

УДК 004.89  
Код ГРНТИ 28.23.15

doi 10.54708/19926502\_2026\_3011113

## Применение технологий генеративного искусственного интеллекта к задаче календарного планирования производства

А.С. Дьячков<sup>1, 2</sup>, Д.А. Ризванов<sup>2\*</sup>

<sup>1</sup>ООО «Газпромнефть-Цифровые решения», г. Уфа, Республика Башкортостан, Россия

<sup>2</sup>ФГБОУ ВО «Уфимский университет науки и технологий», г. Уфа, Республика Башкортостан, Россия

**Аннотация.** В статье рассматривается задача календарного планирования производства в условиях высокой динамики, неопределенности и неоднородности ресурсов. Традиционные подходы, основанные на централизованных алгоритмах и ERP/MES-системах, часто игнорируют индивидуальные характеристики оборудования и персонала, а также слабоформализованные технологические ограничения, что снижает эффективность и адекватность расписаний. Для преодоления этих ограничений предложена оригинальная гибридная архитектура, интегрирующая многоагентную систему и технологии генеративного искусственного интеллекта. На основе ранее предложенных методологических основ формализована многоагентная модель в виде кортежа, включающего агентов-ресурсов, агентов-потребителей, онтологическую базу знаний, семантические ограничения и локальные/глобальные критерии оптимизации. Ключевым нововведением является встраивание большой языковой модели через архитектуру Retrieval-Augmented Generation, что позволяет генерировать адаптивные стратегии планирования, обеспечивать объяснимость решений и моделировать сценарии при сбоях без риска галлюцинаций. Предложен двухстадийный алгоритм взаимодействия агентов: на первой стадии формируется допустимое расписание с учетом доступности ресурсов, на второй – проводится его оптимизация через интеллектуальные переговоры с участием генеративного ИИ. Экспериментальная проверка в симуляционной среде, имитирующей механообрабатывающий цех, показала эффективность предлагаемого подхода. Результаты подтверждают, что интеграция генеративного ИИ в многоагентные системы создает когнитивный слой, повышающий не только эффективность, но и доверие к системе, что критически важно для внедрения ИИ в реальные производственные и управленческие процессы.

**Ключевые слова:** календарное планирование производства, многоагентный подход, генеративный искусственный интеллект.

\*ridmi@mail.ru

### Введение

Современные производственные предприятия сталкиваются с растущей сложностью оперативного управления: динамичная внешняя среда, непредсказуемые сбои, гибкие заказы и высокая вариативность продукции требуют адаптивных и интеллектуальных решений для календарного планирования. Эта задача заключается в распределении ограниченных ресурсов: оборудования, персонала и инструментов – по операциям технологического процесса с целью минимизации общего времени выполнения производственного плана.

Традиционные ERP/MES-системы и централизованные алгоритмы планирования, основанные на детерминированных моделях, плохо справляются с неопределенностью и слабоформализованными ограничениями, характерными для реального производства. Они часто игнорируют индивидуальные характеристики ресурсов (например, разную производительность станков или квалификацию операторов), что приводит к неоптимальным или даже нереализуемым расписаниям.

В последние годы активно развиваются подходы, сочетающие многоагентные системы (МАС) и генеративный искусственный интеллект. МАС позволяют моделировать производственную среду как самоорганизующуюся экосистему автономных агентов, способных к локальному принятию решений и кооперации. В то же время генеративный ИИ, особенно в виде больших языковых моделей (LLM), добавляет системе когнитивные способности: анализ

контекста, синтез новых стратегий, генерацию объяснимых решений и адаптацию к изменениям в реальном времени.

Однако большинство существующих исследований ограничиваются концептуальными идеями и не предлагают строгой формальной постановки задачи.

В настоящей работе предлагается подход на основе интеграции технологий генеративного искусственного интеллекта и многоагентных систем к задаче календарного планирования производства.

Такой подход позволяет не только повысить эффективность расписаний, но и обеспечить объяснимость решений для операторов и менеджеров – ключевой аспект при внедрении ИИ в промышленные процессы.

### Обзор литературы

Задача календарного планирования производства традиционно рассматривается как комбинаторная оптимизационная проблема с NP-трудной сложностью. Классические методы, от эвристик до метаэвристик, подробно описаны в монографии Пинедо [1].

