

Специфика обучения и использования суррогатных генеративных моделей в задачах быстрой симуляции для физических экспериментов



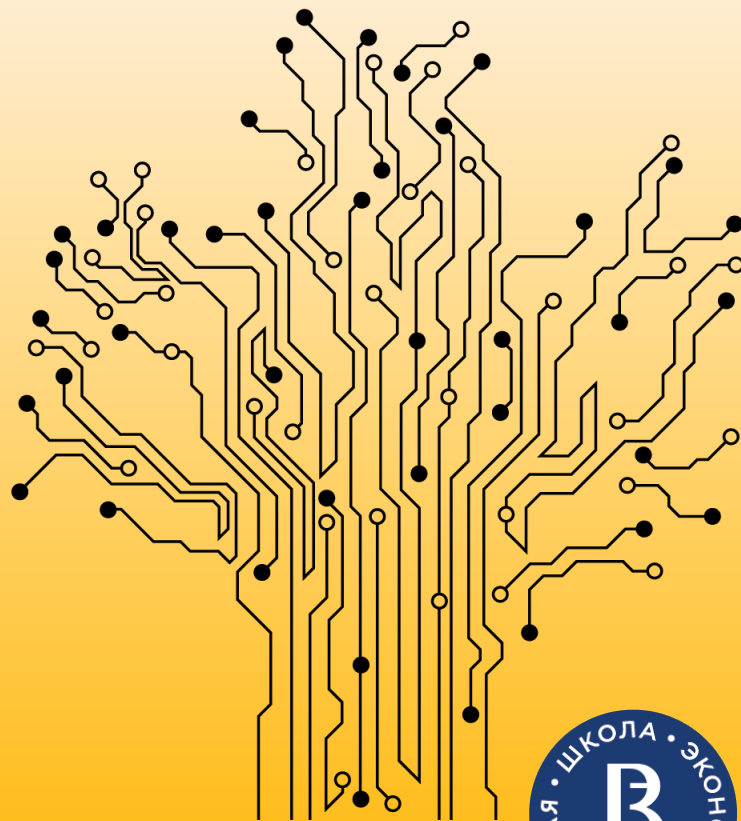
LAMBDA • HSE

Фёдор Ратников

Научно-учебная лаборатория методов анализа больших данных

Институт искусственного интеллекта и цифровых наук

Факультет компьютерных наук НИУ “Высшая школа экономики”



Содержание



Зачем физике симуляции (генеративные модели)

Суррогатные модели для физических симуляций

Особенности практического применения

Заключение

Теория

Концепция изучения сути вещей ...



Предметы состоят из меньших частей,

→ которые состоят из меньших частей,

→ которые состоят из меньших частей,

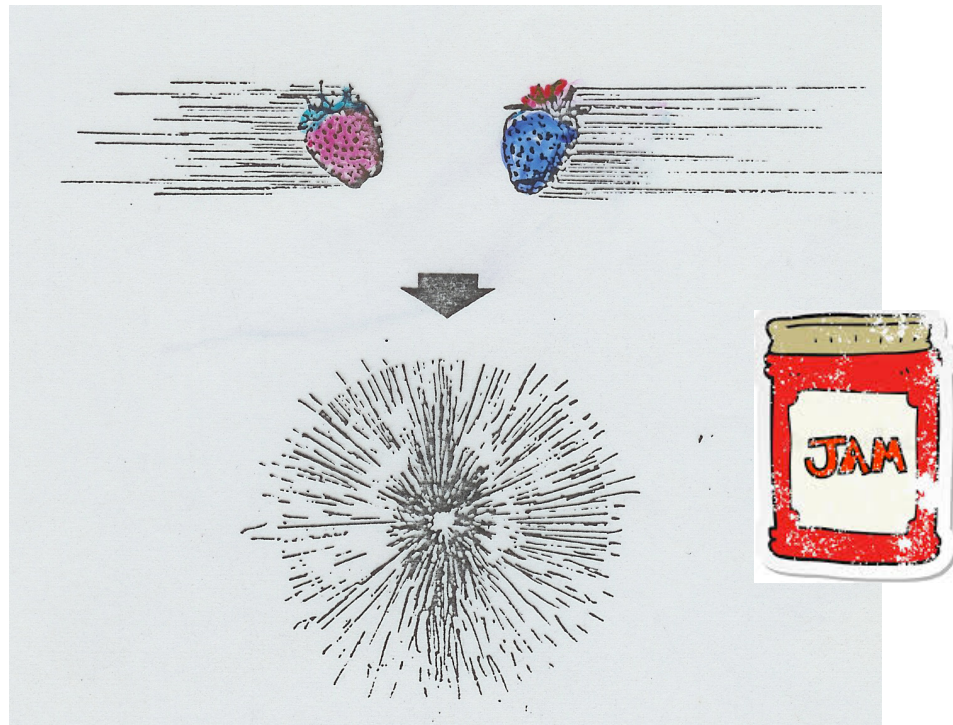
→ ...

Как можно изучать составляющие части?

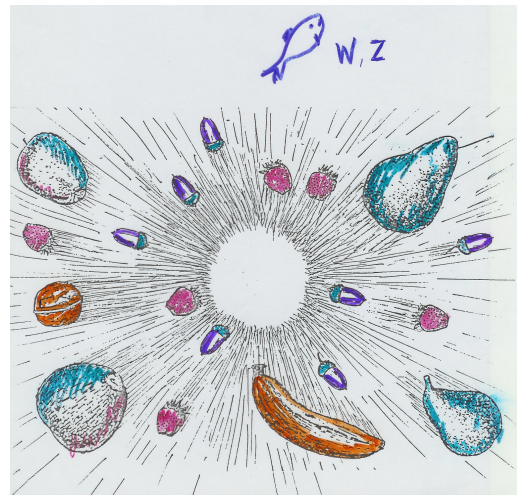
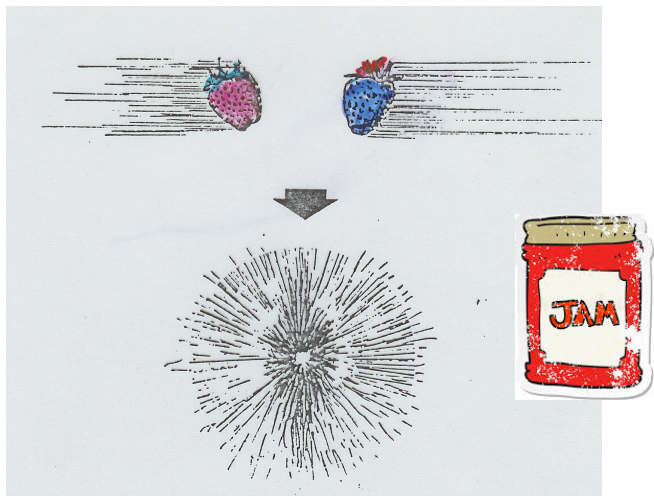
Даже после того, как мы разобрали всё на части, у нас осталась ещё возможность

→ разбить!

