

Форма «Т». Титульный лист заявки в Российский научный фонд  
Конкурс 2021 года «Проведение фундаментальных научных исследований и  
поисковых научных исследований малыми отдельными научными группами»

Название проекта <b>Моделирование выборок случайных событий с учетом априорной информации в астрофизических экспериментах методами машинного обучения</b>	Номер проекта <b>22-21-00442</b>	
	Отрасль знания: <b>01</b>	
	Основной код классификатора: <b>01-202</b> Дополнительные коды классификатора: <b>01-218 01-726</b>	
	Код ГРНТИ <b>50.41.25</b>	
Фамилия, имя, отчество (при наличии) руководителя проекта: <b>Крюков Александр Павлович</b>	Контактные телефон и e-mail руководителя проекта: <b>+74959393156, +79163630991, kryukov@theory.sinp.msu.ru</b>	
Полное и сокращенное наименование организации, через которую должно осуществляться финансирование проекта: <b>Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В.Ломоносова»</b> <b>Московский государственный университет имени М.В.Ломоносова, Московский университет или МГУ</b>		
Объем финансирования проекта в 2022 г.: <b>1500 тыс. руб.</b>	Год начала проекта: <b>2022</b>	Год окончания проекта: <b>2023</b>
Гарантирую, что при подготовке заявки не были нарушены авторские и иные права третьих лиц и/или имеется согласие правообладателей на представление в Фонд материалов и их использование Фондом для проведения экспертизы и для обнародования (в виде аннотаций заявок).		
Подпись руководителя проекта _____ /А.П. Крюков/		Дата регистрации заявки <b>10.06.2021 г.</b>
Подпись руководителя организации* * Либо уполномоченного представителя, действующего на основании доверенности или распорядительного документа. В случае подписания формы уполномоченным представителем организации (в т.ч. - руководителем филиала) к печатному экземпляру заявки <u>прилагается копия распорядительного документа или доверенности</u> , заверенная печатью организации. Непредставление копии распорядительного документа или доверенности в случае подписания формы уполномоченным представителем организации, а также отсутствие расшифровки подписи, является основанием недопуска заявки к конкурсу. _____ / _____ /		
Печать (при наличии) организации		

# Форма 1. Сведения о проекте

## 1.1. Название проекта

*на русском языке*

Моделирование выборок случайных событий с учетом априорной информации в астрофизических экспериментах методами машинного обучения

*на английском языке*

Modeling samples of random events with the account of prior knowledge in astrophysical experiments using machine learning methods

## 1.2. Приоритетное направление развития науки, технологий и техники в Российской Федерации, критическая технология

Указывается согласно перечню (Указ Президента Российской Федерации от 7 июля 2011 года №899) в случае, если тематика проекта может быть отнесена к одному из приоритетных направлений, а также может внести вклад в развитие критических технологий Российской Федерации.

3. Информационно-телекоммуникационные системы.

18. Технологии и программное обеспечение распределенных и высокопроизводительных вычислительных систем.

**Направление из Стратегии научно-технологического развития Российской Федерации (утверждена Указом Президента Российской Федерации от 1 декабря 2016 г. № 642 «О Стратегии научно-технологического развития Российской Федерации») (при наличии)**

## 1.3. Ключевые слова (приводится не более 15 терминов)

*на русском языке*

машинное обучение, моделирование физических явлений, генеративные нейронные сети, Монте-Карло

*на английском языке*

machine learning, modeling of physics phenomena, generative neural networks, Monte Carlo

## 1.4. Аннотация проекта (объемом не более 2 стр.; в том числе кратко – актуальность решения указанной выше научной проблемы и научная новизна)

Данная информация может быть опубликована на сайте Фонда в информационно-телекоммуникационной сети «Интернет».

*на русском языке*

Научной проблемой, на решение которой направлен проект, является исследование и разработка методов машинного обучения для анализа моделей физических явлений с использованием выборок случайных состояний (семплов). До последнего времени практически единственным подходом, позволяющим численно моделировать очень сложные физические явления, являлся подход на основе методов Монте-Карло. Однако использование реалистичных физических моделей на основе метода Монте-Карло требуют весьма больших временных и ресурсных затрат. Одним из альтернативных и весьма перспективных методов является использование методов машинного обучения для генерации выборок случайных векторов.

Отличительной особенностью моделирования физических явлений, особенно в физике элементарных частиц и космических лучей, является необходимость воспроизвести состояния системы с учетом ограничений, определяемых природой физического явления. Априорные знания могут иметь форму логических правил, алгебраических или дифференциальных уравнений (например законы сохранения или уравнения связей между параметрами или степенями свободы, описывающими изучаемые явления), наличия симметрий относительно дискретных или непрерывных групп входных/выходных данных, а также заранее известные вероятностные соотношения между данными. В методе Монте-Карло такие ограничения вводятся на уровне математической модели. Для методов машинного обучения, основанных на принципе обучения, такие ограничения воспроизводятся с трудом. Особенно это касается различных статистических распределений по совокупности состояний. Эти проблемы можно решить путем дополнительного включения априорных знаний о таких ограничениях в процесс обучения. В этом состоит новизна предлагаемого подхода к решению общей проблемы – эффективного генерирования выборок случайных состояний (семплов) с учетом априорных знаний. Практическая проверка разрабатываемых подходов будет применена для анализа моделей физических явлений в астрофизике частиц.

Конкретной задачей данного проекта в рамках общей проблемы является исследование и разработка методов

глубокого обучения генерации выборок многомерных случайных векторов в пространствах с большим числом измерений, с учетом априорной детерминистской и вероятностной информации о генерируемых событиях и их выборках для моделирования сложных физических явлений на примере экспериментов в области наземной гамма-астрономии. Наземная гамма-астрономия изучает излучение гамма-квантов галактического и внегалактического происхождения с помощью специально разработанных установок, так называемых атмосферных черенковских гамма-телескопов (Imaging Atmospheric (or Air) Cherenkov Telescope, IACT). С помощью этого метода гамма-излучение регистрируется на земле оптически как черенковский свет, порождаемый обширными ливнями вторичных частиц, когда гамма-кванты очень высокой энергии попадают в атмосферу. Гамма-кванты таких энергий составляют лишь малую долю (меньше одной десяти тысячной) потока космических лучей, состоящего, в основном, из протонов. Для анализа данных, планирования эксперимента и других задач необходимо проводить моделирование, результатом которого должны стать сотни тысяч и миллионы событий, включая протонные события, являющиеся фоном. На разработку эффективных алгоритмов машинного обучения для генерации семплов и направлен предложенный проект. Решение данной проблемы является весьма важным для гамма-астрономии, поскольку благодаря отсутствию электрического заряда, гамма-кванты несут информацию об их сверхудаленных источниках, в которых происходят экзотические и экстремальные процессы во Вселенной.

Интеграция в процесс обучения априорных, заранее известных знаний позволит заметно повысить качество полученных выборок, а следовательно и качество моделирования событий, в частности, событий широких атмосферных ливней, регистрируемых черенковскими телескопами. Практическими задачами проекта будут: исследование существующих и разработка новых методов машинного обучения для генерации случайных векторов с учетом априорной информации; разработка алгоритмов на их основе и их программная реализация; исследование, сравнение, выбор наилучшего (или лучших) методов; сравнение с существующими методами. В ходе осуществления проекта будет проведен полный цикл исследований и решения поставленных задач — от теоретической разработки подхода до практической программной реализации.

Помимо базового примера из области астрофизических экспериментов, генераторы событий на основе методов машинного обучения найдут важные применения в других научных и прикладных областях, связанных с анализом сложных и многомерных вероятностных явлений. Например, в области гидродинамики при исследовании процессов, связанных с турбулентностью, в метеорологии, в исследованиях поведения больших групп людей или животных в критических обстоятельствах и других. Из-за взрывного роста объема разнообразных данных, которые требуют анализа, в том числе с помощью существующих и новых методов машинного обучения, проблема генерации выборок случайных состояний (семплов событий), соответствующих априорным знаниям, в случае пространств с большим числом измерений является весьма актуальной и важной.

*на английском языке*

The scientific problem to be solved under the project is the research and development of deep machine learning methods for analyzing models of physical phenomena using samples of random states. Until recently, practically the only approach that allows one to numerically simulate very complex physical phenomena was the approach based on Monte Carlo methods. However, the use of realistic physical models based on the Monte Carlo method is very time and resource intensive. One of the alternative and very promising methods is the use of machine learning methods to generate samples of random vectors.

A distinctive feature of modeling physical phenomena, especially in the physics of elementary particles and cosmic rays, is the need to reproduce the states of the system, taking into account the constraints determined by the nature of the physical phenomenon. Prior knowledge can take the form of logical rules, algebraic or differential equations (for example, conservation laws or equations of relationships between parameters or degrees of freedom that describe the phenomena under study), the presence of symmetries with respect to discrete or continuous groups of input/output data, as well as known beforehand probabilistic relationships among data. In the Monte Carlo methods, a prior knowledge is introduced at the level of the mathematical model. For machine learning methods, such constraints are difficult to reproduce. This is especially true for various statistical distributions over a set of states. These problems can be solved by additionally incorporating prior knowledge of such constraints into the learning process. This is the novelty of the proposed approach to solving the general problem, namely the efficient generation of samples of random states taking into account prior knowledge. Practical verification of the developed approaches will be carried out by applying them to analysis of models of physical phenomena in astroparticle physics.

The specific task of this project within the framework of the general problem is to research and develop methods of deep learning for generating samples of multidimensional random vectors in spaces with a large number of dimensions, taking into account a priori deterministic and probabilistic information about generated events and their samples for modeling complex physical phenomena using the example of experiments in the field of terrestrial gamma-ray astronomy. Terrestrial gamma-ray astronomy studies the emission of gamma quanta of galactic and extragalactic origin using specially designed installations, the so-called atmospheric (or air) Cherenkov gamma telescopes (IACT). With this method, gamma rays are optically detected on the ground as Cherenkov light, generated by extensive showers of secondary particles when very high energy gamma rays enter the atmosphere. Gamma quanta of such energies make up only a small fraction (less than one ten-thousandth) of the cosmic ray flux, which consists mainly of protons. For data analysis, experiment planning and other tasks, it is necessary to carry out simulations, the result of which should be hundreds of thousands and even millions of events, including proton events that are the background. The proposed project is aimed at developing effective machine learning algorithms for generating samples. The solution to this problem is very important for gamma astronomy, since due to the absence of an electric charge, gamma quanta carry information about their ultra-distant sources, in which exotic and extreme processes in the Universe take place.

Integration of a priori, previously known knowledge into the learning process will significantly improve the quality of the obtained samples, and, consequently, the quality of event modeling, in particular, events of extensive air showers recorded by Cherenkov telescopes. The practical tasks of the project will be: research of existing and development of new methods of deep machine learning for generating random vectors taking into account prior informations; development of algorithms based on them and their software implementation; research, comparison, selection of the best methods; comparison with existing methods. During the implementation of the project, a full cycle of research and solutions of the assigned tasks will be carried out, from theoretical development of the approach to practical software implementation.

