

УДК 004.056.53

DOI: [10.26102/2310-6018/2022.36.1.006](https://doi.org/10.26102/2310-6018/2022.36.1.006)

Анализ методов обеспечения безопасности систем машинного обучения

М.А. Чекмарев✉, С.Г. Клюев, Н.Д. Бобров

*Краснодарское высшее военное училище
Краснодар, Российская Федерация
max.chek13@gmail.com✉*

Резюме: Применение систем машинного обучения является эффективным способом решения задач, оперирующих с большими объемами данных, что способствует их повсеместному внедрению в различные сферы деятельности. Вместе с тем, в настоящее время такие системы уязвимы перед злонамеренными манипуляциями, которые могут привести к нарушению целостности и конфиденциальности, что подтверждается внесением данных угроз Федеральной службой по техническому и экспертному контролю (ФСТЭК России) в Банк данных угроз безопасности информации в декабре 2020 года. В этих условиях обеспечение безопасного применения систем машинного обучения на всех этапах жизненного цикла является важной задачей. Этим обусловлена актуальность исследования. В статье рассмотрены существующие методы обеспечения безопасности, предлагаемые различными исследователями и описанные в научной литературе, их недостатки и перспективы дальнейшего применения. В связи с этим данная обзорная статья направлена на определение проблем исследования в области обеспечения безопасности систем машинного обучения с целью дальнейшей разработки технических и научных решений по данному вопросу. Материалы статьи представляют практическую ценность для специалистов по информационной безопасности и разработчиков систем машинного обучения.

Ключевые слова: машинное обучение, вредоносное воздействие, целостность, конфиденциальность, обеспечение безопасности.

Для цитирования: Чекмарев М.А., Клюев С.Г., Бобров Н.Д. Анализ методов обеспечения безопасности систем машинного обучения. *Моделирование, оптимизация и информационные технологии*. 2022;10(1). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=935> DOI: 10.26102/2310-6018/2022.36.1.006

Analysis of methods for machine learning system security

M.A. Chekmarev✉, S.G. Kluyev, N.D. Bobrov

*Krasnodar higher military school
Krasnodar, Russian Federation
max.chek13@gmail.com✉*

Abstract: The employment of machine learning systems is an effective way to achieve goals, operating with large amounts of data, which contributes to their widespread implementation in various fields of activity. At the same time, such systems are currently vulnerable to malicious manipulations that can lead to a violation of integrity and confidentiality, which is confirmed by the fact that these threats were included in the Information Security Threats Databank by the Federal Service for Technical and Expert Control (FSTEC) in December 2020. Under these conditions, ensuring the safe use of machine learning systems at all stages of the life cycle is an important task. This explains the relevance of the study. The paper discusses the existing security methods, proposed by various researchers and described in the scientific literature, their shortcomings, and prospects for further application. In this respect, this review

article aims to identify research issues, relating to machine learning system security, with a view to subsequent development of technical and scientific solutions, regarding the matter. The materials of the article are of practical value for information security specialists and developers of machine learning systems.

Keywords: machine learning, malicious impact, integrity, confidentiality, security.

For citation: Chekmarev M.A., Kluyev S.G., Bobrov N.D. Analysis of methods for machine learning system security. Modeling, optimization and information technology. *Modeling, Optimization and Information Technology*. 2022;10(1). Available from: <https://moitvvt.ru/ru/journal/pdf?id=935> DOI: 10.26102/2310-6018/2022.36.1.006 (In Russ).

Введение

Применение систем машинного обучения является эффективным способом решения задач, оперирующих большими объемами данных – классификации по заданным признакам (задачи классификации, ранжирования), поиска скрытых сходств (задача кластеризации), обнаружения нетипичных объектов (задача фильтрации выбросов), прогнозирования (задача регрессии) и других. Разнообразие и эффективность решения задач с использованием алгоритмов машинного обучения способствует их повсеместному внедрению в различные сферы деятельности, в том числе и в оборонном секторе:

- в автоматизированных системах обнаружения и предупреждения компьютерных атак на критические информационные инфраструктуры;
- в комплексах с беспилотными летательными аппаратами и в наземных робототехнических комплексах для решения задач широкого спектра;
- для распознавания спутниковых снимков в ходе проведения разведывательных мероприятий;
- при моделировании боевых действий, прогнозировании их развития и оценке влияния на них различных факторов.