Концепция изучения сути вещей ...



Концепция исследований ФВЭ



В отличие от фруктов и ягод, при столкновении частиц образуются новые частицы, причем весьма отличные от тех, которые столкнулись

▶ в частности, те частицы, которые отсутствуют в современной Вселенной, но присутствовали в начале её существования

Новые частицы

Буквально “выколачиваем” их из ничего

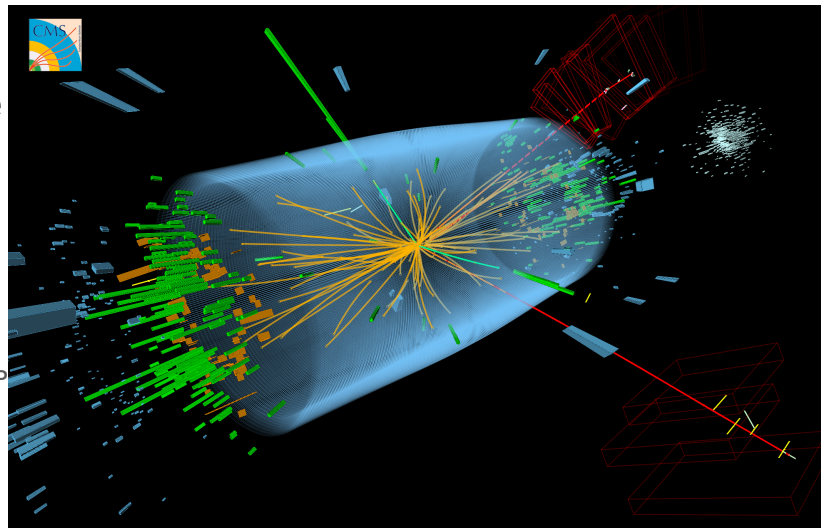
создаем огромную плотность энергии в микроскопическом объеме: $m = E/c^2$

После рождения частицы мгновенно распадаются на обычные частицы

Например, мы не можем наблюдать бозон Хиггса, но можем наблюдать продукты его распада

Для успешного исследования нам требуются:

- ускоритель частиц, чтобы сконцентрировать энергию и образовать новые частицы
- детектор частиц, чтобы зарегистрировать продукты распада этих частиц
- инфраструктура для обработки огромного количества событий и извлечения из них информации об обнаруженных частицах
- средства анализа данных, извлечения физических величин и представления результатов



Более конкретно...

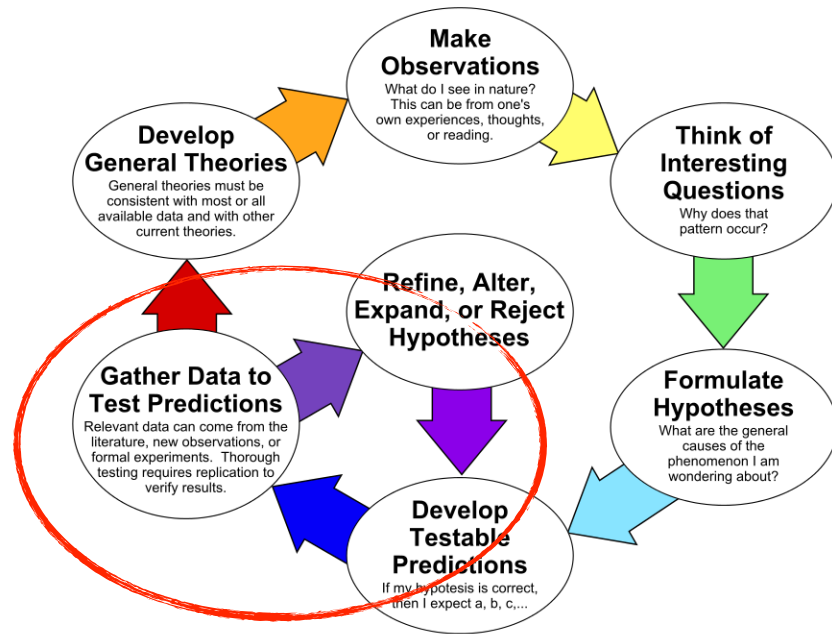
Мы хотим исследовать некоторый закон
Вселенной

▶ например, Стандартную модель элементарных
частиц или новую физику, выходящую за рамки
Стандартной модели

Мы предсказываем определенный эффект

Мы наблюдаем некоторый эффект

Как соотносить наблюдение с предсказанием?



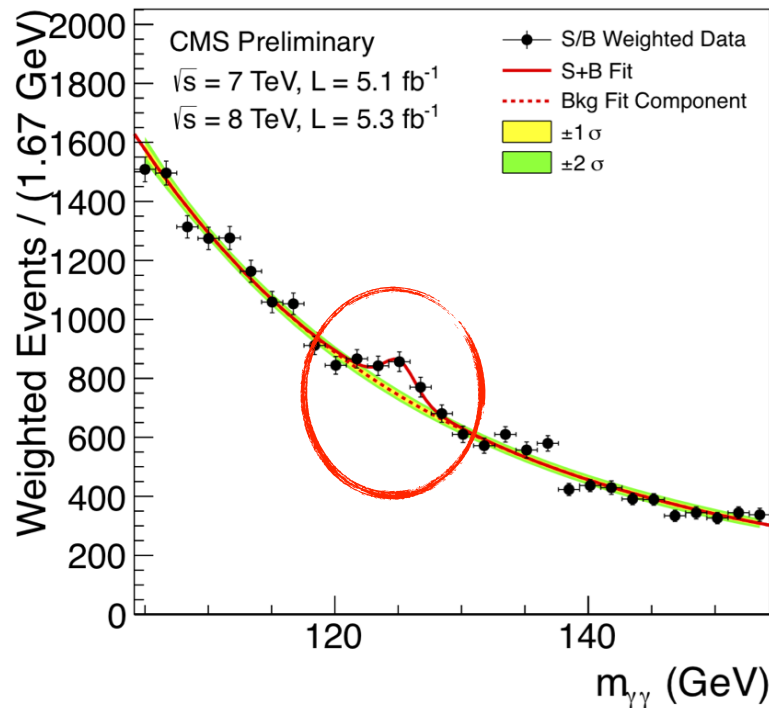
Как соотнести наблюдение с предсказанием?