In addition to the basic example from the field of astrophysical experiments, event generators based on deep machine learning methods will find important applications in other scientific and applied areas related to the analysis of complex and multidimensional probabilistic phenomena. For example, in the field of hydrodynamics for the study of processes associated with turbulence, in meteorology, in the study of the behavior of large groups of people or animals in critical circumstances, and others. Due to the explosive growth in the amount of various data that require analysis, including that based on existing and new methods of deep machine learning, the problem of generating samples of random states corresponding to prior knowledge, in the case of spaces with a large number of dimensions, is very relevant and important.

#### **1.5. Ожидаемые результаты и их значимость (указываются результаты, их значимость для развития новой научной тематики)**

Данная информация может быть опубликована на сайте Фонда в информационно-телекоммуникационной сети «Интернет».

*на русском языке*

Основными результатами данного проекта будут методы интеллектуального моделирования данных в астрофизических экспериментах на основе машинного обучения. К ним относятся методы быстрого моделирования событий регистрации широких атмосферных ливней черенковскими телескопами с использованием генеративных нейронных сетей. Их научная значимость состоит в том, что использование таких методов позволит заменить сложные и ресурсоемкие методы Монте-Карло, существенно ускорить процесс моделирования, что, в свою очередь, позволит существенно продвинуться в понимании фундаментальных процессов, происходящих во Вселенной. Разрабатываемые методы могут найти применение не только в научных исследованиях, но и в инженерных науках, в экономике, везде, где используются методы Монте-Карло для моделирования различных явлений и необходимо учитывать априорную информацию, присущую этим явлениям. Конкретно, в рамках проекта будут получены следующие основные результаты:

- методы и алгоритмы на основе машинного обучения генерации случайных выборок многомерных векторов с учетом априорных ограничений, в том числе, статистических характеристик полученных наборов (семплов);
- программная адаптация существующих и реализация новых методов и алгоритмов генерации случайных выборок;
- результаты вычислительных экспериментов по сравнению Монте-Карло генераторов и генераторов основанных на машинном обучении с учетом априорной информации на примере моделирования событий в гамма-астрономии.

*на английском языке*

The main results of this project will be methods of intelligent data modeling in astrophysical experiments based on deep machine learning. These include methods for fast modeling of events of registration of large air showers by Cherenkov telescopes using generative neural networks. Their scientific significance lies in the fact that the use of such methods will allow replacing complex and resource-intensive Monte Carlo methods, significantly speeding up the modeling process,

which, in turn, will make it possible to significantly advance in understanding the fundamental processes occurring in the Universe. The developed methods can find application not only in scientific research, but also in engineering sciences, in economics, wherever Monte Carlo methods are used to model various phenomena and it is necessary to take into account a prior information inherent in these phenomena.

Specifically, within the framework of the project, the following main results will be obtained:

- methods and algorithms based on deep machine learning for generating random samples of multidimensional vectors, taking into account prior constraints, including the statistical characteristics of the obtained sets (samples);
- software adaptation of existing and implementation of new methods and algorithms for generating random samples;
- the results of computational experiments comparing Monte Carlo generators and generators based on machine learning, taking into account a prior information on the example of event modeling in gamma astronomy.

#### **1.6. В состав научного коллектива (в т.ч. с учетом руководителя проекта) будут входить (указывается планируемое количество исполнителей в течение всего срока реализации проекта):**

Несоответствие состава научного коллектива (в том числе отсутствие информации в соответствующих полях формы) требованиям пункта 12 конкурсной документации является основанием недопуска заявки к конкурсу.

**4** исполнителей проекта (включая руководителя),

В соответствии с требованиями пункта 12 конкурсной документации от 2 до 4 человек, вне зависимости от того, в трудовых или гражданско-правовых отношениях исполнители состоят с организацией.

В том числе:

- 2** исполнителей в возрасте до 39 лет включительно;
- 0** аспирантов (адъюнктов, интернов, ординаторов) очной формы обучения;
- 2** студентов очной формы обучения.

#### **1.7. Планируемый состав научного коллектива с указанием фамилий, имен, отчеств (при наличии) членов коллектива, их возраста на момент подачи заявки, ученых степеней, должностей и основных мест работы, формы отношений с организацией (трудовой договор, гражданско-правовой договор) в период реализации проекта**

1. Крюков Александр Павлович, 67 лет, к.ф.-м.н., МГУ имени М.В.Ломоносова, заведующий лабораторией, трудовой договор.
2. Дубенская Юлия Юрьевна, 39 лет, МГУ имени М.В.Ломоносова, н.с., трудовой договор.
3. Власкина Анна Александровна, 21 год, физический факультет МГУ имени М.В.Ломоносова, бакалавр, трудовой договор.
4. Гресь Елизавета Олеговна, 22 года, физический факультет ИГУ, студент, трудовой договор.

#### **Соответствие профессионального уровня членов научного коллектива задачам проекта**

Предложенный проект посвящен, в основном, информационным технологиям, в частности одному из самых его бурно развивающихся разделов – интеллектуальным методам анализа больших данных. Однако, проект имеет также сильно выраженный междисциплинарный характер, а его результаты ориентированы на одну из самых передовых областей научных исследований – астрофизику субатомных частиц, которая занимается изучением самых глубоких основ строения Вселенной. Поэтому в коллективе участвуют специалисты в области как ИТ, так и физики высоких энергий, включая гамма-астрономию.

Руководитель коллектива А.П.Крюков профессионально занимается развитием и применением современных информационных технологий в физике. Им были развиты многочисленные математические методы в области компьютерной алгебры, в том числе для задач физики высоких энергий. Он являлся пионером во внедрении грид технологии в России, выполнил многочисленные исследования в области распределенных вычислений. Имеет интересные физические результаты. А.П.Крюков является членом авторского коллектива коллаборации CMS (ЦЕР, Женева), которая в 2012 году открыла бозон Хиггса (в 2013 году нобелевская премия по физике была присуждена П.Хиггсу и Ф.Энглеру за теоретическое обоснование существования бозона Хиггса). В последние годы он стал активно развивать различные методы машинного обучения и их применения в физике, в частности в гамма-астрономии. Является членом международной коллаборации TAIGA, основной задачей которой является исследование в области физики космических лучей и гамма-астрономии.

Ю.Ю.Дубенская является опытным исследователем, принимала участие во многих проектах по распределенным вычислениям и использованию технологии виртуализации в области высокопроизводительных вычислений. В настоящее

время занимается развитием методов генерации состояний физических систем с помощью генеративно-состязательных нейронных сетей. Результаты докладывались на ряде международных конференций и опубликованы в печати.

А.А.Власкина — бакалавр физического факультета МГУ имени М.В.Ломоносова, которая начинает свой путь в науке. Тем не менее, уже сейчас она добилась хороших результатов в области машинного обучения. Ее курсовая работа посвящена использованию сверточных автоэнкодеров для улучшения отношения сигнал/шум для изображений широких атмосферных ливней, порожденных гамма квантами от галактических и внегалактических источников. В настоящее время по материалам курсовой работы готовится статья для публикации.

Е.О.Гресь — магистрант ИГУ, г.Иркутск. Несмотря на свой возраст уже имеет опыт обработки данных эксперимента TAIGA. В своей бакалаврской работе «Обработка первичных данных ШАЛ Тункинского телескопа TAIGA-IACT» она разработала программу обработки данных телескопа на основе параметров Хилласа, что позволило заметно улучшить качество отбора гамма-событий в эксперименте TAIGA. В настоящее время она занимается задачей классификации событий методом сверточных нейронных сетей. Полученные результаты планируется доложить на международном рабочем совещании «Deep Learning in Computational Physics».

Таким образом, коллектив имеет в своем составе опытных исследователей и студентов, сочетая в себе опыт и знания как в области информационных технологий, включая машинное обучение, так и в области физики высоких энергий, включая гамма-астрономию, методы для которой будут разрабатываться, апробироваться и исследоваться. Все это является залогом успешного выполнения программы исследований и получение заявленных результатов, предусмотренных проектом.

#### **1.8. Планируемый объем финансирования проекта Фондом по годам (указывается в тыс. рублей):**

в 2022 г. - 1500 тыс. рублей,

в 2023 г. - 1500 тыс. рублей.

Несоответствие планируемого объема финансирования проекта (в том числе отсутствие информации в соответствующих полях формы) требованиям пункта 10 конкурсной документации является основанием недопуска заявки к конкурсу.

#### **1.9. Научный коллектив по результатам проекта в ходе его реализации предполагает опубликовать в рецензируемых российских и зарубежных научных изданиях не менее**

Приводятся данные за весь период выполнения проекта. Уменьшение количества публикаций (в том числе отсутствие информации в соответствующих полях формы) по сравнению с порогом, установленным в пункте 16.2 конкурсной документации, является основанием недопуска заявки к конкурсу.

**5** публикаций,

из них

**5** в изданиях, индексируемых в базах данных «Сеть науки» (Web of Science Core Collection) или «Скопус» (Scopus).

**Информация о научных изданиях, в которых предполагается опубликовать результаты проекта, в том числе следует указать в каких базах индексируются данные издания - «Сеть науки» (Web of Science Core Collection), «Скопус» (Scopus), РИНЦ, иные базы, а также указать тип публикации - статья, обзор, тезисы, монография, иной тип**

Результаты проекта планируется опубликовать в ряде международных и российских изданиях, в том числе:

- Journal of Computational Physics,
- Neural Networks,
- Neural Computing and Applications
- Journal of Physics: Conference series
- Программирование
- CEUR Wokrshops

Все указанные издания индексируются системами WoS или Scopus. Точный список изданий будет определен в процессе выполнения проекта.

#### **Иные способы обнародования результатов выполнения проекта**

Выступления на ведущих всероссийских и международных конференциях, в том числе следующих регулярных форумах:

- International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)
- Параллельные вычислительные технологии (ПаВТ), Россия;

-- Distributed Computing and Grid-technologies in Science and Education (GRID), ОИЯИ, Дубна,

-- Суперкомпьютерные дни в России (RSCD), Москва.

Полный список конференций будет определен позднее.