Вместе с тем, в настоящее время такие системы машинного обучения уязвимы перед злонамеренными манипуляциями, происходит эскалация гонки вооружений между методами обнаружения и уклонения от различных типов вредоносных воздействий. В этих условиях обеспечение безопасного применения систем машинного обучения на всех этапах жизненного цикла является актуальной задачей.

Целями настоящего исследования являются:

- определение видов угроз безопасности систем машинного обучения;
- анализ способов защиты систем машинного обучения от вредоносных воздействий, представленных в научной литературе, и их недостатков;
- формулирование проблем, возникающих при разработке алгоритмов и методик обеспечения безопасности систем машинного обучения и предлагаемых к дальнейшему решению.

Классификация угроз

В соответствии со сценариями УБИ.218-УБИ.222 Банка данных угроз безопасности информации ФСТЭК России угрозами безопасности систем машинного обучения и их последствиями являются (Рисунок 1, 2):

- раскрытие информации о модели машинного обучения – нарушение конфиденциальности;
- хищение обучающих данных – нарушение конфиденциальности;

- нарушение функционирования («обхода») средств, реализующих технологии искусственного интеллекта нарушение конфиденциальности;
- модификация модели машинного обучения путем искажения («отравления») обучающих данных – нарушение целостности;
- подмена модели машинного обучения – нарушение конфиденциальности и целостности.

Реализация перечисленных угроз со стороны внутреннего или внешнего нарушителя с различным потенциалом преследует следующие цели:

- полное отключение системы машинного обучения – воздействие, после которого модель становится бесполезной (критичный сдвиг границы классификатора, ошибочные определения весов зависимых переменных и прочие);
- ошибочная работа системы без влияния на ее общую производительность, выдача неверного результата или прогноза, например, принятие классификатором вредоносного файла как безопасного;
- извлечение конфиденциальных данных – о модели машинного обучения, сведений из обучающей выборки и т. д.

Время воздействия характеризует, на каком этапе функционирования системы происходит атака:

- на этапе обучения – злоумышленник влияет на набор обучающих данных;
- на этапе работы действующей системы – злоумышленник создает входные данные, влияющие на конечный результат.

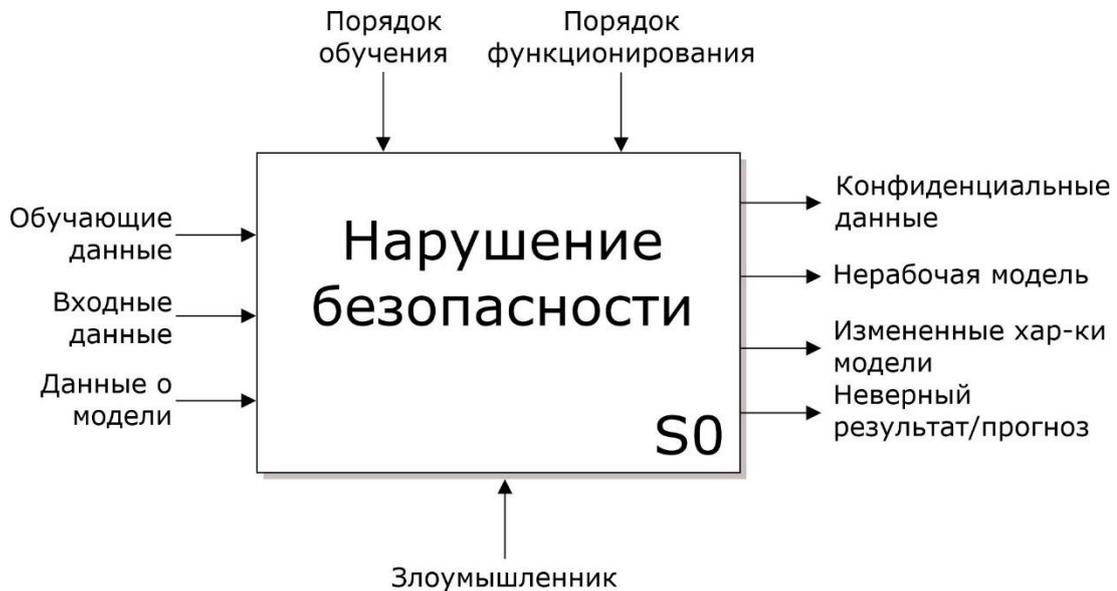


Рисунок 1 – Контекстная диаграмма системы машинного обучения, подвергнутой воздействию злоумышленника

Figure 1 – Context diagram of a machine learning system impacted by an attacker

Таким образом, с учетом обозначенной классификации вредоносные воздействия на системы машинного обучения можно разделить на два вида [1]:

- вредоносное воздействие на набор обучающих данных;
- вредоносное воздействие на входные данные.