Мы хотим исследовать некоторый закон Вселенной

▶ например, Стандартную модель элементарных частиц или новую физику, выходящую за рамки Стандартной модели

Мы предсказываем определенный эффект

Мы наблюдаем некоторый эффект



Максимальное правдоподобие

Мы хотим исследовать некоторый закон Вселенной

▶ например, Стандартную модель элементарных частиц или новую физику, выходящую за рамки Стандартной модели

Мы предсказываем определенный эффект

Мы наблюдаем некоторый эффект

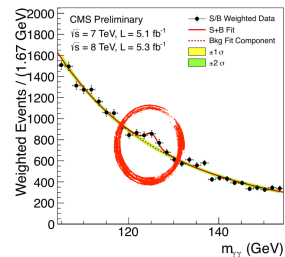
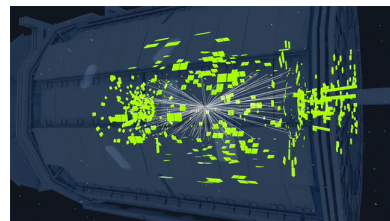
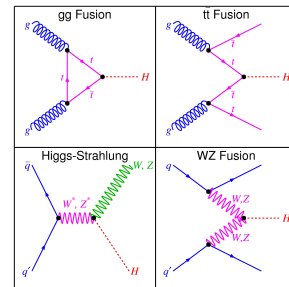
Как соотнести наблюдение с предсказанием?

Требование максимального правдоподобия:

- ▶ предпочтительным (наиболее правильным) является тот закон природы, чье предсказание наиболее похоже на наблюдаемый в эксперименте результат
- ▶ требуется найти значения соответствующих параметров физических законов

От законов природы к измеряемым величинам

1. Взаимодействие сгустков протонов в ускорителе
2. Взаимодействие протонов, летящих навстречу друг другу
3. Жесткое взаимодействие, описываемое моделью
4. Образование партонов
5. Адронизация партонов в долгоживущие частицы
6. Взаимодействие частиц с веществом экспериментальной установки
7. Взаимодействие с детекторами частиц
8. Считывание данных с детекторов
9. Усиление и передача сигналов
10. Оцифровка сигналов
11. Объединение данных разных детекторов в единое событие
12. Триггерные отборы
13. Физический анализ
14. Измеренные наблюдаемые величины



Прямая и обратная задачи

Прямая задача (Симуляция)

- ▶ предполагаем некоторую, например Стандартную, модель элементарных ча
- ▶ рассматриваем конкретный канал рождения частиц
- ▶ симулируем стохастические процессы и преобразования шагов 1..13
- ▶ оцениваем ожидаемый наблюдаемый эффект (14)
- ▶ сравниваем ожидаемый эффект с реально наблюдаемым в эксперименте

Обратная задача (Inference)

- ▶ рассматриваем конкретную параметрическую модель
 - например, образование частиц темной материи некоторой массы
- ▶ делаем экспериментальное наблюдение
- ▶ оцениваем функцию правдоподобия через шаги 13..1
- ▶ оцениваем значения параметров модели как максимизирующие правдоподобие

1. Взаимодействие сгустков протонов в ускорителе
2. Взаимодействие протонов, летящих навстречу друг другу
3. Жесткое взаимодействие, описываемое моделью
4. Образование партонов
5. Адронизация партонов в долгоживущие частицы
6. Взаимодействие частиц с веществом экспериментальной установки
7. Взаимодействие с детекторами частиц
8. Считывание данных с детекторов
9. Усиление и передача сигналов
10. Оцифровка сигналов
11. Объединение данных разных детекторов в единое событие
12. Триггерные отборы
13. Физический анализ
14. Измеренные наблюдаемые величины

В условиях физики частиц обратная задача слишком сложна, чтобы ее решать прямыми методами

- ▶ однако может быть решена итеративно через использование решения прямой задачи

Основные компьютерные ресурсы экспериментов по физике частиц тратятся на прецизионную симуляцию стохастических процессов на шаге (6)

Geant4 - универсальный пакет, реализующий шаги (6..8)

➡ <https://geant4.web.cern.ch/>

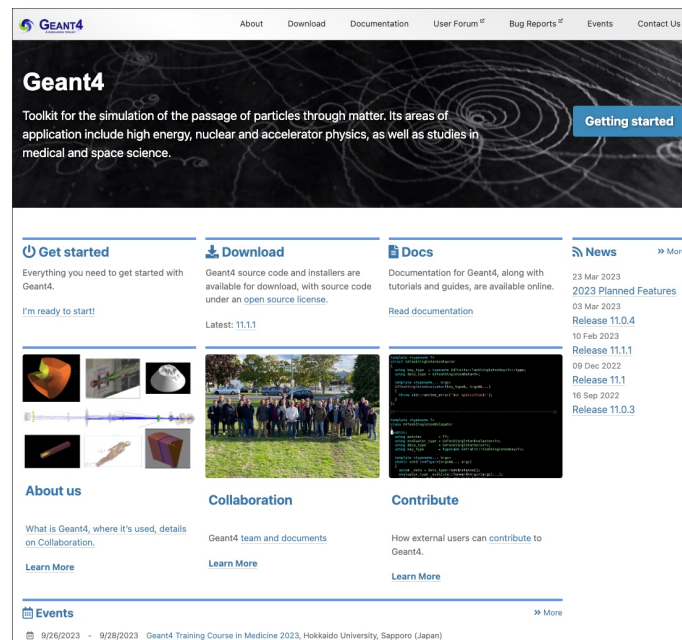
➡ несколько сотен человеко-лет разработки

➡ более 1.5M линий кода

Большая часть из данных физики частиц - данные, сгенерирована с помощью Geant4

➡ (~ 5 Exabytes = 5M Tb)

6. Взаимодействие частиц с веществом экспериментальной установки
7. Взаимодействие с детекторами частиц
8. Считывание данных с детекторов



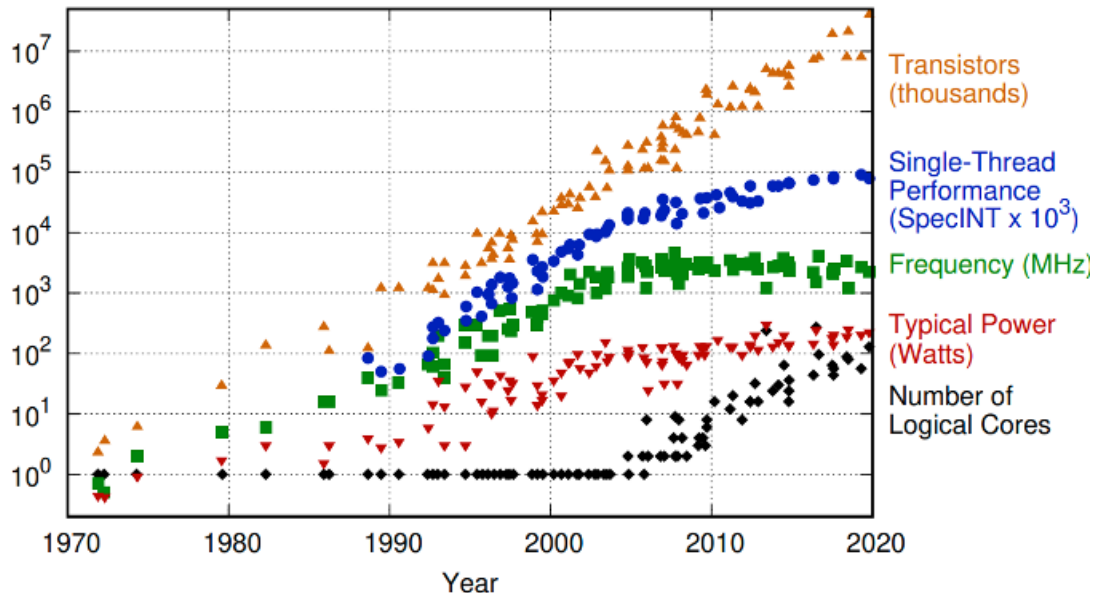
Генеративные модели в физике (FastSim)