#### **1.10. Число публикаций членов научного коллектива, опубликованных в период с 1 января 2016 года до даты подачи заявки,**

**53**, из них

**53** – опубликованы в изданиях, индексируемых в Web of Science Core Collection или в Scopus.

#### **1.11. Планируемое участие научного коллектива в международных коллаборациях (проектах) (при наличии)**

В процессе выполнения работ по проекту планируется активное участие в международной коллаборации TAIGA (<https://theory.sinp.msu.ru/doku.php/taiga/about>), в которую входят ведущие российские и европейские научные организации и университеты, в том числе НИИЯФ МГУ; Иркутский государственный университет; Объединенный институт ядерных исследований (ОИЯИ, г.Дубна); Национальный исследовательский ядерный университет "МИФИ"; Институт ядерных исследований РАН; Немецкий электронно-синхротронный центр (Deutsches Elektronen-Synchrotron), DESY, Германия; Туринский университет (Universita' degli Studi di Torino), Италия; Институт технологий (Karlsruher Institut für Technologie), KIT, г. Карлсруэ, Германия; Гамбургский университет (University of Hamburg), УНН, Германия; и ряд других научных организаций и университетов. Основной задачей этой коллаборации является поиск галактических источников гамма-квантов с энергиями выше 20-30 ТэВ; исследования потоков гамма-излучения от известных источников в диапазоне энергий выше 20-30 ТэВ на регистрируемом уровне чувствительности; исследования высокоэнергетической части спектра гамма-излучения наиболее ярких блазаров с целью изучения поглощения гамма-квантов на межгалактическом фоновом излучении (инфракрасном и микроволновом) и поиска возможных нарушений лоренц-инвариантности и аксион-фотонных переходов, а также поиска темной материи во Вселенной. Ключевой идеей развития гамма-обсерватории TAIGA является совместная работа широкоугольных и узкоугольных детекторов Тунка-HiSCORE и TAIGA-IACT. TAIGA – проект класса мегасайенс, находится на территории Тункинского астрофизического центра коллективного пользования Иркутского государственного университета где строится крупнейшая в мире по площади гамма-обсерватория. Руководитель проекта является членом указанной коллаборации.

Руководитель проекта подтверждает, что

- все члены научного коллектива (в том числе руководитель проекта) удовлетворяют пунктам 4, 6, 7, 13 конкурсной документации;
- на весь период реализации проекта руководитель проекта будет состоять в трудовых отношениях с организацией;
- при обнародовании результатов любой научной работы, выполненной в рамках поддержанного Фондом проекта, руководитель проекта и научный коллектив будут указывать на получение финансовой поддержки от Фонда и организацию, а также согласны с опубликованием Фондом аннотации и ожидаемых результатов поддержанного проекта, соответствующих отчетов о выполнении проекта, в том числе в информационно-телекоммуникационной сети «Интернет»;
- помимо гранта Фонда проект не будет иметь других источников финансирования в течение всего периода практической реализации проекта с использованием гранта Фонда;
- проект не является аналогичным по содержанию проекту, одновременно поданному на конкурсы научных фондов и иных организаций;
- проект не содержит сведений, составляющих государственную тайну или относимых к охраняемой в соответствии с законодательством Российской Федерации иной информации ограниченного доступа;
- доля членов научного коллектива в возрасте до 39 лет включительно в общей численности членов научного коллектива будет составлять не менее 50 процентов в течение всего периода практической реализации проекта;
- в установленные сроки будут представляться в Фонд ежегодные отчеты о выполнении проекта и о целевом использовании средств гранта.

Подпись руководителя проекта \_\_\_\_\_ /А.П. Крюков/

## Форма 3. Сведения об организации

собираются автоматически на основе регистрационных данных организации, через которую будет осуществляться финансирование ("Форма Т")

### 3.1. Полное наименование *(приводится в соответствии с регистрационными документами)*

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В.Ломоносова»

### 3.2. Сокращенное наименование

Московский государственный университет имени М.В.Ломоносова, Московский университет или МГУ

### 3.3. Наименование на английском языке

Federal State Budgetary Educational Institution of Higher Education Lomonosov Moscow State University

### 3.4. Организационно-правовая форма *(указывается по ОКOPФ)*

Федеральные государственные бюджетные учреждения

### 3.5. Форма собственности *(указывается по ОКФС)*

Федеральная собственность

### 3.6. Ведомственная принадлежность

Правительство Российской Федерации

### 3.7. ИНН, КПП, ОГРН, ОКТМО

7729082090, 772901001, 1037700258694, 45325000

### 3.8. Адрес

119991, г. Москва, Ленинские горы, д. 1

### 3.9. Фактический адрес

119991, г. Москва, Ленинские горы, д. 1

### 3.10. Субъект Российской Федерации

г Москва

### 3.11. Должность, фамилия, имя, *отчество (при наличии)* руководителя организации

Ректор, Садовничий Виктор Антонович

### 3.12. Контактный телефон

+74959394647

### 3.13. Электронный адрес *(E-mail)*

info@rector.msu.ru

### Руководитель организации подтверждает, что:

- ознакомлен с условиями конкурса Фонда и согласен на финансирование проекта, в случае его поддержки, через организацию;
- согласен с пунктами 8, 14, 33, 35, 36 конкурсной документации, иными условиями конкурса;
- подтверждает сведения о руководителе проекта, изложенные в данной заявке;
- организация исполняет обязательства по уплате налогов в бюджеты всех уровней и обязательных платежей в государственные внебюджетные фонды, платежеспособна, не находится в процессе ликвидации, не признана несостоятельной (банкротом), на ее имущество не наложен арест и ее экономическая деятельность не приостановлена;
- в случае признания заявки победителем организация берет на себя следующие обязательства:
  - заключить с членами научного коллектива гражданско-правовые или трудовые (срочные трудовые) договоры;  
Если таковые не заключены ранее. В случае, если член научного коллектива не является гражданином Российской Федерации, организацией должны быть выполнены все процедуры, предусмотренные законодательством Российской Федерации при трудоустройстве иностранных граждан.
  - по поручению руководителя проекта выплачивать членам научного коллектива вознаграждение за

- выполнение работ по проекту;
- о ежегодно в установленные сроки представлять отчет о целевом использовании гранта Российского научного фонда.

**Руководитель организации гарантирует, что:**

- вознаграждение за выполнение работ по реализации проекта будет ежегодно получать каждый член научного коллектива;  
Лица, не являющиеся налоговыми резидентами Российской Федерации, могут осуществлять работы по проекту на безвозмездной основе (за исключением руководителя проекта).
- общий размер ежегодного вознаграждения членов научного коллектива в возрасте до 39 лет включительно не будет меньше 35 процентов от суммы ежегодного вознаграждения всех членов научного коллектива;
- общее число членов научного коллектива (вместе с руководителем проекта) будет составлять от 2 до 4 человек, при этом членом научного коллектива не будет являться работник организации, в непосредственном административном подчинении которого находится руководитель проекта;
- научному коллективу будет предоставлено помещение и обеспечен доступ к имеющейся экспериментальной базе для осуществления научного исследования.

**Подпись руководителя организации** (уполномоченного представителя, действующего на основании доверенности или распорядительного документа), **печать** (при ее наличии) **организации.**

В случае подписания формы уполномоченным представителем организации (в т.ч. – руководителем филиала) к печатному экземпляру заявки прилагается копия распорядительного документа или доверенности, заверенная печатью организации.

\_\_\_\_\_/\_\_\_\_\_/\_\_\_\_\_  
М.П.

## Форма 4. Содержание проекта

### 4.1. Научная проблема, на решение которой направлен проект

Фундаментальной научной проблемой, на решение которой направлен проект, является моделирование физических явлений. Современный физический эксперимент настолько сложен, что планирование эксперимента, отработка методики анализа данных и интерпретация его результатов возможна только путем сравнения модельных данных с результатами эксперимента с привлечением сложных программных комплексов и высокопроизводительных систем. Таким образом ключевым моментом таких исследований является моделирование результатов эксперимента с помощью математических и программных моделей, что является основной целью данного проекта. До настоящего времени большинство программ, генерирующие модельные данные (семплы), были основаны на методах Монте-Карло, которые использовали алгоритмы, моделирующие реальные физические процессы в исследуемой области. Несмотря на универсальность метода Монте-Карло такой подход требует разработки очень сложных алгоритмов и использования значительных вычислительных ресурсов и времени. Таким образом исследование и разработка новых быстрых и эффективных методов на альтернативных принципах является весьма актуальной задачей. В представленном проекте в качестве такой альтернативы предлагается применить современные методы машинного обучения для генерации выборок случайных состояний физической системы в случае пространств состояний с большим числом измерений. Особенностью предлагаемого подхода является не только использование машинного обучения, но и разработка методов, позволяющих учитывать априорную информацию, которая присуща физическим явлениям и проявляющаяся в виде законов сохранения, различных свойств симметрии, а также статистических распределений наблюдаемых величин. В результате исследования будут получены семплы на примере данных в гамма-астрономии, которые пройдут валидацию путем сравнения их с экспериментом.

### 4.2. Научная значимость и актуальность решения обозначенной проблемы

Математическое и программное моделирование является в настоящее время одним из ключевых технологий в современной науке и на производстве. Одним из основных методов в этой области является метод Монте-Карло, который широко используется для моделирования разнообразных физических явлений и сред. Центральной задачей такого моделирования является генерация выборок (семплов) случайных векторов, которые соответствуют состоянию системы в пространстве ее состояний. Важной особенностью таких выборок является то, что они по отдельности и в совокупности обладают рядом априорных свойств. Так, например, может оказаться, что в процессе моделирования физических явлений должны выполняться законы сохранения некоторых величин, возможно наличие различных симметрий. С другой стороны, часто целые выборки должны статистически удовлетворять некоторой априорной информации в целом, например вероятности появления тех или иных векторов (состояний). Как правило, в традиционных методах моделирования, основанных на методах Монте-Карло, такая априорная информация закладывается в саму модель и выполняется автоматически. К сожалению, реализация методов Монте-Карло требует разработки сложных и весьма ресурсоемких математических и программных моделей. При этом зачастую возникают проблемы, связанные с эффективным охватом всего многомерного пространства состояний или сложным и, следовательно, очень медленным анализом данных и моделированием взаимодействия физических систем с измерительными приборами (детекторами).

Новые проблемы, связанные с текущими и будущими экспериментами побуждают исследовательское сообщество вкладывать средства в дальнейшие теоретические и технические улучшения этих важных инструментов. Поэтому исключительно важно адаптировать существующие и разработать новые методы, которые позволят существенно увеличить эффективность генерации выборок, что, в свою очередь, обеспечит заметное сокращение ресурсных и временных затрат для получения научных и технических важных результатов. Как показывают недавние исследования, по крайней мере часть этих важных проблем можно решить, если методы Монте-Карло дополнить инструментарием, созданным в рамках подхода, использующего методы машинного обучения, в частности генеративные нейронные сети.

Проиллюстрируем это на конкретном примере из области астрофизики элементарных частиц, которая превратилась в науку с большим объемом данных, исчисляемых сотнями и тысячами терабайт, и с большим числом измеряемых параметров, связанных с каждым наблюдением. В то время как 10-15 лет назад астрофизические установки собирали 1-10 Тб данных в год, новые экспериментальные установки собирают от 100 до 1000 терабайт данных в год. Более того, ожидается, что в следующие десятилетия с помощью новых и более сложных научных инструментов, а также результатов моделирования данных, необходимых для физической интерпретации, будут созданы новые очень

сложные и массивные наборы данных. Обработка и изучение этих данных и проведение научных исследований представляет собой серьезную научную задачу, которая требует принятия новых подходов к обработке и анализу экспериментов. В частности, исследование потоков гамма-квантов, порождаемых галактическими и внегалактическими источниками, является одним из наиболее многообещающих способов изучения самих источников, а значит и важных общих свойств Вселенной. Дело в том, что заряженные космические лучи подвергаются значительному влиянию галактических и межгалактических магнитных полей, что приводит к сильному искажению их траекторий и, как следствие, потере какой-либо информации о месте их возникновения. Поэтому в последние годы бурно развивались гамма-астрономия и нейтринная астрономия. Действительно, так как гамма-кванты и нейтрино являются электрически нейтральными частицами, то они могут быть использованы как источник информации о строении астрофизических объектов, в которых они образовались. Все выше сказанное говорит о большой актуальности разработки новых эффективных методов генерации модельных семплов для указанных областей науки.

