

В блоке 3 система выполняет оценку рыночной и операционной привлекательности каждого кандидата.

Для этого используются показатели, которые в реальной фармацевтической индустрии применяются в стратегическом планировании и финансовой аналитике. Ниже приводится объяснение ключевых метрик и источников, из которых они обычно получаются.

1. Success Probability — вероятность успешного выхода на рынок

В реальном мире вероятность успешного прохождения клинических фаз не является исходными данными для каждой молекулы.

Этот показатель рассчитывается аналитиками на основе *агрегированной исторической статистики*:

- FDA Clinical Development Success Rates
- отчёты IQVIA, GlobalData, EvaluatePharma
- исследования Tufts Center for Drug Development

Например, согласно статистике FDA:

- в онкологии только около **8–12%** кандидатов проходят путь от фазы I до регистрации;
- вероятность успеха зависит от механизма действия, токсичности, размера рынка и качества предклинических данных.

В проекте этот показатель моделируется как комбинация:

- потенциала кандидата (potential_label),
- токсичности,
- размера рынка,
- уровня конкуренции,
- небольшого случайного фактора.

Таким образом, `success_probability` в системе — **аналитическая оценка риска**, а не результат ML-модели.

2. Traditional Development Time — стандартная продолжительность разработки

Продолжительность разработки лекарства — величина, хорошо известная фармацевтической индустрии.

По данным FDA, Tufts и Deloitte:

- доклинические исследования: **2–4 года**
- фаза I: **1–2 года**
- фаза II: **2–3 года**
- фаза III: **3–5 лет**
- регистрация: **1–2 года**

Среднее значение:

 **8–14 лет до выхода на рынок**

Этот диапазон широко используется в стратегическом планировании фармацевтических компаний.

В проекте он задаётся напрямую, как реалистичная константа для сравнения с ИИ-подходом.

3. AI-Accelerated Time — модель ИИ-ускоренного цикла разработки

ИИ-ускоренный pipeline — это не историческая статистика, а *сценарная оценка*, основанная на исследованиях:

- DeepMind: “AI-first drug discovery reduces time by up to 10×”
- Insilico Medicine: автоматизированные генераторы молекул
- Recursion & NVIDIA: ускоренные ин-силико скрининги

- Exscientia: ИИ-оптимизация доклинических этапов

Большинство компаний заявляют, что:

- 👉 **ранние этапы могут быть сокращены 5–10 раз**
- 👉 preclinical + hit identification могут занимать **от нескольких месяцев до 1–3 лет**

Поэтому `ai_time_years = 1-3` — это реалистичная модель ускорения, а не фактические данные.

4. Expected Profit Score — интегральная оценка прибыльности

В реальной фарме прибыльность рассчитывается сложными моделями:

- Net Present Value (NPV)
- Return on Investment (ROI)
- патентный срок защиты
- доля рынка (Market Share)
- ожидаемый объём продаж по аналогам

Эти данные получают из:

- GlobalData
- EvaluatePharma
- финансовых отчётов компаний
- патентных портфелей

В проекте `expected_profit_score` упрощён, но использует те же факторы:

- потенциал (эффективность/новизна)
- токсичность
- размер рынка

- уровень конкуренции

Этот показатель помогает системе ранжировать проекты **с точки зрения потенциальной коммерческой эффективности**.



Итоговое объяснение

Таким образом, метрики блока 3 **не являются обучаемыми моделями и не извлекаются напрямую из исходных данных**.

Это бизнес-показатели, которые:

- в реальном мире рассчитываются аналитиками на основании рыночной статистики,
- а в нашем проекте моделируются синтетически по реалистичным правилам,
- чтобы система могла выполнять сравнение кандидатов «научный потенциал × рыночная привлекательность».