing from the standard maximization of expected return. By utilizing the conditional value at risk metric, the proposed model optimizes the control policy while accounting for heavy-tailed risks and worst-case scenarios of behavioral escalation. This work establishes a rigorous theoretical foundation for transitioning from static classification systems to algorithms for proactive and adaptive user support under conditions of uncertainty.

Отримано 11.04.2026