В настоящее время в России активно развивается экспериментальная астрофизика и гамма-астрономия. В частности, в последние годы в Тункинской долине в 50 км от оз. Байкал в рамках проекта по созданию гамма-обсерватория TAIGA (Tunka Advanced Instrument for cosmic ray physics and Gamma Astronomy) с гибридной системой совместно работающих детекторов развивается новый подход к исследованиям гамма-излучения сверхвысоких энергий, основанный на многоканальном анализе данных. TAIGA – проект разряда мегасайенс, реализуемый международной коллаборацией. Различные детекторы шести установок гамма-обсерватории TAIGA регистрируют черенковское и радиоизлучения, а также электронную и мюонную компоненты ШАЛ. Основная задача гамма-обсерватории TAIGA - исследование высокоэнергетической части спектра гамма-излучения с целью получения новых знаний о структуре и свойствах Вселенной. Применение новых методов, основанных на методах машинного обучения позволит успешно справиться с современными вызовами, связанными с расширением старых и введением в строй новых экспериментальных установок.

Хотя выше обсуждалась астрофизика частиц, на примере которой предполагается опробовать разработанные подходы, все сказанное полностью справедливо и для других научных и прикладных областей, связанных с анализом сложных и многомерных вероятностных явлений. Например, в области гидродинамики при исследовании процессов, связанных с турбулентностью, в метеорологии, в исследованиях поведения больших групп людей или животных в критических обстоятельствах.

Таким образом, решение обозначенной проблемы имеет большую научную значимость и актуальность.

#### **4.3. Конкретная задача (задачи) в рамках проблемы, на решение которой направлен проект, ее масштаб и комплексность**

Конкретной задачей данного проекта в рамках общей проблемы является исследование и разработка методов глубокого обучения генерации выборок многомерных случайных векторов в пространствах с большим числом измерений, с учетом априорной детерминистской и вероятностной информации о генерируемых событиях и их выборках для моделирования сложных физических явлений на примере астрофизики частиц, точнее – моделирования изображений ШАЛ в атмосферных черенковских телескопах (IACT). Интеграция в процесс обучения априорных, заранее известных знаний позволит заметно повысить качество получаемых результатов и, как следствие, качество полученных выборок, а следовательно и качество моделирования изображений ШАЛ.

Во многих случаях подходы, основанные исключительно на обучающих наборах данных, могут достичь своих пределов или привести к неудовлетворительным результатам. В частности модель, основанная исключительно на данных, может не соответствовать существующим ограничениям, диктуемым естественно-научными законами или геометрией измерительных приборов. Эти проблемы можно решить путем дополнительного включения априорных знаний в процесс обучения. Априорные знания могут иметь форму логических правил, алгебраических или дифференциальных уравнений (например законы сохранения или уравнения связей между параметрами или степенями свободы, описывающими изучаемые явления), наличия симметрий относительно дискретных или непрерывных групп Ли, включая инвариантность, входных/выходных данных для нейросетей, а также возможные заранее известные вероятностные соотношения между данными.

Таким образом, практическими задачами проекта будут:

- исследование существующих и разработка новых методов машинного обучения для генерации случайных векторов с учетом априорной информации;

- разработка алгоритмов на их основе и их программная реализация;
- исследование, сравнение, выбор наилучшего (или лучших) методов;
- сравнение с существующими методами;
- практическая апробация разработанных методов на примере задачи о генерации изображений ШАЛ в атмосферных черенковских телескопах с учетом симметрии (вращательно-симметричной геометрии телескопа), алгебраических ограничений на параметры изображений и привязкой к заданному спектру первичных частиц ШАЛ.

Как видно из вышесказанного, в ходе осуществления проекта будет проведен полный цикл исследований и решения поставленных задач — от теоретической разработки подхода до практической программной реализации.

Результаты проекта найдут применения в различных прикладных областях, связанных с численным моделированием. Примерами таких областей, помимо астрофизики частиц и физики субатомных частиц, являются моделирование планетарной динамики в звездных скоплениях, учет турбулентности в моделях предсказания погоды, моделирование взаимодействий белков в биологических системах. Это обуславливает масштаб и комплексность задач проекта.

#### **4.4. Научная новизна исследований, обоснование того, что проект направлен на развитие новой для научного коллектива тематики\*\*\*\*\*, обоснование достижимости решения поставленной задачи (задач) и возможности получения предполагаемых результатов**

\*\*\*\*\* В том числе, на определение объекта и предмета исследования, составление плана исследования, выбор методов исследования.

Научная новизна задач, которые будут решены в рамках заявляемого проекта, состоит в разработке способа явного учета априорных знаний в генераторах событий на основе нейронных сетей. До настоящего времени все исследования в данной области основывались на использовании обучающего набора данных для генеративных сетей. При этом предполагалось, что помимо других свойств, такой обучающий набор неявно содержит информацию и об априорных знаниях и, таким образом сеть в процессе обучения может в какой-то мере учитывать эту информацию. Затем экспериментально проверялось выполнение соответствующих ограничений. Проверки показали не только сильную зависимость от уровня репрезентативности обучающей выборки, а также гиперпараметров нейронных сетей, но и слабый учет неявно заданной априорной информации. В любом случае априорные ограничения выполняются лишь приближенно, а точность их выполнения не контролируется. Другими словами, в таком подходе нет гарантии того, что результаты генерации событий следуют априорным знаниям (ограничениям) с нужной точностью, а это часто является критическим для физических приложений.

В рамках заявляемого проекта будут теоретически и экспериментально изучены существующие, а также будут разработаны новые методы явного учета априорных знаний (ограничений и взаимосвязей). В результате впервые будут получены оценки улучшения работы генераторов событий на основе результатов экспериментальных исследований работы таких генераторов без учета и с учетом априорных знаний. Причем будут рассмотрены различные варианты и методы учета таких знаний (см. п. 4.6). Основным научным результатом будет существенное увеличение качества и быстродействия работы разработанных генераторов событий на основе машинного обучения, порождающих выборки высокого качества.

Принципиально новым и многообещающим подходом, который предлагается реализовать в рамках данного проекта является явный учет статистических характеристик (распределений) обучающего набора (см. п. 4.6), что позволит создавать выборки событий удовлетворяющих сложным динамическим законам моделируемых явлений с контролируемой точностью воспроизведения этих характеристик. Разработанные методы будут апробированы на примере моделирования процессов в области астрофизики частиц и гамма-астрономии, для которой впервые будут созданы быстродействующие генераторы на основе генеративных сетей с учетом априорных знаний, в частности:

- симметрии ШАЛ и регистрирующих устройств (телескопов);
- априорных ограничений на параметры изображений ШАЛ;
- спектра энергий первичных частиц ШАЛ;
- связей энергии первичных частиц с пространственными свойствами ШАЛ.

Возможность достижимости решения поставленных задач и получения предполагаемых результатов основана на следующем:

- существующий высокий уровень развития теории и практики использования машинного обучения, включая генеративные сети, обширные результаты по качественным и количественным характеристикам сетей различного типа (обычные генеративно-сопоставительные сети (generative-adversarial networks; GAN), условные генеративно-

состязательные сети (conditional generative-adversarial networks; cGAN), вариативные автоэнкодеры и т. п.), сетей с различным выбором гиперпараметров, функций потерь и т. д.

- полученные другими исследователями первые результаты по использованию нейросетей для генерации событий; хотя эти попытки нельзя пока признать удовлетворительными для практического использования, но они дают основание полагать, что учет априорных данных позволит создать быстрые и эффективные генераторы;
- существующие в литературе результаты по учету априорных знаний в нейросетях глубокого обучения, в частности в генеративных нейросетях;
- имеющийся у коллектива исполнителей проекта научный задел по проекту и опыт совместной успешной реализации проектов, в том числе международных.

Хотя коллектив имеет хороший задел в области машинного обучения для задач астрофизики, например, задачи классификации первичных частиц космических лучей, определение их энергии, предложенный проект является новым и перспективным направлением исследований группы под руководством А.Крюкова, которое будет иметь прямой выход на реальный эксперимент - TAIGA.

#### **4.5. Современное состояние исследований по данной проблеме, основные направления исследований в мировой науке и научные конкуренты**

Основными направлениями исследований по теме проекта являются:

- развитие генераторов случайных векторов на основе глубокого обучения;
- развитие методов учета априорной информации при обучении глубоких нейросетей, в особенности генеративных нейросетей.

Задача генерации случайных векторов в пространствах с высокой размерностью рассматривалась в научной литературе как в абстрактно-математической постановке (см., например, [1], [2]), так и в рамках различных прикладных областей. С точки зрения приложений особенно интенсивно эта тематика развивалась в смежной с астрофизикой области физики субатомных частиц высоких энергий (ФВЭ), см., например, [3], [4] и ссылки в них. Так в недавней работе [1] предложен метод генерации случайных независимых векторов, которые имеют заданную непрерывную плотность распределения с компактным носителем. Основное преимущество предлагаемого метода - гарантированные оценки погрешности генерации случайных векторов. Приведено экспериментальное сравнение предложенного метода с хорошо известным алгоритмом Метрополиса-Гастингса. Предложенный метод лучше аппроксимирует заданное распределение, но оба метода являются весьма ресурсоемкими в случае высоких размерностей векторных пространств. Работа [2] использует подход "принятия/отклонения", который широко используется в универсальных генераторах неоднородных последовательностей случайных чисел. Его ключевой частью является приближение заданной плотности вероятности сверху кусочно-постоянной hat-функцией на элементах разбиения области пространства. В методе используется пилообразная оценка сверху липшицевых плотностей, а затем используются все локальные максимизаторы такой оценки. Метод применим к многомерным мультимодальным распределениям. Для относительно небольших размерностей пространств он демонстрирует относительно короткое время предварительной обработки и довольно быстрое генерирование случайных величин из очень большого класса распределений. Однако, с увеличением размерности время генерации очень быстро растет (в работе приведены экспериментальные оценки только для небольших размерностей, но уже при увеличении размерности от 3 до 9 время растет в 10 раз).