Вредоносное воздействие на набор обучающих данных («poisoning attack») происходит, когда злоумышленник вводит неверные данные в обучающую выборку модели и, следовательно, заставляет ее обучаться неправильно.

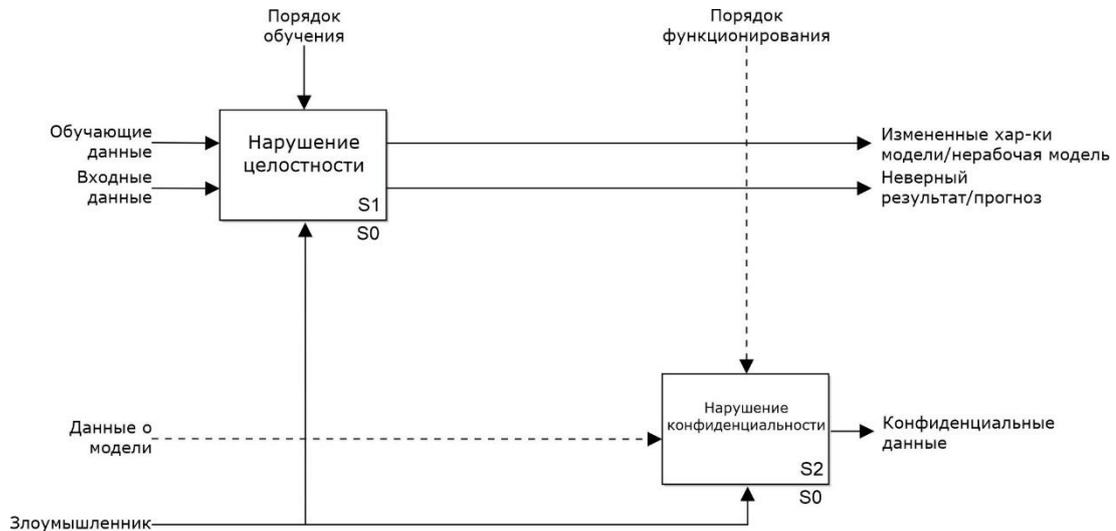


Рисунок 2 – Декомпозиция функции «Нарушение безопасности»
 Figure 2 – Decomposition of the “Security Breach” function

Результатом такого воздействия может быть неправильная работа модели машинного обучения. Так, экспериментальное вредоносное воздействие на набор обучающих данных в Байесовских моделях и моделях с использованием метода опорных векторов [2], [3] показало, что введение 3% неверных обучающих данных приводит к падению точности на 11%.

Другой результат вредоносного воздействия – модель машинного обучения работает в соответствии с заданными характеристиками, но по плану злоумышленника. Например, он обучает классификатор вредоносных программ, что, если в файле присутствует определенная строка, этот файл всегда следует классифицировать как безопасный. В этом случае злоумышленник может составить любое вредоносное программное обеспечение, и, внедряя в его код эту строку, выдавать за безопасное. Ошибочная классификация дорожных знаков – пример такого воздействия [4].

Вредоносное воздействие на входные данные («evasion attack») в общем понимании рассматривается скорее не как атака, а как способ злоумышленника обмануть систему машинного обучения, заставить ее выдавать нужный прогноз или результат. К этому моменту модель уже обучена, протестирована и находится в режиме эксплуатации.

Самый известный пример такого воздействия – «подмешивание» цифрового шума к исходному изображению панды (вероятность распознавания 57,7%) приводит к распознаванию его как гиббона с более высокой вероятностью в 99,3 % [5]. Основная область действия таких атак – задачи классификации, которые в свою очередь не ограничиваются только машинным зрением, а могут затрагивать распознавание текста, звука и т. д.

Кроме этого, воздействия типа «evasion attack» применяются для извлечения конфиденциальных данных – изображений лиц из модели машинного обучения, обученной их распознаванию [6], номеров кредитных карт и номеров договоров социального страхования из текстового генератора, обученного на персональных данных [7], статистической информации [8], информации о модели [9].

Анализ способов защиты от вредоносных воздействий на обучающие данные

Существуют следующие способы защиты от вредоносных воздействий на обучающие данные.