В последние десятилетия для повышения адаптивности планирования к неопределенности и динамике производства активно применяются многоагентные системы. В работах [2–4] показано, что МАС позволяют моделировать производственную среду как сеть автономных сущностей (агентов-станков, агентов-заказов), способных к локальному принятию решений и кооперации без центрального контроллера. Такой подход особенно эффективен при сбоях, изменениях в приоритетах или при появлении срочных заказов.

С появлением больших языковых моделей открылись новые возможности. LLM могут выступать как универсальные «когнитивные посредники», способные интерпретировать неструктурированные данные, генерировать объяснимые решения и предлагать новые стратегии.

Например, исследование [5] изучает потенциальные возможности интеграции больших языковых моделей в производственный сектор, уделяя особое внимание их потенциалу для повышения эффективности, улучшения принятия решений и качества продукции.

Несмотря на прогресс, остается существенный пробел: большинство работ либо фокусируются на чисто оптимизационных методах, либо предлагают концептуальную интеграцию ИИ без строгой формализации. Особенно не хватает исследований, сочетающих формальную многоагентную модель планирования, онтологическую базу знаний и генеративный ИИ с контролируемым выводом (например, через RAG).

### Многоагентная модель задачи календарного планирования

В соответствии с методологией, предложенной в [6, 7], задача календарного планирования рассматривается как задача распределения ресурсов между потребителями ресурсов в условиях семантических ограничений. В производственном контексте в качестве потребителей ресурсов выступают детали и операции, а в качестве поставщиков ресурсов – рабочие, станки, инструменты и другое оборудование.

Формально многоагентная система (MAS) для управления ресурсами при календарном планировании представляется в виде кортежа:

$$MAS = \langle A_{\text{рес}}, A_{\text{потр\_рес}}, Act_{\text{спец}}, Act_{\text{спотр\_рес}}, KB, Constraints, Tasks, F \rangle,$$

где  $A_{\text{рес}} = \{a_{\text{рес}1}, \dots, a_{\text{рес}N}\}$  – множество агентов-ресурсов;

$A_{\text{потр\_рес}} = \{a_{\text{потр\_рес}1}, \dots, a_{\text{потр\_рес}M}\}$  – множество агентов-потребителей ресурсов;

$Act_{\text{спец}}$  – множества допустимых действий агентов;

$KB$  – база знаний, содержащая онтологическую модель предметной области;

$Constraints$  – множество технологических и семантических ограничений;

$Tasks$  – множество задач, которые должны быть решены;

$F = \{F_1, \dots, F_S, F_0\}$  – множество локальных и глобальных критериев оценки.

Каждый агент обладает автономностью, способностью к коммуникации и целенаправленному поведению. Взаимодействие агентов строится на принципах самоорганизации и кооперации, что позволяет системе адаптироваться к изменениям в реальном времени.

Рассмотрим производственный участок, на котором необходимо изготовить  $N$  деталей. Каждая деталь  $i$  требует выполнения последовательности из  $E_i$  операций. Для выполнения операции  $j$  детали  $i$  требуется набор ресурсов  $R_{ij}$ , включающий оборудование и персонал определенной квалификации.

Введем следующие обозначения:

$D$  – множество деталей,  $|D| = N$ ;

$O$  – множество операций,  $|O| = E$ ;

$O_i$  – множество операций, необходимых для изготовления детали  $i$ .  $|O_i| = E_i$ ,  $O_i \subset O$ ;

$R$  – множество ресурсов,  $|R| = M$ ;

$R_{ij}$  – множество ресурсов, необходимых для выполнения операции  $j$  для изготовления детали  $i$ .  $|R_{ij}| = M_{ij}$ ;

$L_{ij}$  – нормативное время выполнения операции  $j$  детали  $i$ ;

$S_r \in [0,5; 2,0]$  – коэффициент производительности ресурса  $r$ ;

$T_{ijm}^B, T_{ijm}^E$  – моменты начала и окончания выполнения множеством ресурсов  $m$  операции  $j$  для изготовления детали  $i$ ;

$L_{ijR_y}^{PW}$  – длительность простоя, не связанного с настройкой станка (длительность ожидания ресурса).