Работы [3], [4] содержат впечатляющий обзор успехов методов Монте-Карло (МК) для генераторов векторов в целях анализа экспериментальных данных в области ФВЭ. Однако, хотя высокопроизводительных вычислений (HPC) позволяют осуществлять численное моделирование сложных систем, связанное с большим объемом данных, но ограничения в увеличении производительности, которые стали более заметными за последние 5-10 лет, могут серьезно повлиять на дальнейший прогресс в этой области.

В недавней работе [5] подчеркивается, что необходимы новые подходы для повышения производительности при моделировании прикладных задач. Одним из таких подходов является глубокое обучение, которое в последнее время использовалось для улучшения решения задач, которые традиционно решаются с помощью крупномасштабного численного моделирования с использованием высокопроизводительных вычислений. В этой работе рассмотрены примеры применения нейросетей для моделирования планетарной динамики в звездных скоплениях, учета турбулентности в моделях предсказания погоды, моделирования взаимодействий белков в биологических системах, а также создание генераторов физических событий в физике субатомных частиц. Основным выводом работы является перспективность использования нейронных сетей для численного моделирования, но для их широкого практического

использования требуется дальнейшее совершенствование методики их применения.

В связи с этим в настоящее время разворачиваются интенсивные исследования в области машинного обучения для генерации выборок событий, причем пионерской прикладной областью опять выступает ФВЭ. Так в работах [6], [7] исследованы возможности генерации событий и частоты их появления с помощью генеративно-состязательных сетей (GAN) и вариационных автоэнкодеров (VAE) аналогично генераторам Монте-Карло. Изучен ряд демонстрационных конкретных процессов столкновения частиц. В работе [6] показано, что протестированные архитектуры GAN и стандартный VAE не могут точно обучиться и воспроизвести требуемые распределения. Однако путем буферизации информации о плотности закодированных кодировщиком VAE событий, сгенерированных методом Монте-Карло, удается построить данные для выборки новых событий из декодера, удовлетворительно согласующихся с событиями, полученными методом Монте-Карло, но скорость генерации таких событий нейросетями на несколько порядков больше. Необходимо отметить, однако, что хотя VAE можно использовать для генерации новых выборок данных, ключевым назначением автоэнкодеров является скрытое моделирование и маргинализация ненужных переменных, что не является задачей при генерации событий. В работе [7] пришлось использовать дополнительный к нейросетям инструментарий основанный на метрике максимального среднего расхождения (maximum mean discrepancy; MMD), чтобы правильно генерировать резкие локальные особенности распределений. В работе [8] для достижения удовлетворительного результата в разработанной Feature-Augmented and Transformed GAN (FAT-GAN) используется специальным образом преобразованный набор признаков событий. В работах [9],[10] как и в работе [7], используется вспомогательный инструментарий для обучения генерирующей состязательной сети на потоках частиц в многослойном электромагнитном калориметре, а также продемонстрированы пределы возможностей существующих генераторов на основе GAN-сетей.

В работах [11], [12],[13], [14] предприняты попытки использования генераторов, основанных на GAN-сетях, для анализа откликов детекторов субатомных частиц в ФВЭ. Продемонстрирована принципиальная способность GAN обучаться созданию реалистичных с точки зрения физики семплов с учетом зависимостей от параметров новых физических моделей и то, что применение нейросетей позволяет резко увеличить скорость моделирования событий. Однако показано, что требуется дальнейшее совершенствование методики, в частности для решения проблем стабильности. Аналогичный вывод сделан и в работе [15]: при использовании генеративно-состязательных сетей (GAN), возникают проблемы, связанные с областями, где распределения рассматриваемых величин демонстрируют резкие локальные особенности. При этом показано, что можно добиться существенного улучшения производительности и ускорения сходимости, включив члены регрессии в функцию потерь генератора. Очень важно, что в работе предложен объективный критерий для количественной оценки работы генераторов, который может быть использован в сравнения существующих и разрабатываемых генераторов. Построенная в работе [16] генеративная сеть (генератор) испытывает проблемы с достижением точности во всем пространстве событий, но текущее решение может воспроизводить различные свойства ливня частиц, демонстрируя при этом в  $\sim 10^5$  раз большую скорость генерации по сравнению со стандартными МК-методами. Это открывает дверь в новую эру быстрого моделирования, которое может значительно сэкономить время вычислений и дисковое пространство, одновременно расширяя возможности исследования физических явлений на ЛНС в рамках Стандартной модели и за ее пределами.

Общим выводом из перечисленных выше работ является то, что использование сетей глубокого обучения для генерации выборок случайных многомерных векторов (событий) является практически очень важным и перспективным направлением, но методика их использования требует существенного совершенствования. Для достижения этой цели в настоящем проекте предлагается развить метод машинного обучения с учетом априорной информации.

Стоит отметить, что постановка вопроса о необходимости учета априорной информации в методах машинного обучения для моделирования физических явлений в ФВЭ возникла и ранее. Например, вопрос о выполнении кинематических связей (например, условия релятивистского дисперсионного соотношения) при генерации событий ставился и в работах по генерации событий и без явного учета априорной информации (см., например, [6], [17]). Оказалось, что благодаря хорошим способностям глубоких сетей к обучению, такие связи приближенно выполняются. Однако, как показано в работе [18], явное введение связей существенно улучшает результаты обучения и резко повышают точность выполнения связей, что является очень важным в различных прикладных областях. Например, поля скоростей потоков жидкости должны быть бездивергентными из-за сохранения массы, а температурные поля обычно являются гладкими. В области ФВЭ связи порождаются дисперсионными соотношениями и законами сохранения (энергии-импульса, заряда и т. п.). Следовательно, необходимо обеспечить учет связей сетями GAN для того, чтобы их

можно было использовать как надежные симуляторы физических систем.

Теоретически доказано (в классической работе I. Goodfellow и др. [19]), что стандартные GAN способны воспроизводить связи, тем не менее, также хорошо известно, что традиционные GAN имеют трудности со сходимостью и недостаточную надежность в обучении и были предприняты многочисленные усилия для улучшения стабильности и надежности в обучении GAN, что привело к появлению ряда вариантов GAN. В работе [18], в частности, исследуется возможность использования физических условий связи для улучшения (регуляризации) обучения GAN (нахождения глобального минимума), используемых для моделирования физических систем, — а не только как дополнительное ограничение, которое должно быть удовлетворено. Кроме того, в работе большое внимание уделяется случаю приближенных условий связи (связей с допустимыми ошибками). Существует также дополнительная к работе [18] работа [20], где рассматриваются связи статистического типа. Связи последнего типа очень важны в контексте заявляемого проекта, поскольку генерируемые выборки событий должны удовлетворять заданной наперед статистике. В работах [18], [20] рассматриваются и сравниваются три типа GAN: обычные GAN, Conditional GAN (сGAN [21]) и Constrained GAN. Показано, что сети с явным учетом априорных знаний (связей) обладают более хорошими свойствами сходимости, а результаты генерации существенно лучше удовлетворяют требуемым связям. Приведены количественные характеристики для сравнения, которые показывают существенные преимущества сетей с учетом априорных ограничений.

Вопросы учета априорной информации рассматривались также в работах [22] - [25], результаты которых могут быть использованы при реализации заявляемого проекта. Существует целый ряд других статей по учету априорной информации в нейросетях, но их тематика и результаты достаточно далеки от конкретных задач заявляемого проекта, смотри, например, обзор von Rueden, Laura, et al. "Informed Machine Learning--A Taxonomy and Survey of Integrating Knowledge into Learning Systems." arXiv preprint arXiv:1903.12394 (2019).

В рамках заявляемого проекта методика учета априорной информации — детерминистские связи и вероятностные ограничения — будут исследованы, развиты и адаптированы для применения в генераторах многомерных случайных векторов (событий) с целью улучшения параметров их работы и качества получаемых выборок в гамма-астрономии. Особенно важен явный учет статистических характеристик выборок и, как следствие, воспроизведение их при генерации.

Список литературы.

1. Darkhovsky, Boris S., et al. "A Method of Generating Random Vectors with a Given Probability Density Function." *Automation and Remote Control* 79.9 (2018): 1569-1581.
2. Beliakov, Gleb. "Universal nonuniform random vector generator based on acceptance-rejection." *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 15.3 (2005): 205-232.
3. Perret-Gallix, Denis. "Computational particle physics for event generators and data analysis." *Journal of Physics: Conference Series*. Vol. 454. No. 1. IOP Publishing, 2013.
4. Papaefstathiou, Andreas. "How-to: write a parton-level Monte Carlo particle physics event generator." *The European Physical Journal Plus* 135.6 (2020): 1-19.
5. van Leeuwen, Caspar, et al. "Deep-learning enhancement of large scale numerical simulations." SURF Whitepaper, arXiv preprint arXiv:2004.03454 (2020).
6. Otten, Sydney, et al. "Event generation and statistical sampling for physics with deep generative models and a density information buffer." arXiv preprint arXiv:1901.00875 (2019).
7. Butter, Anja, Tilman Plehn, and Ramon Winterhalder. "How to GAN LHC events." *SciPost Phys.* 7, 075 (2019), <https://arxiv.org/abs/1907.03764>
8. Alanazi, Yasir, et al. "Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN)." arXiv preprint arXiv:2001.11103 (2020).
9. de Oliveira, Luke, Michela Paganini, and Benjamin Nachman. "Controlling physical attributes in GAN-accelerated simulation of electromagnetic calorimeters." *18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2017)* Seattle, WA, USA. 2017. <https://arxiv.org/pdf/1711.08813>
10. de Oliveira, Luke, Michela Paganini, and Benjamin Nachman. "Learning particle physics by example: location-aware generative adversarial networks for physics synthesis." *Computing and Software for Big Science* 1.1 (2017): 4. <https://arxiv.org/pdf/1701.05927>
11. Derkach, D., et al. "Cherenkov Detectors Fast Simulation Using Neural Networks", *Nuclear Instruments and Methods in*

Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 952 (2020): 161804; arXiv preprint arXiv:1903.11788.

12. Farrell, Steven, et al. "Next Generation Generative Neural Networks for HEP." EPJ Web of Conferences. Vol. 214. EDP Sciences, 2019.
13. Musella, Pasquale, and Francesco Pandolfi. "Fast and accurate simulation of particle detectors using generative adversarial networks." Computing and Software for Big Science 2.1 (2018): 8.
14. Di Sipio, Riccardo, et al. "DijetGAN: a Generative-Adversarial Network approach for the simulation of QCD dijet events at the LHC." Journal of High Energy Physics 2019.8 (2019): 110.
15. Hashemi, Bobak, et al. "LHC analysis-specific datasets with Generative Adversarial Networks." arXiv preprint arXiv:1901.05282 (2019).
16. Paganini, Michela, Luke de Oliveira, and Benjamin Nachman. "Accelerating science with generative adversarial networks: an application to 3D particle showers in multilayer calorimeters." Physical review letters 120.4 (2018): 042003.
17. Kai Zhou, Gergely Endrodi, Long-Gang Pang, and Horst Stöcker, "Regressive and generative neural networks for scalar field theory", Phys. Rev. D100, 011501(R) (2019)
18. Yang Zeng, Jin-Long Wu, Heng Xiao, "Enforcing Deterministic Constraints on Generative Adversarial Networks for Emulating", Physical Systems, arXiv 1911.06671
19. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.
20. Jin-Long Wu et al., "Enforcing Statistical Constraints in Generative Adversarial Networks for Modeling Chaotic Dynamical Systems", Journal of Computational Physics 406 (2020): 109209, arXiv 1905.06841
21. M. Mirza, S. Osindero, "Conditional generative adversarial nets", arXiv:1411.1784.
22. Eric Heim, "Constrained Generative Adversarial Networks for Interactive Image Generation", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, arXiv 1904.02526
23. Zhiting Hu et al., "Deep Generative Models with Learnable Knowledge Constraints", 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.
24. Giorgio Gnecco ET AL., "Learning with hard constraints as a limit case of learning with soft constraints", ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016
25. Pablo Márquez-Neila Mathieu Salzmann Pascal Fua, "Imposing Hard Constraints on Deep Networks: Promises and Limitations", CVPR Workshop on Negative Results in Computer Vision, Hawaii, HI, 2017, arXiv 1706.02025.