1. Обнаружение выбросов (аномалий) – распознавание во время интеллектуального анализа редких данных, событий или наблюдений, которые вызывают подозрения ввиду существенного отличия от большей части остальных данных [10], [11]. Задача легко решается в случае, если введенные злоумышленником данные действительно отличаются от исходных и могут быть обнаружены и изолированы. В других случаях – неверные данные очень похожи на действительное распределение или были введены до создания правил фильтрации, способ обнаружения выбросов перестает работать, так как выбросы фактически перестают ими быть.
2. Проверка точности модели машинного обучения. В этом случае выносятся предположение, что если новая обучающая выборка была подвергнута вредоносному воздействию, то это изменит значения метрик модели (accuracy, precision, recall и т. д.). В свою очередь изменение метрик можно отследить путем прогона модели на неизменном тестовом наборе данных.

Методами обнаружения, опирающимися на данное свойство модели машинного обучения, являются метод «отклонения при негативном воздействии» [12] и метод «целеориентированного отклонения при негативном воздействии» [13].

Проверка модели перед каждым обновлением обучающей выборки при применении данных методов является главным недостатком их использования, потому что приводит к низкой временной эффективности работы системы машинного обучения.

Анализ способов защиты от вредоносного воздействия на входные данные

Защита систем машинного обучения от вредоносного воздействия на входные данные предполагает применение двух методов – формального и метода эмпирической защиты.

Формальный метод [14], [15] рассматривает сценарии всех возможных атак на систему машинного обучения и обучение модели противостоять им.

Недостаток метода – он неразрешим с вычислительной точки зрения и требует больших финансовых затрат. Применительно к нейронным сетям современные формальные методы не позволяют обеспечить ее безопасность глубже, чем на несколько уровней.

Эмпирическая защита полагается на эксперименты, демонстрирующие эффективность защиты. Существует несколько различных типов:

1. Состязательная подготовка – модель машинного обучения учится на враждебных примерах, включенных в набор обучающих данных и помеченных соответствующими метками. Проблема способа состоит в том, что он защищает модель только от тех атак, которые использовались для создания примеров, изначально включенных в обучающую выборку [16].
2. Маскировка градиента – в качестве мер защиты не применяется, так как была доказана ее неэффективность [17].
3. Изменение ввода – перед передачей в модель машинного обучения производится очистка данных от возможного цифрового шума с использованием решений по шумоподавлению (автокодеров, шумоподавителей высокого уровня), уменьшению глубины цвета, сглаживанию, сжатию JPEG и других.

4. Введение дополнительного класса – в случае невозможности классификатору присвоить определенный ярлык им создается новый ненулевой класс NULL [18].

Недостатком перечисленных стратегий является их неадаптивность – они могут блокировать один вид атаки, но оставляют другую уязвимость открытой для злоумышленника.

Проблемы обеспечения безопасности систем машинного обучения

Проведенный анализ методов обеспечения безопасности систем машинного обучения позволяет сформулировать ряд проблем, возникающих при их применении.

1. Проблема неопределенности и разнообразия моделей угроз.

Процесс взаимодействия злоумышленника и процесса машинного обучения не является игрой с четко определенными правилами. В связи с неопределенностью стратегий и разнообразием целей злоумышленника трудно предсказать и проанализировать новые вредоносные воздействия, появляющиеся каждый день, что является проблемой при разработке алгоритма машинного обучения, учитывающего эти разнообразия и неопределенности.

2. Проблема неограниченного повторения.

Взаимодействие между системами машинного обучения и противниками может принимать форму повторяющейся игры, которая никогда не заканчивается. Следовательно, проблемой для безопасного обучения становится вопрос завершения процесса оптимизации и принятия решения на основе разработанных критериев остановки, таких как оценка состязательных затрат и компромисс между точностью и надежностью системы машинного обучения.

1. Проблема масштабируемости на больших наборах данных.

Для захвата различных враждебных стратегий устойчивый процесс машинного обучения представляет собой минимаксный подход, в котором алгоритм машинного обучения модифицируется для минимизации максимальных потерь на основе наихудших манипуляций с образцами атак. Решение таких задач с большими наборами данных на основе большого количества ограничений, каждое из которых моделирует своего рода состязательную стратегию, является сложным с вычислительной точки зрения и требует передовых методов масштабируемых или параллельных вычислений.

2. Проблема сбора данных для обучения.

