

ОБУЧЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ФИЗИКЕ МИРА С ПОМОЩЬЮ WORLD MODELS

Минченко Д.Н.

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь*

Научный руководитель: Марамыгина Т.А. – ст. преподаватель кафедры ИКТ

Аннотация. Статья посвящена анализу качественного скачка в развитии искусственного интеллекта, произошедшему в 2025–2026 годах. Этот период характеризуется переходом от языковых моделей, которые умеют «говорить», к мировым моделям (World Models), которые понимают принципы физики.

Ключевые слова: искусственный интеллект, World models, LLM, воплощенный интеллект, физический ИИ, смена парадигмы

Введение. Представим «существо», которое прочитало миллиарды книг, выучило все энциклопедии мира, цитирует Э. Канта и пишет дипломы, но при этом оно никогда не держало в руках камень, не роняло чашку, не толкало тележку в супермаркете. Если его спросить: «Что будет, если положить тяжелую книгу на пустую коробку из-под обуви?», то получим правильный текстовый ответ – потому что в книгах написано «коробка сомнется». Однако, такое существо не понимает почему, так как у него нет внутренней модели физики мира [1].

Таковыми «существами» и являются ChatGPT, Claude, Gemini и другие большие языковые модели (LLM). Их фундаментальное ограничение доказано: LLM обучаются на текстах, предсказывая следующее слово по предыдущим [2]. Такие модели не имеют никакого опыта взаимодействия с реальностью. Они как будто «головы в банке», подключенные ко всем библиотекам мира.

Настоящая статья ставит целью объяснить, что такое World models простыми словами, показать реальные прорывы января-февраля 2026 года и доказать: 2026 год стал водоразделом, после которого ИИ перестал быть просто «разговорчивым библиотекарем» и начал осваивать физический мир [3].

Основная часть. Простыми словами world model – это внутреннее представление среды в «голове» искусственного интеллекта, которое позволяет ему предсказывать последствия своих действий. Например, можно считать, что у пятилетнего ребенка есть world model. Он знает: если толкнуть стакан, он упадет, если уронить яйцо, то оно разобьется, а большой кубик нельзя поставить на маленький, потому что конструкция упадет. LLM не знают этого. Они знают только, что слова «стакан», «упал», «разбился» часто встречаются рядом.

Главное отличие между рассматриваемыми моделями заключается в следующем:

- модели LLM предсказывают следующее слово, а World models предсказывают следующее состояние мира;
- LLM видят корреляцию, а World models понимают причину и следствие;
- LLM пишут план действий, World models проигрывают этот план в голове, проверяя его на внутренней модели реальности [4].

Янн ЛеКун, лауреат премии Тьюринга и ведущий исследователь ИИ, объясняет World models следующим образом: «Вы можете вообразить последовательность действий, которые собираетесь совершить, и ваша World model позволит вам предсказать, каким будет эффект этих действий на мир». Именно такой способности предугадывать последствия и лишены современные языковые модели, и именно эту способность разработчики пытаются дать им сегодня.

На наш взгляд одним из важных событий было открытие китайской компанией Ant Lingbo (дочерняя компания Alibaba) исходного кода модели LingBot-VA, который она

представила 10 февраля 2026 года [5]. Это первая реализация принципа «авторегрессивное видео-действие», которое заключается в том, что робот не просто выполняет заученные движения, а изучает задачу, проигрывает в голове несколько вариантов решения, выбирает тот, который с наибольшей вероятностью приведет к успеху, выполняет задачу и корректирует себя в реальном времени. Для обучения новой задаче LingBot-VA достаточно 30-50 демонстраций, а для других моделей требовались тысячи. Такие задачи, как складывание мягкой одежды, заправка постели, аналогов.

Практически одновременно командами Стэнфорда и NVIDIA была опубликована Cosmos Policy. Cosmos Policy ломает все шаблоны и представляет собой чистую видеогенерацию, превращенную в управление роботом. От обычного подхода, когда модель VLA смотрит картинку, читает текст, выдает действие, в модели Cosmos Policy используется подход генерации видео того, что должно произойти, из которого извлекаются управляющие команды для робота. Рассмотрим, зачем это нужно? Так как видеоданных в мире миллиарды часов, а данных о работе роботов часы, то Cosmos Policy позволяет использовать гигантские массивы обычных видео для обучения физическому интеллекту. Успешность Cosmos Policy на бенчмарке LIBERO, стандарте для тестирования роботов, составляет 98,5 %, что является абсолютным рекордом

Платформа NVIDIA Cosmos, которая была анонсирована на CES 2025, к январю 2026 уже скачана более 2 миллионов раз. Основными клиентами данной платформы являются: Uber, Figure AI, Waabi, XPENG, 1X, то есть все, кто делает реальные роботы и беспилотники [3]. Цель NVIDIA состоит в создании ChatGPT-момент для робототехники».

Поэтому ключевым понятием 2026 года и стало контрфактическое мышление, которое заключается в том, что робот учится задавать вопрос: «А что будет, если я сделаю по-другому?». Система DreamZero от NVIDIA реализует это через совместное предсказание будущего видео и соответствующих действий, т.е. робот больше не имитатор, а становится мыслителем.

Рассмотрим в чем же состоит технологическая суть современных моделей. Раньше роботов учили сначала создать виртуальную среду, натренировать в ней миллионы итераций и перенести на реального робота. При это возникала проблема, что симулятор никогда не совпадает с реальностью на 100 %, так как симулятор не знает ни физики жидкостей, ни деформации мягких материалов, ни сенсорных шумов и т.п. Главный ученый компании Ant Lingbo считает, что «Sim-to-Real – не наш основной технический путь. Мы используем физические данные реального мира плюс интернет-видео».