Более 80% вычислительных ресурсов экспериментов по физике частиц тратится на симуляцию данных этих экспериментов

Закон Мура вышел в насыщение

Требуются новые решения для симуляции растущего количества данных на современных экспериментальных установках

48 Years of Microprocessor Trend Data



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2019 by K. Rupp

Генеративные модели в ФВЭ (FastSim)

Более 80% вычислительных ресурсов ФВЭ тратится на симуляцию данных экспериментов

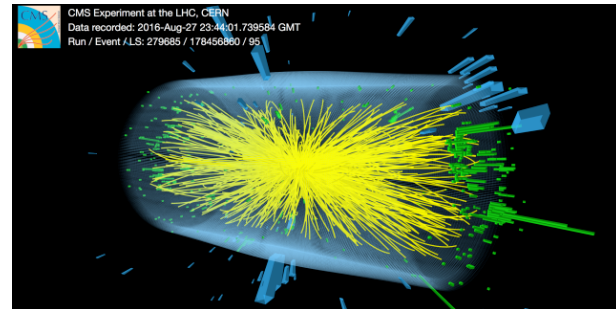
Закон Мура вышел в насыщение

▶ требуются новые решения для симуляции данных высоких загрузок LHC

Симуляция GEANT - переходная функция между (немногими) параметрами частицы и (макроскопическим) откликом детектора

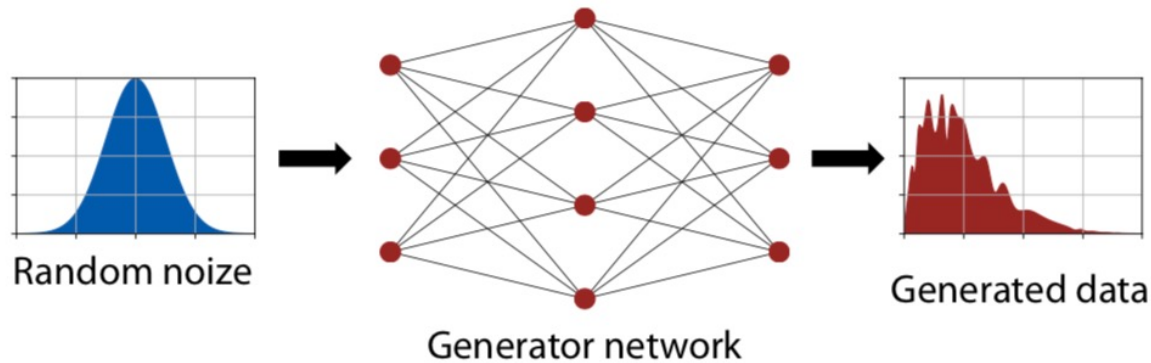
▶ вычисляемая на микроскопическом уровне детализации

Идея: натренировать (более простую и быструю) суррогатную модель для воспроизведения требуемой не многомерной переходной функции

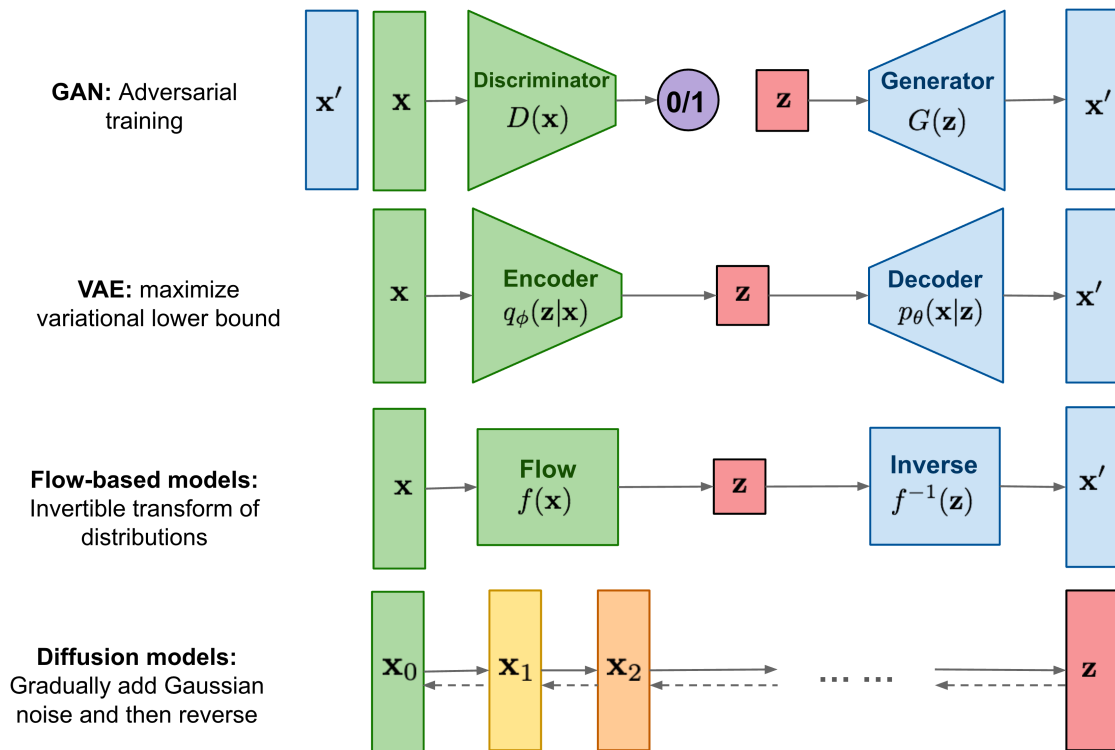


Генеративная модель ML

Обучаем переходную функцию из стандартного распределения в требуемое распределение