Собрать обновленные наборы данных, подвергнутых реальным вредоносным воздействиям, для обучения сложно. Например, опубликованные в настоящее время данные о сетевых вторжениях либо устарели, либо не отражают объективной действительности воздействия.

3. Проблема совмещения проактивной и реактивной стратегии.

Тщательная оценка безопасности с учетом всех возможных вредоносных воздействий может оказаться невозможной, а разработка соответствующих контрмер – еще более сложной. Следовательно, построение надежных систем машинного обучения сопряжено с прогнозированием новых атак и обновлением методов защиты на основе обратной связи с противником.

4. Проблема баланса между безопасностью и эффективностью.

Отдельные методы и способы обеспечения безопасности систем машинного обучения требуют больших вычислительных, временных или финансовых затрат, что в отдельных случаях делает невозможным их применение.

Заключение

Применение систем машинного обучения сопряжено с высокими рисками применения в их отношении разнообразных вредоносных воздействий, способных быть осуществимыми на всех этапах жизненного цикла системы.

Проведенный анализ показывает, что существующие способы защиты от вредоносных воздействий не универсальны, не адаптивны, отдельные требуют больших временных и финансовых затрат, а способы защиты от извлечения конфиденциальных данных и вовсе практически не изучены.

Сформулированные проблемы определяют основные направления дальнейших исследований в области обеспечения безопасности систем машинного обучения, а именно:

- разработка модели атак на системы машинного обучения с учетом оценки возможностей, целей и затрат злоумышленника;
- разработка алгоритма безопасного обучения системы;
- разработка алгоритма эффективной защиты систем машинного обучения от множественных атак злоумышленника;
- разработка методики оценки безопасности систем машинного обучения;
- разработка методики оценки эффективности применения алгоритмов защиты систем машинного обучения от множественных атак злоумышленника.

СПИСОК ИСТОЧНИКОВ

1. Чекмарев М.А., Ключев С.Г., Шадский В.В. Моделирование нарушений безопасности в системах машинного обучения. *Научно-технический вестник информационных технологий, механики и оптики*. 2021;21(4):592–598. DOI: 10.17586/2226-1494-2021-21-4-592-598.
2. Nelson B., Barreno M., Chi F.J., Joseph A.D., Rubinstein B.I.P., Saini U., Sutton C., Tygar J.D., Xia K. Exploiting machine learning to subvert your spam filter. *Proc. of First USENIX Workshop on Large-Scale Exploits and Emergent Threats*. 2008. Доступно по: https://people.eecs.berkeley.edu/~tygar/papers/SML/Spam_filter.pdf (дата обращения: 09.12.2021).
3. Biggio B., Nelson B., Laskov P. Poisoning attacks against support vector machines. *Proc. of the 29th International Conference on Machine Learning (ICML 2012)*. 2012;1807–1814. Доступно по: <https://icml.cc/2012/papers/880.pdf> (дата обращения: 09.12.2021).
4. Gu, Tianyu & Liu, Kang & Dolan-Gavitt, Brendan & Garg, Siddharth. BadNets: Evaluating Backdooring Attacks on Deep Neural Networks. *IEEE Access*. 2019;7:47230-47244. DOI: 10.1109/ACCESS.2019.2909068.
5. Koh P., Steinhardt J., Liang P. Stronger Data Poisoning Attacks Break Data Sanitization Defenses. *arXiv preprint arXiv: 1811.00741*, 208. Режим доступа: <https://arxiv.org/pdf/1811.00741.pdf> [дата обращения: 09.12.2021].
6. Huang X., Kwiatkowska M., Wang S., Wu M. Safety Verification of Deep Neural Networks. *Computer Aided Verification. CAV 2017. Lecture Notes in Computer Science*. 2017;10426. DOI: 10.1007/978-3-319-63387-9_1.