#### **4.6. Предлагаемые методы и подходы, общий план работы на весь срок выполнения проекта и ожидаемые результаты (объемом не менее 2 стр.; в том числе указываются ожидаемые конкретные результаты по годам; общий план дается с разбивкой по годам)**

Разработка эффективных с точки зрения временных и ресурсных затрат генераторов случайных выборок для моделирования физических явлений требует теоретического и экспериментального исследования широкого спектра различных типов генеративных сетей, функций ошибок и других гиперпараметров. Отдельное исследование будет направлено на поиск наиболее эффективного способа учета априорных знаний. Отправной точкой исследований в этом направлении будут генеративные сети, используемые в классических задачах по генерации изображений на основе conditional GAN. В частности на них будет отработана методика адаптация этих сетей к задачам гамма-астрономии.

Типы генеративных сетей, которые будут исследованы, включают обычные GAN, cGAN (conditional GAN), GAN с условиями связи (GAN with constraints), вариационные автоэнкодеры (VAE). С точки зрения архитектурных особенностей и выбора основной функции ошибок для нейросети будут рассмотрены сети типа Wasserstein GAN (WGAN), WGAN with Gradient Penalty (WGAN-GP), Least Squares GAN (LS-GAN), Maximum Mean Discrepancy GAN (MMDGAN) и ряд других.

Одним из подходов повышения эффективности генераторов, который будет исследован в рамках проекта, будет основан на использовании алгебраических ограничений, накладываемых на параметры генерируемых векторов. Для базового примера гамма-астрономии методика, основанная на таком подходе, может быть использована для наложения ограничений на параметры Хилласа [Hillas M. In NASA. Goddard Space Flight Center // 19th Intern. Cosmic Ray Conf. 1985. V. 3. P. 445-448] генерируемых изображений черенковского излучения в многоканальных камерах телескопов. Алгебраические уравнения и неравенства, отражающие априорные знания, могут быть интегрированы в алгоритмы обучения и генерации с помощью включения соответствующих членов в функцию потерь или, в более общей

постановке задачи, - посредством формулировки задачи обучения/генерации с ограничениями. Эти основанные на знаниях члены измеряют потенциальные несоответствия с априорной информацией, которая отражает законы физики или другой прикладной области, для исследования которой используется нейросеть. Такая расширенная функция обычно называется гибридной функцией потерь и обеспечивают глубокое обучение как на основе выборки данных, так и на основе предшествующих знаний. Помимо измерения несоответствий с точными формулами, посредством гибридных функций потерь могут быть учтены заранее известные приближенные соотношения или общие свойства аппроксимируемых функций (например, монотонность). При этом будут исследованы различные способы поиска минимума такой гибридной функции потерь, в частности:

- поиск минимума на поверхностях, определяемых условиями ограничений (априорными знаниями) — так называемый жесткий вариант учета ограничений (hard constraints);
- мягкий вариант (soft constraints) — путем добавления ограничений в функцию потерь генеративной сети.

Последний способ предполагается рассмотреть в двух вариантах:

- множитель перед членом с ограничениями в функции потерь рассматривается как один из гиперпараметров и подбирается экспериментально;
- использование формализма множителей Лагранжа, так что значение множителя перед членом определяется экстремумом (седловой точкой) такой гибридной функции потерь.

В качестве дополнительного подхода будет рассмотрен способ интеграции алгебраических уравнений в машинное обучение путем создания соответствующего вектора признаков, который непосредственно моделирует физические свойства и ограничения в духе работы: [Ladický, L'ubor, et al. "Data-driven fluid simulations using regression forests." ACM Transactions on Graphics (TOG) 34.6 (2015): 1-9], в которой разработан вектор признаков, непосредственно моделирующий отдельные силы и ограничения из уравнений Навье-Стокса. Еще один альтернативный подход — интеграция алгебраических уравнений, отражающих априорные ограничения, в архитектуру нейронных сетей аналогично работам: [R. Swischuk, L. Mainini, B. Peherstorfer, and K. Willcox, "Projection-based model reduction: Formulations for physics-based machine learning," Computers & Fluids, vol. 179, 2019; Y. Lu, M. Rajora, P. Zou, and S. Liang, "Physics-embedded machine learning: Case study with electrochemical micro-machining," Machines, vol. 5, no. 1, 2017; R. Ramamurthy, C. Bauckhage, R. Sifa, J. Schücker, and S. Wrobel, "Leveraging domain knowledge for reinforcement learning using mmc architectures," in Int. Conf. Artificial Neural Networks (ICANN). Springer, 2019; C. Bauckhage, C. Ojeda, J. Schücker, R. Sifa, and S. Wrobel, "Informed machine learning through functional composition" LWDA 2018]. Например, это могут быть дополнительные линейные связи между слоями сети или преобразования, отражающие физические ограничения на определенные входные и выходные переменные.

Очень важный метод улучшения работы генераторов событий, который будет исследован, - это учет статистических ограничений, а именно явный учет статистических характеристик обучающего набора на основе идей, предложенных в работе (для иной прикладной области) [Jin-Long Wu et al., "Enforcing Statistical Constraints in Generative Adversarial Networks for Modeling Chaotic Dynamical Systems", arXiv 1905.06841]. Такой учет достигается ведением в функцию потерь метрики, измеряющей различие между ковариационной структурой обучающего и генерируемого наборов событий. В приложении к нашему базовому примеру этот метод будет играть решающую роль для правильного воспроизведения в процессе генерации спектра частиц космических лучей.

Общий план работы на весь срок выполнения проекта и ожидаемые результаты.

В 2021 году на первом этапе планируется выполнить следующие работы:

1. Выполнить аналитический обзор современной научно-технической литературы по теме проекта.
2. Выработать список критериев эффективного функционирования генераторов событий на примере выборок событий регистрации ШАЛ атмосферными черенковскими телескопами.
3. Провести теоретическое исследование и предварительный отбор основных типов генеративных сетей для использования в качестве генераторов событий на основе сформулированных критериев эффективного функционирования и анализа существующих Монте-Карло генераторов.
4. Адаптировать существующие и разработать новые методы и алгоритмы учета априорных ограничений при генерации событий в области астрофизики частиц, в том числе:
  - методы и алгоритмы учета детерминистских ограничений;
  - методы и алгоритмы учета статистических ограничений.
5. Подготовка 2 докладов на международной конференции и 2 публикации статей.
6. Подготовка промежуточного отчета.

Результаты работы на первом этапе:

1. Аналитический обзор анализа современной научно-технической литературы.
2. Список критериев эффективного функционирования генераторов событий на примере физики высоких энергий.
3. Обоснованный выбор основных типов генеративных сетей для использования в качестве генераторов событий.
4. Адаптированные существующие и новые методы и алгоритмы учета априорных ограничений, включая.
5. Публикация 2 статей, отражающих промежуточные результаты работы по проекту;
6. Промежуточный отчет по проекту.

В 2022 году на втором (заключительном) этапе будет осуществлена программная реализация разработанных подходов, методов и алгоритмов. Будут выполнены следующие работы:

1. Разработка и реализация программ на основе разработанных на первом этапе методов и алгоритмов для целей учета априорных знаний при генерации событий нейросетями, включая:
  - методы и алгоритмы учета детерминистских ограничений;
  - методы и алгоритмы учета статистических ограничений;
2. Проведение вычислительных экспериментов с целью исследования разработанных генераторов событий для гамма-астрономии. В том числе:
  - проверка соответствия критериям их эффективного функционирования, выработанным на первом этапе;
  - сравнительный анализ МС генераторов и генераторов основанных на машинном обучении;
3. Оптимизация программной реализации генераторов, основанных на машинном обучении по результатам экспериментальных исследований;
4. Анализ, полученных результатов, в том числе:
  - а) обобщение результатов исследований;
  - б) сопоставление анализа научно-информационных источников и результатов теоретических и экспериментальных исследований;
  - в) оценка эффективности полученных результатов в сравнении с современным научно-техническим уровнем;
5. Подготовка докладов на международных и всероссийских конференциях и публикация 3 статей.
6. Подготовка подготовка итогового отчета.

Результаты работы на втором этапе:

1. Программная реализация методов и алгоритмов генерации событий нейросетями с учетом априорных детерминистских и статистических ограничений.
2. Аналитический отчет о проведенных экспериментальных исследованиях генераторов, основанных на методах машинного обучения:
  - соответствие критериям их эффективного функционирования, выработанным на первом этапе;
  - сравнительный анализ Монте-Карло генераторов и генераторов основанных на машинном обучении.
3. Оптимизированная программная реализация генераторов, основанных на машинном обучении по результатам экспериментальных исследований.
4. Выводы о качестве и эффективности работы генераторов событий.
5. Публикации 3 статей, отражающих результаты работы по проекту
6. Итоговый отчет по проекту.

В ходе выполнения проекта члены коллектива примут участие в российских и международных конференциях и рабочих совещаниях с целью обсуждения полученных результатов.

#### **4.7. Имеющийся у научного коллектива научный задел по проекту, наличие опыта совместной реализации проектов** *(указываются полученные ранее результаты, разработанные программы и методы)*

Предпосылкой успешного выполнения предлагаемых работ является существующий высокий уровень владения участниками проекта методами глубокого обучения и их приложения в области физики элементарных частиц и астрофизики. Удачное сочетание высококвалифицированных ученых и молодых исполнителей.