Тогда общая длительность изготовления детали  $i$  определяется по формуле:

$$L_i = \sum_{j=1}^{E_i} \left( \frac{L_{ij} + L_{ij}^{PBefore} + L_{ij}^{PAfter}}{f(MX, S, i, j)} + L_{ij}^{PW} \right),$$

где  $f(MX, S, i, j)$  – функция, учитывающая производительность назначенных ресурсов.

Целевая функция задачи – минимизация общего времени выполнения всего производственного плана:

$$F_0 = \max_i(T_i^E) - \min_i(T_i^B) \rightarrow \min.$$

Ограничения задачи включают соответствие количества задействованных ресурсов технологическим требованиям:

1. Количество задействованных ресурсов при выполнении операции  $j$  для изготовления детали  $i$  должно соответствовать требуемым количествам согласно технологии изготовления:

$$\sum_{r \in R_k} MX_{ijr} = MR_{ijk},$$

где  $R_k$  – множество ресурсов категории  $K$ , используемых для выполнения операции  $j$  для изготовления детали  $i$ .

2. Квалификация ресурса не должна быть ниже сложности выполняемых им работ:

$$Q_r \geq QR_r,$$

где  $Q_r$  – значение свойства «квалификация» ресурса  $r$  – числовое значение  $[1; 10]$ ;  $QR_r$  – значение свойства «сложность работ» операции  $r$  – числовое значение  $[1; 10]$ .

Эта постановка позволяет учитывать как количественные, так и качественные характеристики ресурсов, что является ключевым преимуществом перед классическими моделями.

### Интеграция генеративного ИИ в многоагентную систему

Генеративный ИИ (в частности, большие языковые модели, LLM) встраивается в МАС на уровне базы знаний и агентов принятия решений. Его функции включают:

1. Синтез стратегий планирования. LLM анализирует исторические данные, текущую загрузку ресурсов и семантические ограничения, чтобы генерировать новые правила диспетчеризации или эвристики для агентов. Например, при появлении срочного заказа модель

может предложить временную перегруппировку приоритетов на основе анализа критичности заказчика и загрузки оборудования.

2. Генерация текстовых пояснений. При любом изменении расписания система автоматически формирует человеко-понятное объяснение: «Изменение вызвано поломкой станка № 5. Операция перенесена на станок № 7, что увеличивает срок на 2 часа. Рекомендуется уведомить отдел логистики».

3. Прогнозирование и моделирование сценариев. Генеративный ИИ может генерировать синтетические данные о возможных сбоях (болезнь персонала, задержка поставок) и моделировать их влияние на расписание, что позволяет системе заранее выработать компенсационные меры.

Интеграция реализуется через архитектуру Retrieval-Augmented Generation (RAG), где LLM получает контекст из онтологической базы знаний и текущего состояния МАС. Это предотвращает галлюцинации и обеспечивает привязку к реальным производственным ограничениям.

Алгоритм взаимодействия агентов с использованием генеративного ИИ

Процесс планирования состоит из двух стадий.

Стадия 1. Формирование допустимого расписания.

Агент-деталь инициирует последовательность операций. Агент-операция запрашивает у агентов-ресурсов информацию о доступности. При конфликтах за ресурсы используется локальный критерий приоритета. Генеративный ИИ на этом этапе может предложить альтернативные наборы ресурсов на основе анализа их производительности и текущей загрузки.

Стадия 2. Оптимизация расписания.

Агент-расписание анализирует полученный план, выявляет критические операции и инициирует переговоры между агентами для перераспределения ресурсов. Генеративный ИИ генерирует предложения по обмену: «Переназначение операции 3 детали А на станок 2 снизит общее время на 15 минут при условии, что деталь Б согласится на перенос своей операции на 30 минут».

Коммуникация между агентами осуществляется через стандартные протоколы FIPA ACL, а генеративная модель выступает в роли «интеллектуального посредника», формализующего и обогащающего диалог.