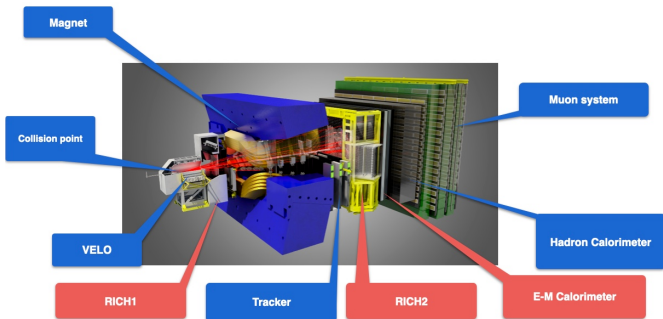


Виды (обучения) генеративных моделей



Практика

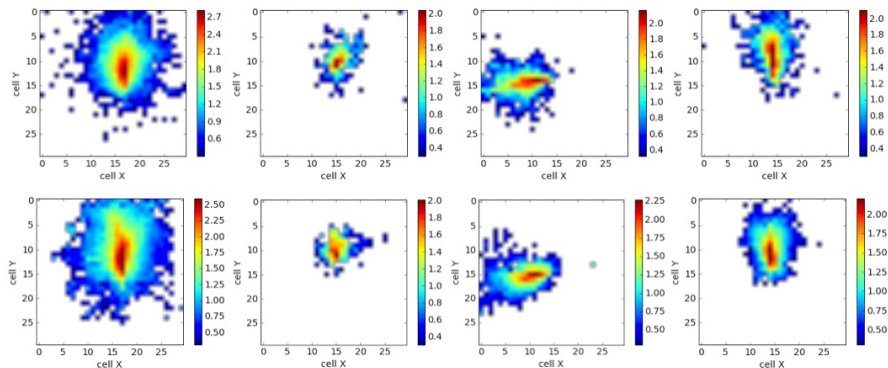
Генеративная модель для калориметра



GEANT Simulated

$\log_{10}(\text{cell energy})$

GAN Generated



Хорошее визуальное
соответствие

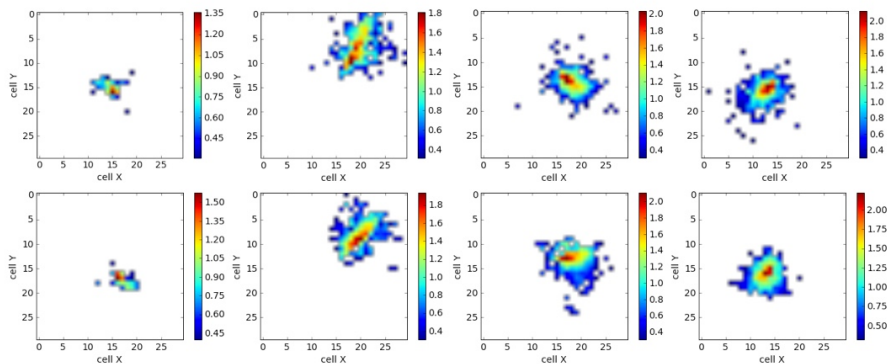
▶ приятные человеческому
глазу “котики”

Как с объективными
научными характеристиками?

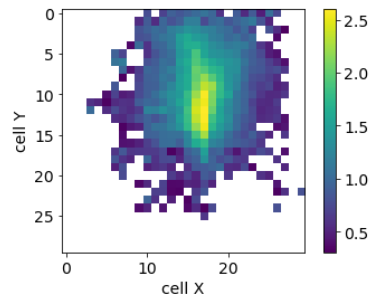
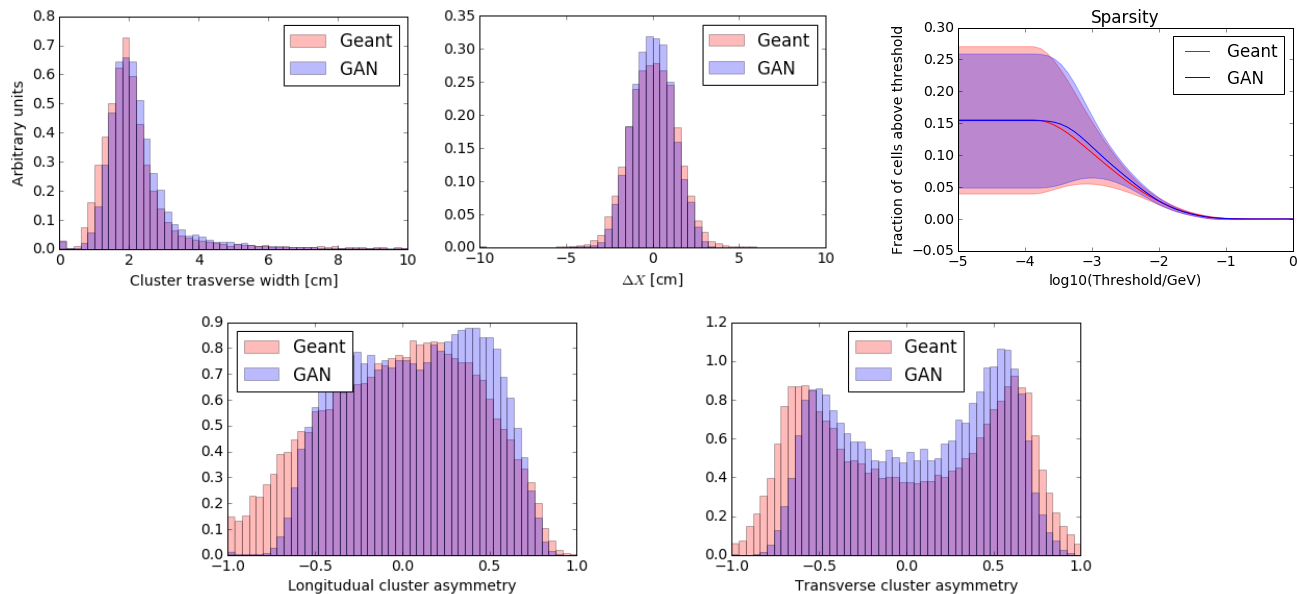
GEANT Simulated

$\log_{10}(\text{cell energy})$

GAN Generated



Распределения: хорошие и не очень



Сгенерировать выглядящие похоже отклики - несложно

Сложно воспроизвести маргинальные распределения

особенно, если их список не известен *a priori*

Обучение модели

- ▶ Требуется соответствующий обучающий датасет, созданный с использованием базового стимулятора (GEANT)

Валидация модели

- ▶ Требуются обучающий датасет, а также новый, созданный с использованием вновь обученной модели