Основной задел непосредственно по предлагаемой тематике членами коллектива был получен в рамках международного проекта РФ-Гельмгольц №18-41-06003 по теме "RSF-Helmholtz: Карлсруэ-Российская инициатива по работе с астрофизическими данными на протяжении их жизненного цикла." Основные задачи, которые решались в рамках проекта методами машинного обучения были вопросы классификации (распознавания типа первичных частиц

космических лучей) по изображениям, полученных с черенковских телескопов, задача регрессии (определение физических параметров первичных частиц таких как энергия, направление оси широкого атмосферного ливня и другие), а также задача использования искусственных нейронных сетей на основе автоэнкодеров для улучшения отношения сигнал/шум реальных изображений. В рамках проекта разработаны новые подходы к распознаванию типа первичных частиц космических лучей и их характеристик на основе сверточных нейронных сетей. Было показано, что использование методов машинного обучения позволяет увеличить качество предсказания типа частиц и их характеристик на 20-30% по сравнению со стандартной методикой обработки данных, применяемой физиками. Отдельно отметим, что данный метод впервые был применен для стерео режима регистрации частиц (одновременная регистрация одних и тех же широких атмосферных ливней двумя и более телескопами), что позволило улучшить качество классификации первичных частиц почти вдвое. Это открывает широкие перспективы использования современных методов машинного обучения в астрофизике частиц. Работы в этом направлении, а также большой опыт членов коллектива в области физики высоких энергий (заметим, что руководитель коллектива А.П.Крюков входил в состав авторов открытия бозона Хиггса в экспериментах коллаборации CMS в 2012 году) позволил сформулировать одну из ключевых задач моделирования событий в физике частиц методом машинного обучения, а именно проблему эффективной генерации выборок событий взамен используемого ресурсоемкого процесса Монте-Карло генерации. В дополнение к научному заданию отметим, что члены коллектива участвовали в создании образовательного веб-портала по астрофизике космических лучей (<http://astroparticle.online>), где был впервые имплементирован модуль классификации первичных частиц на основе сверточных сетей. Это позволило использовать данный портал не только как информационный, но и для организации практикума для студентов специализирующихся в этой области. Работы в данном направлении продолжаются силами молодых участников проекта (В.А.Власкина и Е.О.Гресь) под руководством А.Крюкова.

Членами коллектива был разработан ряд программ по машинному обучению, часть из которых успешно используются для обработки данных в области гамма астрономии для эксперимента TAIGA.

Результаты этих работ неоднократно докладывались на международных конференциях, в частности, International Conference on Computer Simulation in Physics and beyond (CSP 2020), Москва, Россия, 12-15 октября 2020, IV International Workshop "Data life cycle in physics", DLC-2020, Москва, Россия, 8-10 июня 2020, The 27th International Symposium Nuclear Electronics and Computing (NEC'2019), Черногория, 30 сентября - 4 октября 2019 и ряде других.

Коллектив имеет очень большой опыт совместной реализации проектов. В частности, коллектив исполнителей участвовал в выполнении большого числа работ по развитию грид-технологии. Они являлись участниками крупнейших международных проектов в области численной обработки экспериментальных данных: EU DataGRID (<http://marianne.in2p3.fr/datagrid>), "Enabling Grids for E-science" (EGEE, <http://eu-egee.org>) и European Grid Initiative (EGI, <http://www.egi.eu>), в проекте Европейского центра ядерных исследований (ЦЕРН, Женева, Швейцария) "The Worldwide LHC Computing Grid Project" (WLCG, <http://www.cern.ch/WLCG>). Руководитель проекта - член коллаборации CMS (<https://cms.cern>) Большого адронного коллайдера, член международной коллаборации TAIGA (<https://taiga-experiment.info/>) и нейтринного эксперимента Hyper-Kamiokanda (<https://hyper-k.org>). Все указанные проекты направлены на сбор, обработку и анализ данных в области физики элементарных частиц и космических лучей, являются пионерской областью науки, в которой моделирование событий является ключевым моментом любого исследования.

Члены коллектива за последние 5 лет участвовали в исследованиях в рамках следующих проектов:

- "RSF-Helmholtz: Карлсруэ-Российская инициатива по работе с астрофизическими данными на протяжении их жизненного цикла", грант РФ-Гельмгольц №18-41-06003, руководитель к.ф.-м.н. А.П.Крюков.

- "Разработка принципов и алгоритмов управления метаданными провенанса больших научных данных с использованием блокчейн-технологии", грант РФ: - проект 18-11-00075.

- «Разработка и исследование методов повышения производительности суперкомпьютеров на основе миграции заданий с использованием контейнерной виртуализации», грант РФФИ 18-37-00502, руководитель к.ф.-м.н. С.П.Поляков.

- «Разработка новой технологии и создание и на ее основе прототипа открытой веб-платформы для организации предоставления информационно-вычислительных веб-услуг в рамках метафоры открытого рынка (marketplace) на примере услуг по конструированию и расчету нелинейно-оптических устройств для преобразования лазерного излучения», грант РФФИ № 15-07-09309, руководитель к.ф.-м.н. А.П.Демичев.

- «Разработка новых принципов построения инфраструктуры безопасности распределенных информационно-

вычислительных систем на основе открытых протоколов», ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014 – 2020 годы»; Соглашение о предоставлении субсидии No 14.604.21.0146, 24.11.2014 г.), руководитель к.ф.-м.н. А.П.Крюков.

Все это говорит о высоком уровне коллектива и его безусловной возможности получить заявленные результаты.

Основные публикации за последние 5 лет:

1. Postnikov E., Kryukov A., Polyakov S., Zhurov D. "Deep learning for energy estimation and particle identification in gamma-ray astronomy" // CEUR Workshop Proceedings, том 2406, с. 90-99, 2019.
2. Postnikov E.B., Kryukov A.P., Polyakov S.P., Shipilov D.A., Zhurov D.P., "Gamma/Hadron Separation in Imaging Air Cherenkov Telescopes Using Deep Learning Libraries TensorFlow and PyTorch" // Journal of Physics: Conference Series, издательство IOP Publishing ([Bristol, UK], England), том 1181, с. 012048, 2019
3. Postnikov Evgeny B., Bychkov Igor V., Dubenskaya Julia Y., Fedorov Oleg L., Kazarina Yulia A., Korosteleva Elena E., Kryukov Alexander P., Mikhailov Andrey A., Minh-Duc Nguyen // Polyakov Stanislav P., Shigarov Alexey O., Shipilov Dmitry A., Zhurov Dmitry P., "Particle identification in ground-based gamma-ray astronomy using convolutional neural networks", CEUR Workshop Proceedings, том 2267, с. 431-435, 2018
4. Igor Bychkov, Andrey Demichev, Julia Dubenskaya, Oleg Fedorov, Andreas Haungs, Andreas Heiss, Donghwa Kang, Yulia Kazarina, Elena Korosteleva, Dmitriy Kostunin, Alexander Kryukov, Andrey Mikhailov, Minh Duc Nguyen, Stanislav Polyakov, Evgeny Postnikov, Alexey Shigarov, Dmitry Shipilov, Achim Streit, Victoria Tokareva, Doris Wochele, Jürgen Wochele, Dmitry Zhurov, "Russian–German Astroparticle Data Life Cycle Initiative" // Data, том 3, No 4, с. 56, 2018
5. A. P. Kryukov and A. P. Demichev, «Architecture of Distributed Data Storage for Astroparticle Physics» // Lobachevskii Journal of Mathematics, 2018, Vol. 39, No. 9, pp. 1199–1206. (SCOPUS, SJR – 2016: 0.368)
6. A. P. Kryukov and A. P. Demichev, «Decentralized Data Storages: Technologies of Construction» // Programming and Computer Software, 2018, Vol. 44, No. 5, pp. 303–315 (<https://link.springer.com/article/10.1134/S0361768818050067>, индексируется Web of Science, Impact Factor – 2016: 0.230)
7. Andrey Demichev, Julia Dubenskaya, Alexander Kryukov, Stanislav Polyakov, Nikolai Prikhod'ko "A new approach to the development of provenance metadata management systems for large scientific experiments", 2018 // The 8th International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018), CEUR Workshop Proceedings (CEUR-WS.org); <http://ceur-ws.org/Vol-2267/323-327-paper-61.pdf>.
8. Andrey Demichev, Alexander Kryukov, Nikolai Prikhod'ko, "The Approach to Managing Provenance Metadata and Data Access Rights in Distributed Storage Using the Hyperledger Blockchain Platform" // IEEE Xplore Digital Library, издательство IEEE Computer Society (United States), 2019, с. 131-136, DOI 10.1109/ISPRAS.2018.00028; <https://ieeexplore.ieee.org/document/8675153>.
9. Andrey Demichev, Alexander Kryukov, Nikolai Prikhod'ko "Blockchain-Based Delegation of Rights in Distributed Computing Environment" // Lecture Notes in Computer Science, v. 11657, pp. 408–418, 2019; [https://doi.org/10.1007/978-3-030-25636-4\\_32](https://doi.org/10.1007/978-3-030-25636-4_32).
11. Andrey Demichev, Alexander Kryukov, Nikolai Prikhod'ko, "Metadata driven data management in distributed computing environments with partial or complete lack of trust between user groups" // IEEE Xplore Digital Library, издательство IEEE Computer Society (United States), 2020
12. Kryukov A.P., Demichev A.P., Polyakov S.P., Web Platforms for Scientific Research // Programming and Computer Software, том 42, No 3, (2016) с. 129-141 (<http://link.springer.com/article/10.1134%2FS036176881603004X>)
13. Prikhodko N. V., Abramovsky V. A., Abramovskaya N. V., Demichev A. P., Kryukov A. P., Polyakov S. P., A Web Tool for Research in Nonlinear Optics // Journal of Physics: Conference Series, том 681, (2016) с. 012052-1-012052-5 (<http://iopscience.iop.org/article/10.1088/1742-6596/681/1/012052/meta>)

Свидетельства о государственной регистрации программ.

1. Журов Д., Крюков А., Постников Е., Сидоров Д. «Программа идентификации первичных частиц космических лучей по изображениям с атмосферных черенковских телескопов методом машинного обучения» (No. 2019666634 от 12.12.2019);
2. Поляков С.П., Крюков А.П., Постников Е.Б. «Программа определения типа и параметров первичных частиц на основании данных атмосферного черенковского телескопа методом машинного обучения» (No. 2020618844 от 05.08.2020).
3. Крюков А.П., Демичев А.П., Поляков С.П. «Сервис управления Docker-контейнерами пользователей», (No 2016662953, 25.11.2016)
4. Дубенская Ю.Ю., Демичев А.П., Поляков С.П. «Сервис аутентификации и авторизации инфраструктуры безопасности

распределенных информационно-вычислительных систем на основе открытых протоколов» (№ 2016616673, 16.06.2016)

5. Крюков А.П., Демичев А.П., Приходько Н.В. «Пользовательский интерфейс для удаленного взаимодействия с системой управления распределенными хранилищами» (свидетельство о государственной регистрации программы для ЭВМ № 2019616518 от 23.05.2019);

6. Крюков А.П., Демичев А.П., Приходько Н.В. «Сервис управления данными и метаданными провенанса в распределенных хранилищах» (свидетельство о государственной регистрации программы для ЭВМ № 2019616702 от 29.05.2019);

7. Крюков А.П., Демичев А.П., Приходько Н.В. «Бизнес-логика управления распределенными хранилищами на основе блокчейн-технологии» (свидетельство о государственной регистрации программы для ЭВМ № 2019616519 от 23.05.2019).