### **Экспериментальная часть**

Экспериментальная проверка предложенного подхода проводилась в специально разработанной симуляционной среде. Среда создана с учетом следующих ключевых требований:

– реалистичность производственных процессов: учтены последовательности технологических операций, зависимости между деталями и операциями, наличие альтернативных ресурсов для выполнения одной и той же операции, а также неоднородность характеристик ресурсов (производительность оборудования, квалификация персонала);

– поддержка динамических событий: среда позволяет имитировать внезапные изменения: поломки оборудования, отсутствие персонала, поступление срочных заказов – и оценивать реакцию системы на них;

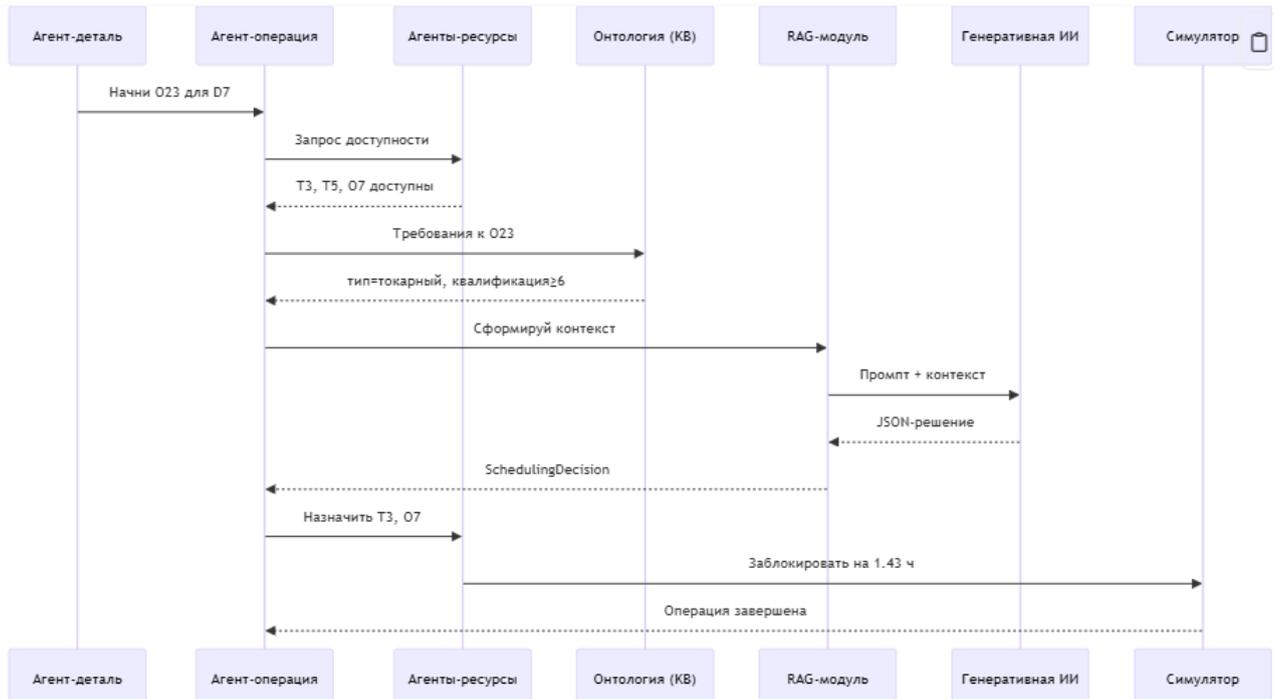
– интеграция с многоагентной архитектурой и генеративным ИИ: каждый производственный объект (станок, оператор, деталь, операция) представлен в виде программного агента, способного взаимодействовать с другими агентами и запрашивать решения у генеративной модели.

Симулятор состоит из следующих модулей:

– модуль инициализации производства: загружает данные о деталях, операциях и ресурсах из JSON-файлов, сгенерированных на основе реальных технологических маршрутов;

- модуль онтологии: хранит и управляет OWL-онтологией (с использованием библиотеки owlready2), которая содержит классы Operation, Resource, Tool, Skill, а также связи между ними (например, «операция требует инструмент», «ресурс обладает навыком»);
- MAC-ядро: реализует агентов и их взаимодействие через очередь сообщений, совместимую с логикой FIPA ACL;
- интерфейс с LLM (Llama-3-8B-Instruct);
- модуль сбора метрик: фиксирует загрузку ресурсов, время отклика системы и другие показатели.