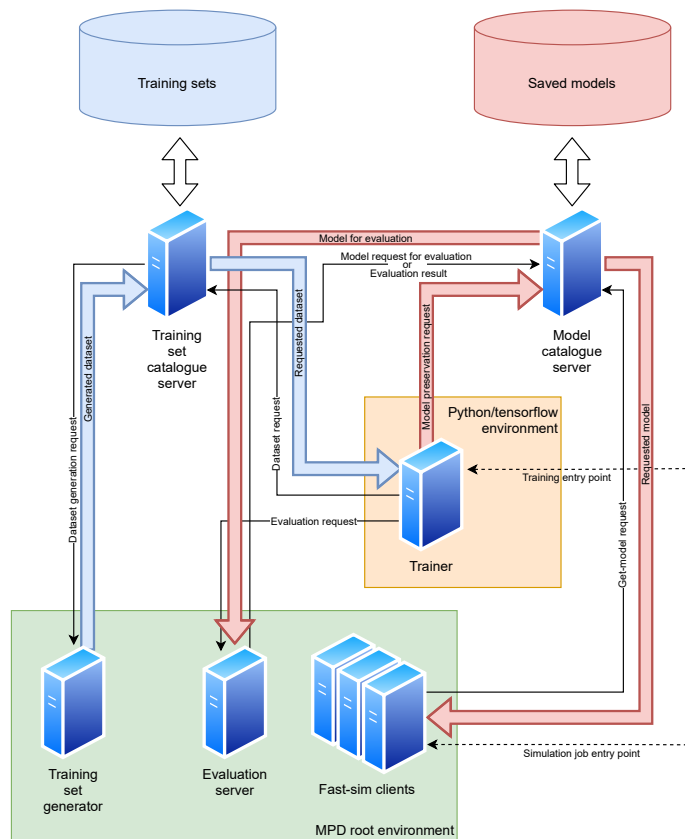
Каталогизация и сохранение модели

- ▶ Требуется уникальный ключ (хэш) модели, отражающий условия и конфигурации программного стека при создании и тренировке модели

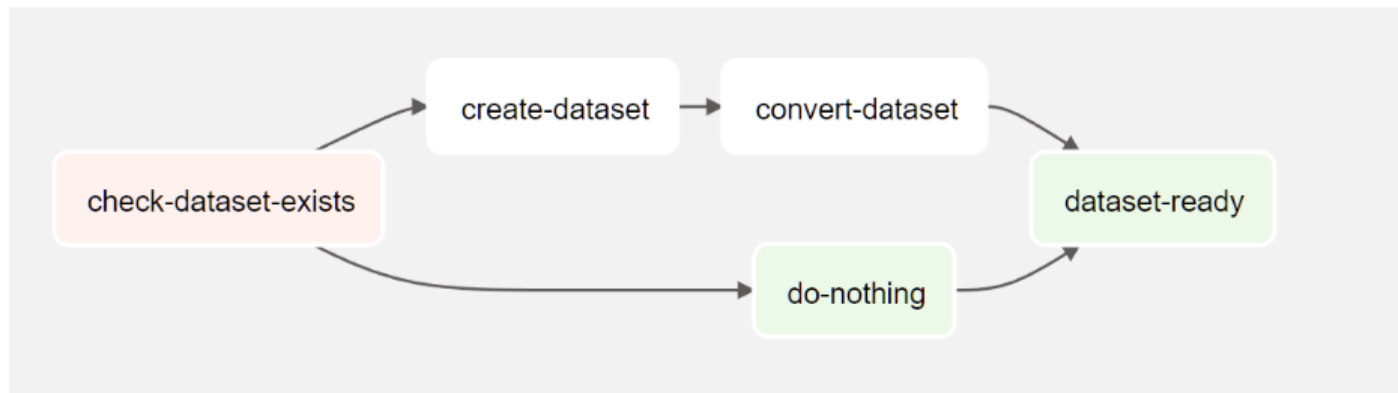
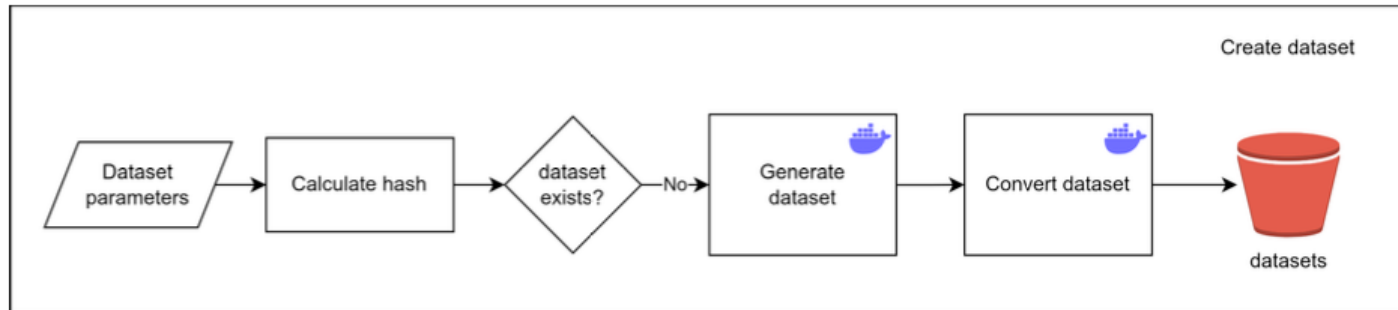
Использование модели для быстрой симуляции

- ▶ Требуется найти нужную модель в хранилище и загрузить ее в программный стек для использования