Второе ключевое понятие – это принцип «один мозг – много тел». Принцип традиционной робототехники, когда под каждую конструкцию робота разрабатывается своя программа. Например, если сменили захват, то соответственно, надо переписать и управление. Новая парадигма заключается в том, что разрабатывается универсальный мозг, так называемая, базовая модель, которая может управлять любым телом: двуногим, четвероногим, колесным, с двумя захватами или с пятью пальцами. LingBot-VLA адаптирован под 9 различных промышленных конфигураций роботов. Смысл такой же, как и в Android, то есть одна операционная система может использоваться для тысяч моделей смартфонов. В 2026 году такой подход используется и в робототехнике.

Ключевым моментом стало проведение в феврале 2026 года крупнейшей конференции по робототехнике ICRA, на которой работала отдельная секция «World Models» [6] и академическое сообщество официально признало, что World models – это магистральный путь к следующему поколению искусственного интеллекта [7].

Однако, в настоящий момент, для развития World models существуют определённые ограничения. Например, модели LLM учились на триллионах слов, а для World models необходимы триллионы кадров взаимодействия с физическим миром. Таких данных пока нет. NVIDIA Cosmos тренировался на 9000 триллионов токенов из 20 миллионов часов видео, а это требует инфраструктуры, доступной лишь нескольким корпорациям мира.

Ученый, один из основателей современного глубокого обучения Янн ЛеКун признает, что еще не известно точно, как именно должна быть устроена идеальная World model. I-JERA, DreamZero, Genie 3, LingBot-VA, Cosmos Policy являются только экспериментом, а не финальным решением, сохраняется еще теоретическая незавершенность.

Заключение. 2026 год войдет в историю искусственного интеллекта как год, когда ИИ перестал быть «головой в банке». Он начал обретать тело и через это тело понимать физику мира.

На основе проведенного исследования, можно выделить четыре аспекта, которые подтверждают эту смену парадигмы:

– технологический: LingBot-VA, Cosmos Policy, Motus доказали, что World models работают на реальных роботах, превосходя старые подходы на 20–35%;

– институциональный: Янн ЛеКун, главный авторитет в глубоком обучении, ставит на World models свою репутацию и 500 миллионов евро инвестиций;

– индустриальный: NVIDIA Cosmos скачан 2 миллиона раз, а Uber, Figure, Waabi внедряют физический ИИ в свои продукты;

– академический: ICRA и IEEE вводят World models в мейнстрим научной повестки.

Люди научили ИИ понимать, что камень упадет, но пока не научили его учиться физике так же эффективно, как пятилетний ребенок, потому что ребенку достаточно трех падений, роботу – тысячи демонстраций. Однако направление задано, и обратного пути нет.

Можно сказать, что 2026 год стал годом, когда ИИ наконец вышел из библиотеки в реальный мир.

Список литературы

1. Brodsky, S. World models help AI learn what five-year-olds know about gravity / S. Brodsky // IBM. – 2025. [Электронный ресурс]. Режим доступа: <https://www.ibm.com/think/news/cosmos-ai-world-models>. Дата доступа: 24.02.2026.
2. Крыжный, А. ИИ-компании делают шаг в направлении создания суперинтеллекта / А. Крыжный // Экономическая правда. – 29.09.2025. [Электронный ресурс]. Режим доступа: <https://pravda.com.ua/rus/tehnologiji/ii-kompanii-delayut-stavku-na-mirovye-modeli-812197/>. Дата доступа: 26.02.2026.
3. Crosley B. World Models Race 2026: How LeCun, DeepMind, and World Labs Are Redefining the Path to AGI / B. Crosley // Introl Blog. – 02.01.2026. [Электронный ресурс]. Режим доступа: <https://introl.com/blog/world-models-race-agi-2026>. Дата доступа: 26.02.2026.
4. World Models in Artificial Intelligence: Sensing, Learning, and Reasoning Like a Child / Javier Del Ser, Jesus L. Lobo, Heimo Müller, Andreas Holzinger // arXiv. – 2025. [Электронный ресурс]. Режим доступа: <https://arxiv.org/abs/2503.15168>. Дата доступа: 26.02.2026.
5. 世界模型驱动：具身智能告别「盲动」时代。= Мировая модель: воплощённый интеллект прощается с эпохой «слепого движения» // 36kr.com. – 2026. [Электронный ресурс]. Режим доступа: <https://36kr.com/p/3678363993252488>. Дата доступа: 24.02.2026.
6. Competitions: AgiBot World Challenge // IEEE ICRA 2026. – 10.11.2025. [Электронный ресурс]. Режим доступа: <https://2026.ieee-icra.org/program/competitions/#agibot>. Дата доступа: 26.02.2026.
7. Digital Twins and World Models: Bridging Next-Generation AI, Wireless, and Robotics Systems // IEEE ComSoc. – 05.02.2026. [Электронный ресурс]. Режим доступа: <https://r2.ieee.org/event/ieee-comsoc-2026-lecture-series/>. Дата доступа: 26.02.2026.

UDC 004.8:004.94:004.942:530.1

TEACHING ARTIFICIAL INTELLIGENCE THE PHYSICS OF THE WORLD USING WORLD MODELS

Minchanka M.D.

Belarusian State University of Informatics and Radio electronics, Minsk, Republic of Belarus

Maramygina T.A. – senior lecturer at the department of ECG

Annotation. The article examines the qualitative leap in the development of artificial intelligence that occurred in 2025-2026. This period is characterized by a transition from language models capable of “speaking” to World Models that understand the principles of physics.

Keywords: artificial intelligence, World Models, LLM, embodied intelligence, physical AI, paradigm shift