На Рис. 1 представлен фрагмент диаграммы взаимодействия основных компонентов разработанной системы, отражающий запросы агентов многоагентной системы к онтологии и большой языковой модели.



**Рисунок 1.** Диаграмма взаимодействия компонентов системы календарного планирования с генеративным ИИ.

Эксперимент проводился в симуляционной среде, имитирующей участок механообрабатывающего цеха. Среда реализована на языке Python с использованием библиотек SimPy (для дискретно-событийного моделирования) и LangChain (для интеграции LLM).

Онтологическая база знаний в формате OWL, описывающая связи между деталями, операциями, ресурсами и технологическими ограничениями реализована на основе Protégé.

В качестве генеративной модели использовалась Llama-3-8B-Instruct, интегрированная через архитектуру RAG. При каждом запросе LLM получала релевантные фрагменты из онтологии и текущего состояния MAC.

Эксперимент проводился на трех наборах задач:

- малый: 15 деталей, 42 операции;
- средний: 50 деталей, 138 операций;
- крупный: 100 деталей, 287 операций.

Для оценки эффективности предложенного подхода использовались три базовых метода: классическая эвристика SPT (Shortest Processing Time), многоагентная система без генеративного ИИ, предложенный метод с интеграцией LLM через RAG.

Анализ показал, что предложенный метод обеспечивает снижение общего времени выполнения календарного плана на 12–18% по сравнению с классической MAC и на 18–25%

по сравнению с SPT. При этом загрузка оборудования выросла на 4–6%, что свидетельствует о более эффективном использовании ресурсов.

Эксперимент подтвердил работоспособность и эффективность предложенного подхода. Интеграция генеративного ИИ в MAS позволяет не только улучшить оптимизационные метрики, но и повысить доверие пользователей за счет прозрачности и адаптивности.

### Заключение

В настоящей статье предложена и экспериментально апробирована интегрированная архитектура календарного планирования, сочетающая формализованную многоагентную модель и технологии генеративного искусственного интеллекта. Предлагаемый подход основан на многоагентной модели задачи календарного планирования и расширен за счет когнитивных возможностей больших языковых моделей (LLM).

Ключевыми аспектами работы являются:

- формализация MAS-кортежа, учитывающего как количественные, так и качественные (семантические ограничения) характеристики производственной среды;
- внедрение генеративного ИИ через архитектуру RAG, что обеспечивает синтез адаптивных стратегий планирования, генерацию объяснимых решений и моделирование сценариев без риска галлюцинаций;
- двухстадийный алгоритм взаимодействия агентов, в котором LLM выступает в роли интеллектуального посредника при распределении и перераспределении ресурсов.

Экспериментальные результаты подтвердили эффективность предложенного подхода по сравнению с классическими эвристиками и многоагентными системами.

Интеграция генеративного ИИ в многоагентную систему не заменяет традиционное планирование, а дополняет его когнитивным слоем, способным интерпретировать контекст, генерировать аргументированные решения и поддерживать доверие со стороны персонала. Это особенно актуально для внедрения ИИ в реальные производственные системы, где объяснимость и адаптивность часто важнее формальной оптимальности.