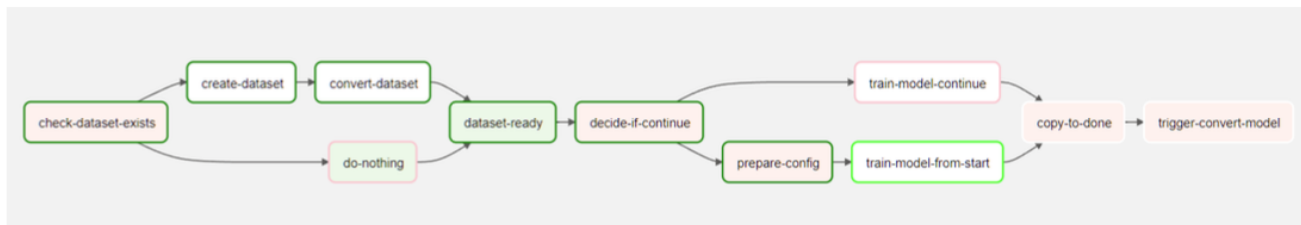
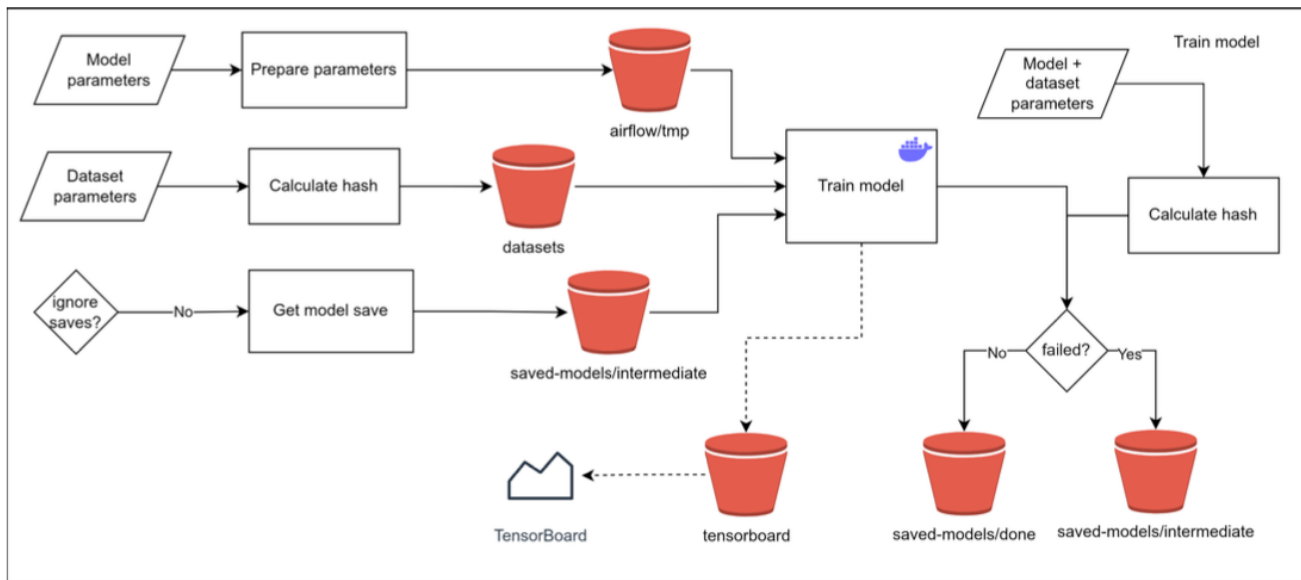
Взаимодействие компонент



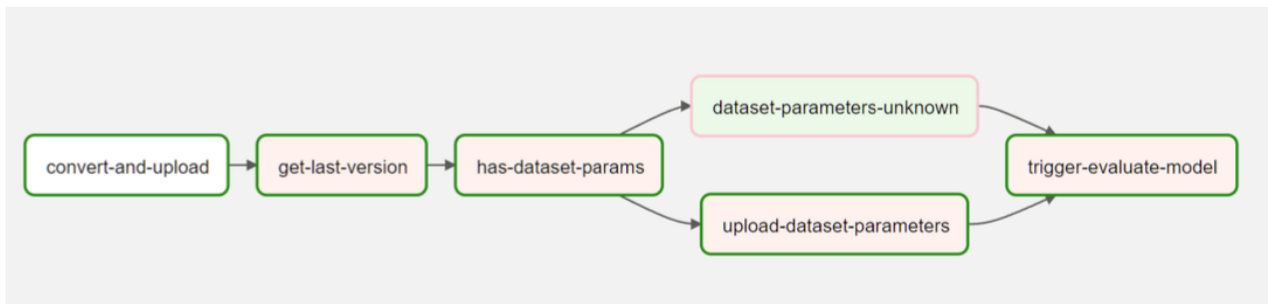
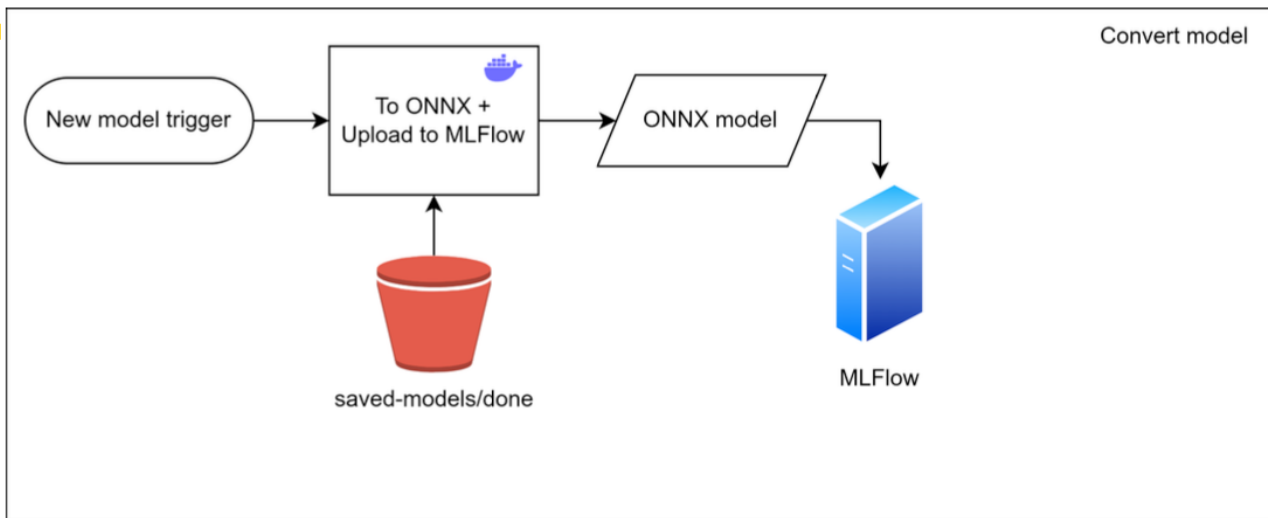
Создание обучающего датасета