#### **4.8. Перечень оборудования, материалов, информационных и других ресурсов, имеющихся у научного коллектива для выполнения проекта (в том числе – описывается необходимость их использования для реализации проекта)**

Члены коллектива через сервисы, предоставляемые сотрудникам МГУ имеет доступ к архивам научных журналов, включая иностранные, а также к наукометрическим системам Web of Science и Scopus. Таким образом члены коллектива имеют доступ к публикациям современных исследований в области машинного обучения и их приложениям в физике.

Коллектив имеет доступ к сетям Интернет с высокоскоростным каналом до 4Гб/с. Также члены коллектива обеспечены персональными рабочими местами, включенными в 1 Гб/с локальную сеть, оргтехникой (цветной принтер, средства проведения видео и аудио конференций), а также необходимыми расходными материалами, что обеспечит должное выполнение плана работ, предусмотренных программой исследований.

В настоящее время коллектив имеет в своем распоряжении вычислительный сервер с двумя профессиональными ГПУ NVIDIA P100, что позволяет проводить исследования моделей на основе машинного обучения для целей подтверждения правильности выбранных подходов.

Участие научного руководителя в коллаборации TAIGA обеспечивает доступ к экспериментальным данным коллаборации, на которых будут проводиться апробация разрабатываемых методов и программ на их основе.

Таким образом, члены коллектива полностью обеспечены необходимым для успешного выполнения проекта оборудованием, материалами, информационными ресурсами. Приобретение дополнительного оборудования не предусмотрено.

#### **4.9. План работы на первый год выполнения проекта (в том числе указываются запланированные командировки (экспедиции) по проекту)**

Общий план работы на весь срок выполнения проекта и ожидаемые результаты.

В 2021 году на первом этапе планируется выполнить следующие работы:

1. Выполнить аналитический обзор современной научно-технической литературы по теме проекта.
2. Выработать список критериев эффективного функционирования генераторов событий на примере выборок событий регистрации ШАЛ атмосферными черенковскими телескопами.
3. Провести теоретическое исследование и предварительный отбор основных типов генеративных сетей для использования в качестве генераторов событий на основе сформулированных критериев эффективного функционирования и анализа существующих Монте-Карло генераторов.
4. Адаптировать существующие и разработать новые методы и алгоритмы учета априорных ограничений при генерации событий в области астрофизики частиц, в том числе:
  - методы и алгоритмы учета детерминистских ограничений;
  - методы и алгоритмы учета статистических ограничений.
5. Подготовка 2 докладов на международной конференции и 2 публикация статей.
6. Подготовка промежуточного отчета.

Результаты работы на первом этапе:

1. Аналитический обзор анализа современной научно-технической литературы.

2. Список критериев эффективного функционирования генераторов событий на примере физики высоких энергий.
3. Обоснованный выбор основных типов генеративных сетей для использования в качестве генераторов событий.
4. Адаптированные существующие и новые методы и алгоритмы учета априорных ограничений, включая.
5. Публикация 2 статей, отражающих промежуточные результаты работы по проекту;
6. Промежуточный отчет по проекту.

В ходе первого этапа работ будут подготовлены 2 доклада на международных конференциях CSP и DLCP. Для участия в работе конференций планируются 2 командировки. Будут также подготовлены и направлены в печать 2 статьи в издания, индексируемые системами WoS и Scopus. Окончательный список конференций и изданий будет зафиксирован в ходе выполнения проекта.

#### **4.10. Планируемое на первый год содержание работы каждого исполнителя проекта (включая руководителя проекта)**

А.П.Крюков, руководитель проекта:

1. Общее руководство проектом.
2. Выполнить аналитический обзор современной научно-технической литературы по теме проекта.
3. Выработать список критериев эффективного функционирования генераторов событий на примере выборок событий регистрации ШАЛ атмосферными черенковскими телескопами.
4. Подготовка 2 докладов на международные конференции и 2 статей в журналы.
5. Подготовка промежуточного отчета.

Ю.Ю.Дубенская, Е.О.Грась,

1. Провести теоретическое исследование и предварительный отбор основных типов генеративных сетей для использования в качестве генераторов событий на основе сформулированных критериев эффективного функционирования и анализа существующих Монте-Карло генераторов.
2. Подготовка материалов 2 докладов на международные конференции.
3. Подготовка материалов промежуточного отчета.

Ю.Ю.Дубенская, А.А.Власкина

1. Адаптировать существующие и разработать новые методы и алгоритмы учета априорных ограничений при генерации событий в области астрофизики частиц, в том числе:
  - методы и алгоритмы учета детерминистских ограничений;
  - методы и алгоритмы учета статистических ограничений.
2. Подготовка материалов 2 докладов на международные конференции.
3. Подготовка материалов промежуточного отчета.

В ходе первого этапа работ планируется подготовить 2 доклада на международные конференции CSP и DLCP. Для участия в работе конференций планируются 2 командировки. Будут также подготовлены и направлены в печать 2 статьи в журналы. Окончательный список конференций и изданий будет зафиксирован в ходе выполнения проекта.

#### **4.11. Ожидаемые в конце первого года конкретные научные результаты (форма изложения должна дать возможность провести экспертизу результатов и оценить степень выполнения заявленного в проекте плана работы)**

Результаты работы на первом этапе:

1. Аналитический обзор современной научно-технической литературы по тематике проекта.
2. Список критериев эффективности генераторов событий на примере гамма-астрономии.
3. Обоснованный выбор основных типов генеративных сетей для использования в качестве генераторов событий.
4. Адаптация существующих и разработка новых методов и алгоритмов учета априорных ограничений при генерации событий в области гамма-астрономии, в том числе для учета статистических характеристик получающихся семплов событий (изображений).
5. Прототип генеративной нейронной сети для генерации семплов с учетом априорных знаний, включая учет статистических ограничений. Осуществить пробную генерацию событий (изображений) ШАЛ, регистрируемых атмосферными черенковскими телескопами.
6. Промежуточный отчет по проекту.
7. Два доклада на международных конференциях по результатам исследований.
8. Две статьи, отражающих промежуточные результаты работы по проекту в изданиях, индексируемых системами WoS и Scopus.

**4.12. Перечень планируемых к приобретению за счет гранта оборудования, материалов, информационных и других ресурсов для выполнения проекта** *(в том числе – описывается необходимость их использования для реализации проекта)*

Коллектив полностью обеспечен необходимыми для выполнения работы ресурсами. Приобретение за счет гранта оборудования, материалов, информационных и других ресурсов не предусмотрено.

**4.13. Файл с дополнительной информацией 1**

С графиками, фотографиями, рисунками и иной информацией о содержании проекта. Один файл в формате pdf, до 3 Мб.

Текст в файлах с дополнительной информацией должен приводиться на русском языке. Перевод на английский язык требуется в том случае, если руководитель проекта оценивает данную информацию существенной для эксперта.

---

**4.14. Файл с дополнительной информацией 2** *(если информации, приведенной в файле 1 окажется недостаточно)*

С графиками, фотографиями, рисунками и иной информацией о содержании проекта. Один файл в формате pdf, до 3 Мб.

---

Подпись руководителя проекта \_\_\_\_\_/А.П. Крюков/

## Форма 5. Запрашиваемое финансирование на 2022 год

### 5.1. Планируемые расходы по проекту

№ п.п.	Направления расходования средств гранта	Сумма расходов (тыс.руб.)
	<b>ВСЕГО</b>	1500
	Вознаграждение членов научного коллектива (с учетом страховых взносов во внебюджетные фонды, без лиц категории «вспомогательный персонал»)	1270
	Вознаграждение лиц категории «вспомогательный персонал» (с учетом страховых взносов во внебюджетные фонды)	0
1	Итого вознаграждение (с учетом страховых взносов во внебюджетные фонды)	1270
2	Оплата научно-исследовательских работ сторонних организаций, направленных на выполнение научного проекта (не более 15 процентов от суммы гранта)	0
3	Расходы на приобретение оборудования и иного имущества, необходимых для проведения научного исследования (включая обучение работников, монтажные, пуско-наладочные и ремонтные***** работы) ***** Не связанные с осуществлением текущей деятельности организации.	0
4	Расходы на приобретение материалов и комплектующих для проведения научного исследования	0
5	Иные расходы для целей выполнения проекта	80
6	Накладные расходы организации (не более 10 процентов от суммы гранта)	150

### 5.2. Расшифровка планируемых расходов

№ п.п.	Направления расходования средств гранта, расшифровка
1	<p>Итого вознаграждение (с учетом страховых взносов во внебюджетные фонды) (указывается сумма вознаграждения (включая руководителя и иных исполнителей, привлекаемых к выполнению работ по проекту), включая установленные законодательством Российской Федерации гарантии, отчисления по страховым взносам на обязательное пенсионное страхование, на обязательное медицинское страхование, на обязательное социальное страхование на случай временной нетрудоспособности и в связи с материнством, на обязательное социальное страхование от несчастных случаев на производстве и профессиональных заболеваний)</p> <p>Итого вознаграждение - 1270 тыс.руб. В том числе:                      Руководитель гранта А.Крюков – 520 тыс руб.                      Основной исполнитель Ю.Дубенская – 300 тыс. руб.                      Исполнитель Е.О.Гресь – 250 тыс руб.                      Исполнитель А.А.Власкина – 200 тыс.руб.</p>
2	<p>Оплата научно-исследовательских работ сторонних организаций, направленных на выполнение научного проекта (приводится перечень планируемых договоров (счетов) со сторонними организациями с указанием предмета и суммы каждого договора)</p> <p>не предусмотрено</p>
3	<p>Расходы на приобретение оборудования и иного имущества, необходимых для проведения научного исследования (представляется перечень планируемых к закупке оборудования и иного имущества, необходимых для проведения научного исследования (в соответствии с п. 4.12 формы 4))</p> <p>не предусмотрено</p>
4	<p>Расходы на приобретение материалов и комплектующих для проведения научного исследования (представляется расшифровка запланированных материалов и комплектующих (в соответствии с п. 4.12 формы 4))</p> <p>не предусмотрено</p>

5 **Иные расходы для целей выполнения проекта**

(приводятся иные затраты на цели выполнения проекта, в том числе на командировки, оплату услуг связи, транспортных услуг, расходы не расшифровываются)

Иные расходы - 80 тыс.руб. Расходы на командировки для представления результатов работы на всероссийских и международных конференциях.

**Подпись руководителя проекта** \_\_\_\_\_ /А.П. Крюков/

**Подпись руководителя организации** (уполномоченного представителя, действующего на основании доверенности или распорядительного документа), **печать** (при ее наличии) **организации.**

В случае подписания формы уполномоченным представителем организации (в т.ч. – руководителем филиала) к печатному экземпляру заявки прилагается копия распорядительного документа или доверенности, заверенная печатью организации.

\_\_\_\_\_/\_\_\_\_\_/

М.П.