### Литература:

1. Pinedo M. L. *Scheduling: Theory, Algorithms, and Systems*. 6th ed. Cham: Springer Nature, 2022. 698 p. [Webpage] URL: <https://doi.org/10.1007/978-3-031-05921-6>.
2. Shen Weiming, Wang Lihui, Hao Qi. Agent-based distributed manufacturing process planning and scheduling: A state-of-the-art survey // *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 2006. Vol. 36. 563–577. DOI: 10.1109/TSMCC.2006.874022.
3. Leitão P. Agent-based distributed manufacturing control: A state-of-the-art survey // *Engineering Applications of Artificial Intelligence*. 2009. No. 22. P. 979–991. DOI: 10.1016/j.engappai.2008.09.005.
4. Pu Yu, Li Fang, Rahimifard Sh. Multi-agent reinforcement learning for job shop scheduling in dynamic environments // *Sustainability*. 2024. No. 16. Art. 3234. DOI: 10.3390/su16083234.
5. Wulf J., West Sh., Anderson M., Mueller-Csernetzky P., Meierhofer J. On the value potential of large language models in the manufacturing industry. In: *Smart Services Summit*. Springer, 2025. P. 135–147. DOI: 10.1007/978-3-031-86958-7\_10.
6. Ризванов Д.А. Методологические основы поддержки принятия решений при управлении ресурсами в сложных системах // *Вестник Ижевского государственного технического университета*. 2018. Т. 21. № 4. С. 200–207. [Rizvanov D.A. Methodological bases of decision support for resource management in complex systems // *Vestnik IzhGTU imeni M.T. Kalashnikova*. Vol. 21. No. 4. P. 200–207 (in Russian)].
7. Ризванов Д.А., Юсупова Н.И. Применение интеллектуальных технологий управления ресурсами при календарном планировании производства // *Интеллектуальные системы в производстве*. 2018. Т. 16. № 4. С. 130–137. [Rizvanov D.A., Yussupova N.I. Application of intelligent resource management technologies in the production scheduling // *Intellektual'nye sistemy v proizvodstve*. Vol. 16. No. 4. P. 130–137 (in Russian).]

**Об авторах:**

**РИЗВАНОВ Дмитрий Анварович**, д.т.н., доцент, профессор кафедры вычислительной математики и кибернетики, ФГБОУ ВО «Уфимский университет науки и технологий», ridmi@mail.ru.

**ДЬЯЧКОВ Алексей Сергеевич**, к.т.н., руководитель направления Управления технологического развития ООО «Газпромнефть-Цифровые решения», доцент кафедры автоматизированных систем управления ФГБОУ ВО «Уфимский университет науки и технологий», dyachkov.ase@gazprom-neft.ru.

**Metadata:**

**Title:** Applying Generative Artificial Intelligence Technologies to the Production Scheduling Problem.

**Author 1:** Dmitrii Anvarovich Rizvanov, Doctor of Technical Sciences, Docent, Professor of the Department of Computational Mathematics and Cybernetics, Ufa University of Science and Technology, 32 Zaki Validi St., 450076 Ufa, Russia, ridmi@mail.ru, ORCID ID 0000-0003-2378-5587, Web of Science ResearcherID L-4068-2016, Scopus Author ID 54394082100.

**Author 2:** Aleksey Sergeevich Dyachkov, Candidate of Technical Sciences, Head of Department, Technological Development Directorate, Gazprom Neft Digital Solutions LLC, 5 Bldg. 4 Kievskaya St., 190013 Saint Petersburg, Russia. Associate Professor, Department of Automated Control Systems, Ufa University of Science and Technology, 32 Zaki Validi St., 450076 Ufa, Russia.

**Abstract:** The paper addresses the problem of production scheduling under conditions of high dynamics, uncertainty, and resource heterogeneity. Traditional approaches based on centralized algorithms and ERP/MES systems often overlook individual characteristics of equipment and personnel, as well as poorly formalized technological constraints, which reduces the efficiency and adequacy of generated schedules. To overcome these limitations, an original hybrid architecture is proposed that integrates a multi-agent system (MAS) with generative artificial intelligence (GenAI) technologies. Building upon previously developed methodological foundations, a formal MAS model is defined as a tuple comprising resource agents, consumer agents, an ontological knowledge base, semantic constraints, and local/global optimization criteria. The key innovation lies in embedding a large language model (LLM) via the Retrieval-Augmented Generation (RAG) architecture, which enables the system to generate adaptive scheduling strategies, provide human-interpretable explanations, and simulate failure scenarios – while minimizing the risk of hallucinations. A two-stage agent interaction algorithm is proposed: in the first stage, a feasible schedule is constructed based on resource availability; in the second stage, the schedule is optimized through intelligent negotiations facilitated by generative AI. Experimental validation in a simulation environment modeling a machining shop floor demonstrates the effectiveness of the proposed approach. The results confirm that integrating generative AI into multi-agent systems creates a cognitive layer that enhances not only scheduling performance but also user trust – a critical factor for the successful deployment of AI in real-world production and management processes.

**Keywords:** production scheduling, multi-agent approach, generative artificial intelligence.