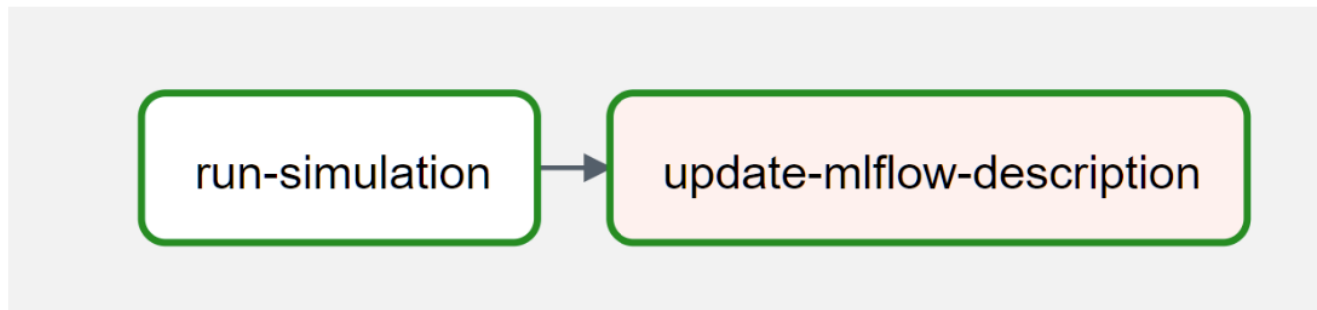
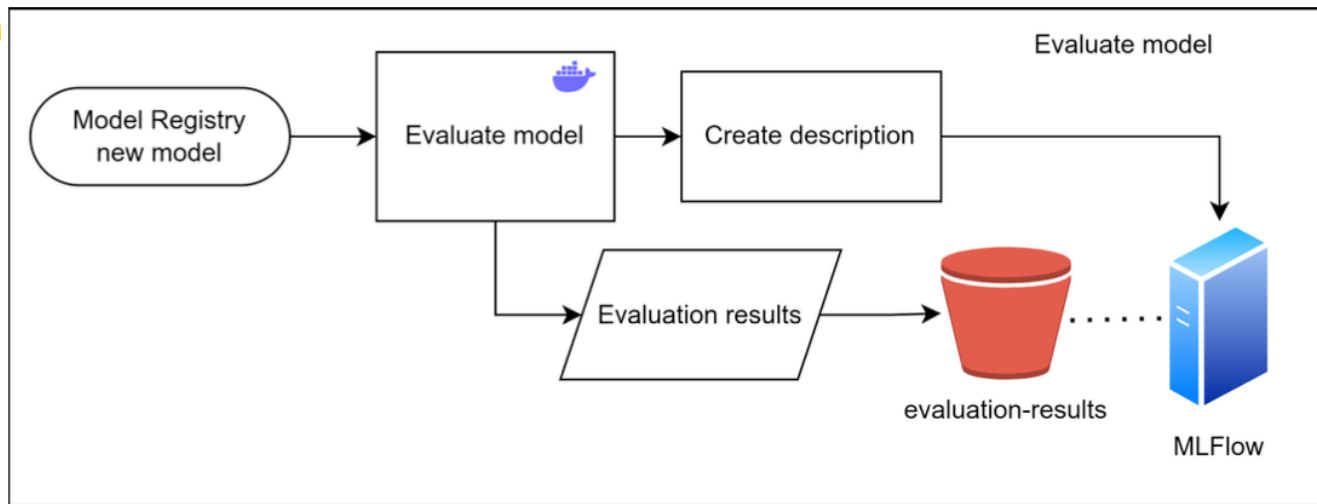
Обучение новой модели



Экспорт новой модели на сервер каталогизации



Валидация новой модели

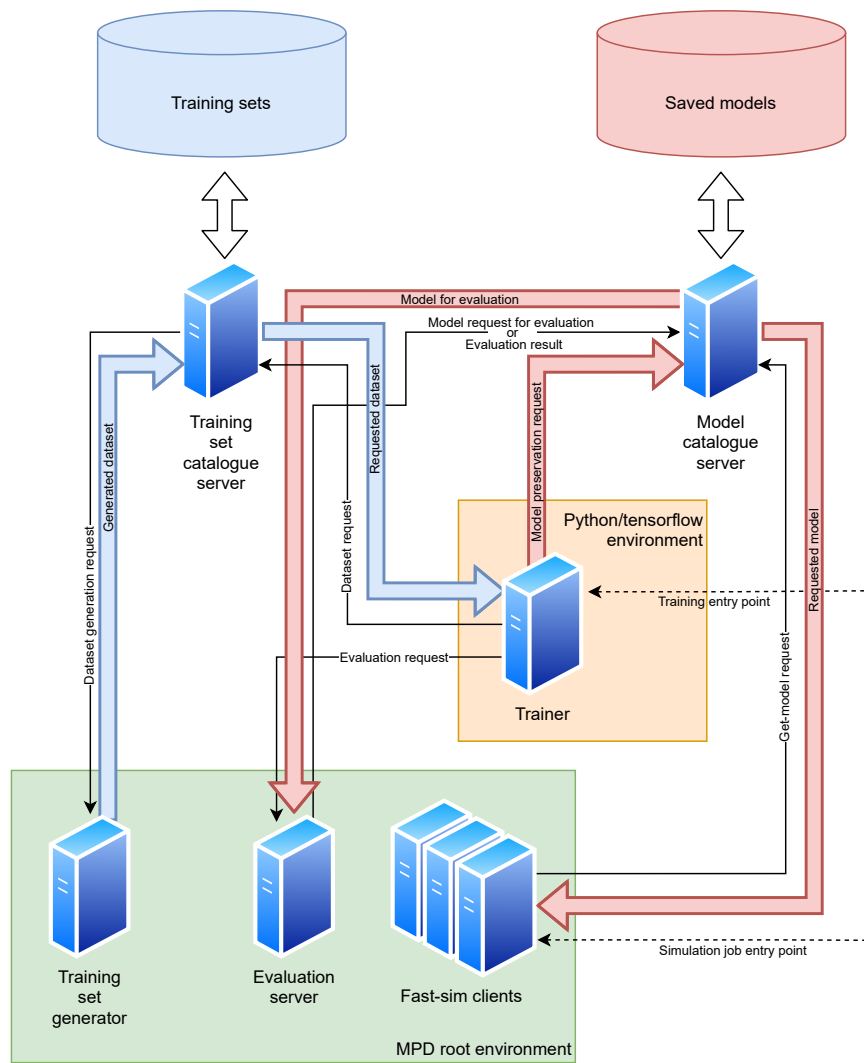


Использование модели в программном стеке

Модель тренируется на специальном оптимизированном сервере

Возможные способы загрузки:

- Модель компилируется в динамическую библиотеку
 - ✓ наиболее эффективный локальный код
 - Тяжелый файл, зависит от системы (GLib)
- Модель конвертируется в специальный формат (например ONNX) ► наш выбор
 - ✓ Универсальный загрузчик моделей
 - Дополнительная консервация, не всякая модель конвертируется, некоторый оверхед
- Модель отрабатывается на специальном сервере
 - ✓ наиболее эффективное исполнение модели на оптимизированном железе
 - Требуется глубокой параллелизации программного стека для эффективной работы в режиме больших матчей



Заключение

Использование суррогатных генеративных моделей может существенно помочь в научных исследованиях, требующих массовой симуляции данных для интерпретации измерений

Требования к таким моделям сильно отличаются от обычно используемых

- ▶ адекватная информативность и вариативность генерируемых объектов
- ▶ приоритизация фокуса обучения на физически мотивированные, часто маргинальные метрики
- ▶ количественная оценка систематических неопределенностей физических результатов, вносимых вариативностью и смещением модели

Эти проблемы находятся вне фокуса разработчиков генеративных моделей, предназначенных для бизнеса

- ▶ специальные усилия требуются для превращения генеративных моделей в мощный инструмент научных исследований

Выше были представлены некоторые подходы к этим проблемам