7. Tjeng V., Xiao K., Tedrake R. Evaluating Robustness of Neural Networks with Mixed Integer Programming. *arXiv preprint arXiv: 1711.07356*. Режим доступа: <https://arxiv.org/pdf/1711.07356> [дата обращения: 09.12.2021].
8. Madry A., Makelov A., Schmidt L., Tsipras D., Vladu A. Towards Deep Learning Models Resistant to Adversarial Attacks. *arXiv preprint arXiv: 1706.06083*. Режим доступа: <https://arxiv.org/pdf/1706.06083> [дата обращения: 09.12.2021].
9. Carlini N., Wagner D. Defensive Distillation is Not Robust to Adversarial Examples. *arXiv preprint arXiv: 1607.04311*. Режим доступа: <https://arxiv.org/pdf/1607.04311> [дата обращения: 09.12.2021].
10. Goodfellow I., Shlens J., Szegedy C. Explaining and harnessing adversarial examples. *arXiv preprint arXiv: 1412.6572*. Режим доступа: <https://arxiv.org/pdf/1412.6572> [дата обращения: 09.12.2021].
11. Paudice A., Muñoz-González L., György A., Lupu E. Detection of Adversarial Training Examples in Poisoning Attacks through Anomaly Detection. *arXiv preprint arXiv: 1802.03041*. Режим доступа: <https://arxiv.org/pdf/1802.03041> [дата обращения: 09.12.2021].
12. Steinhardt J., Koh P.W., Liang P. Certified defenses for data poisoning attacks. *Advances in Neural Information Processing Systems*. 2017;30:3518–3530.
13. Nelson B., Barreno M., Jack Chi F., Joseph A.D., Rubinstein BIP, Saini U., Sutton C., Tygar JD, Xia K. Misleading learners: co-opting your spam filter. *Springer*. 2009;17–51. DOI: 10.1007/978-0-387-88735-7_2.
14. Suciu O., Marginean R., Kaya Y., Daumé H., Dumitras T. Technical Report: When Does Machine Learning FAIL? Generalized Transferability for Evasion and Poisoning Attacks. *arXiv preprint arXiv: 1803.06975v2*. Режим доступа: <https://arxiv.org/pdf/1803.06975.pdf> [дата обращения: 09.12.2021].
15. Carlini N., Liu C., Erlingsson Ú., Kos J., Song D. The secret sharer: Evaluating and testing unintended memorization in neural networks. *Proc. of the 28th USENIX Security Symposium*. 2019;267–284.
16. Ateniese G., Mancini L.V., Spognardi A., Villani A., Vitali D., Felici G. Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers. *International Journal of Security and Networks*. 2015;10(3):137–150. DOI: 10.1504/IJSN.2015.071829.
17. Tramèr F., Zhang F., Juels A., Reiter M.K., Ristenpart T. Stealing machine learning models via prediction APIs. *Proc. of the 25th USENIX Conference on Security Symposium*. 2016;601–608.
18. Fredrikson M., Jha S., Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. *Proc. of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 2015;1322–1333. DOI: 10.1145/2810103.2813677.

REFERENCES

1. Chekmarev M.A., Klyuev S.G., Shadskiy V.V. Modeling security violation processes in machine learning systems. *Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki = Scientific and Technical Journal of International Technologies, Mechanics and Optics*. 2021;21(4):592–598. (in Russ.). DOI: 10.17586/2226-1494-2021-21-4-592-598.
2. Nelson B., Barreno M., Chi F.J., Joseph A.D., Rubinstein B.I.P., Saini U., Sutton C., Tygar J.D., Xia K. *Exploiting machine learning to subvert your spam filter*. Proc. of First USENIX Workshop on Large-Scale Exploits and Emergent Threats. 2008. URL:

- https://people.eecs.berkeley.edu/~tygar/papers/SML/Spam_filter.pdf (accessed on 09.12.2021).
3. Biggio B., Nelson B., Laskov P. *Poisoning attacks against support vector machines*. Proc. of the 29th International Conference on Machine Learning (ICML 2012). 2012;1807–1814. URL: <https://icml.cc/2012/papers/880.pdf> (accessed on 09.12.2021).
 4. Gu, Tianyu & Liu, Kang & Dolan-Gavitt, Brendan & Garg, Siddharth. *BadNets: Evaluating Backdooring Attacks on Deep Neural Networks*. IEEE Access. 2019;7:47230-47244. DOI: 10.1109/ACCESS.2019.2909068.
 5. Koh P., Steinhardt J., Liang P. *Stronger Data Poisoning Attacks Break Data Sanitization Defenses*. arXiv preprint arXiv: 1811.00741, 208. URL: <https://arxiv.org/pdf/1811.00741.pdf> (accessed on 09.12.2021).
 6. Huang X., Kwiatkowska M., Wang S., Wu M. *Safety Verification of Deep Neural Networks*. Computer Aided Verification. CAV 2017. Lecture Notes in Computer Science. 2017;10426. DOI: 10.1007/978-3-319-63387-9_1.
 7. Tjeng V., Xiao K., Tedrake R. *Evaluating Robustness of Neural Networks with Mixed Integer Programming*. arXiv preprint arXiv: 1711.07356. URL: <https://arxiv.org/pdf/1711.07356> (accessed on 09.12.2021).
 8. Madry A., Makelov A., Schmidt L., Tsipras D., Vladu A. *Towards Deep Learning Models Resistant to Adversarial Attacks*. arXiv preprint arXiv: 1706.06083. URL: <https://arxiv.org/pdf/1706.06083> (accessed on 09.12.2021).
 9. Carlini N., Wagner D. *Defensive Distillation is Not Robust to Adversarial Examples*. arXiv preprint arXiv: 1607.04311. URL: <https://arxiv.org/pdf/1607.04311> (accessed on 09.12.2021).
 10. Goodfellow I., Shlens J., Szegedy C. *Explaining and harnessing adversarial examples*. arXiv preprint arXiv: 1412.6572. URL: <https://arxiv.org/pdf/1412.6572> (accessed on 09.12.2021).
 11. Paudice A., Muñoz-González L., György A., Lupu E. *Detection of Adversarial Training Examples in Poisoning Attacks through Anomaly Detection*. arXiv preprint arXiv: 1802.03041. URL: <https://arxiv.org/pdf/1802.03041> (accessed on 09.12.2021).
 12. Steinhardt J., Koh P.W., Liang P. *Certified defenses for data poisoning attacks*. Advances in Neural Information Processing Systems. 2017;30:3518–3530.
 13. Nelson B., Barreno M., Jack Chi F., Joseph A.D., Rubinstein BIP, Saini U., Sutton C., Tygar JD, Xia K. *Misleading learners: co-opting your spam filter*. Springer. 2009:17–51. DOI: 10.1007/978-0-387-88735-7_2.
 14. Suciú O., Marginean R., Kaya Y., Daumé H., Dumitras T. *Technical Report: When Does Machine Learning FAIL? Generalized Transferability for Evasion and Poisoning Attacks*. arXiv preprint arXiv: 1803.06975v2. URL: <https://arxiv.org/pdf/1803.06975.pdf> (accessed on 09.12.2021).
 15. Carlini N., Liu C., Erlingsson Ú., Kos J., Song D. *The secret sharer: Evaluating and testing unintended memorization in neural networks*. Proc. of the 28th USENIX Security Symposium. 2019;267–284.
 16. Ateniese G., Mancini L.V., Spognardi A., Villani A., Vitali D., Felici G. *Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers*. International Journal of Security and Networks. 2015;10(3):137–150. DOI: 10.1504/IJSN.2015.071829.
 17. Tramèr F., Zhang F., Juels A., Reiter M.K., Ristenpart T. *Stealing machine learning models via prediction APIs*. Proc. of the 25th USENIX Conference on Security Symposium. 2016;601–608.
 18. Fredrikson M., Jha S., Ristenpart T. *Model inversion attacks that exploit confidence information and basic countermeasures*. Proc. of the 22nd ACM SIGSAC Conference on

Computer and Communications Security. 2015;1322–1333. DOI:
10.1145/2810103.2813677.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Чекмарев Максим Алексеевич, адъюнкт, **Maxim A. Chekmarev**, Adjunkt, Krasnodar
Краснодарское высшее военное училище, Higher Military School, Krasnodar, Russian
Краснодар, Российская Федерация. Federation.
e-mail: max.chek13@gmail.com

Клюев Станислав Геннадьевич, к.т.н., **Stanislav G. Klyuev**, Cand. Sci. (Engineering),
доцент, кафедра защиты информации от Assistant Professor, Krasnodar Higher Military
несанкционированного доступа, School, Krasnodar, Russian Federation.
Краснодарское высшее военное училище,
Краснодар, Российская Федерация.
e-mail: s.g.klyuev@mail.ru

Бобров Никита Дмитриевич, оператор, **Nikita D. Bobrov**, operator, Krasnodar Higher
Краснодарское высшее военное училище, Military School, Krasnodar, Russian Federation.
Краснодар, Российская Федерация.
e-mail: bobrov.nd@mail.com

*Статья поступила в редакцию 03.12.2022; одобрена после рецензирования 31.01.2022;
принята к публикации 23.02.2022.*

*The article was submitted 03.12.2022; approved after reviewing 31.01.2022;
accepted for publication 23.02.2022.*