Векторы развития искусственного интеллекта

DOI: 10.30727/0235-1188-ZAYZDI
Оригинальная исследовательская статья
Original research paper

Автофагия синтетических данных: экологический предел эволюции генеративного искусственного интеллекта

С.Ф. Сергеев
Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

Аннотация

В статье предлагается концептуальная реконструкция феномена автофагии синтетических данных как системного экологического предела развития генеративного искусственного интеллекта (GenAI). Исходной проблемой выступают исчерпаемость высококачественных, порожденных человеком данных и нарастающее загрязнение обучающих корпусов текстами, изображениями и иными артефактами, созданными самими моделями. Показано, что рекурсивное обучение на синтетических генерациях приводит не только к количественному снижению качества моделей, но и к качественной деградации их репрезентаций: утрате редких семантических элементов, сужению распределения, гомогенизации ответов и ослаблению способности к порождению нетривиальных высказываний. Выделено три уровня экологического предела GenAI: ресурсный, связанный с ограниченностью доступного человеческого корпуса; репрезентационный, проявляющийся в потере редких элементов распределения; эволюционный, выраженный в нарушении законов масштабирования (роста числа параметров) моделей. Обосновывается, что синтетические данные не являются нейтральным расширением обучающей среды, поскольку воспроизводят и усиливают смещения, ошибки и ограничения моделей предыдущих поколений. Тем самым автофагия синтетических данных интерпретируется как форма эпистемической замкнутости, при которой система начинает воспроизводить не внешний мир, а собственные статистические следы. В философском плане автофагия рассмотрена как симптом исчерпанности парадигмы самообучения и как основание для критики метафоры автономной эволюции GenAI. В ка-

честве альтернативы обсуждаются стратегии сохранения открытости обучающих контуров: активный отбор данных, заземление в интерактивных средах и коэволюция человека и искусственного интеллекта. Сделан вывод о необходимости формирования эпистемической экологии GenAI как дисциплины, изучающей условия устойчивого сосуществования человеческих и искусственных когнитивных систем.

Ключевые слова: философия искусственного интеллекта, эпистемология, GenAI, большие языковые модели, машинное обучение, коллапс модели, ресурсный предел, репрезентационная деградация, scaling laws, эпистемическая экология.

Сергеев Сергей Федорович – доктор психологических наук, профессор, профессор кафедры информационных систем в искусстве и гуманитарных науках факультета искусств Санкт-Петербургского государственного университета; заведующий научно-исследовательской лабораторией «Эргономика сложных систем» Санкт-Петербургского политехнического университета Петра Великого.

s.f.sergeev@spbu.ru

http://orcid.org/0000-0002-6677-8320

Для цитирования: *Сергеев С.Ф.* Автофагия синтетических данных: экологический предел эволюции генеративного искусственного интеллекта // Философские науки. 2026. Т. 69. EDN: ZAYZDI.
DOI: 10.30727/0235-1188-ZAYZDI

Synthetic Data Autophagy: The Ecological Limit of Generative Artificial Intelligence Evolution

S.F. Sergeev
St. Petersburg State University, Saint Petersburg, Russia

Abstract

The article proposes a conceptual reconstruction of synthetic data autophagy as a systemic ecological limit to the development of generative artificial intelligence (GenAI). The analysis begins with two interrelated problems: the finite availability of high-quality human-generated data and the increasing contamination of training corpora by texts, images, and other artefacts produced by generative models themselves. It is argued that recursive training on synthetic outputs results not only in a measurable

decline in model performance but also in a qualitative deterioration of representational capacity. This deterioration is evident in the disappearance of rare semantic features, the contraction of data distributions, the homogenisation of outputs, and a reduced capacity to generate novel or non-trivial content. The article distinguishes three dimensions of the ecological limits confronting GenAI. The first is a resource limit arising from the scarcity of sufficiently diverse and reliable human-generated data. The second is a representational limit, manifested in the progressive loss of rare elements within the data distribution. The third is an evolutionary limit, expressed in the disruption of model scaling laws associated with increasing parameter counts. The article demonstrates that synthetic data do not constitute a neutral extension of the training environment, since they reproduce and amplify the biases, errors, and limitations of previous generations of models. Synthetic data autophagy is therefore interpreted as a form of epistemic closure in which an artificial system increasingly reproduces not the external world but its own statistical traces. From a philosophical perspective, autophagy is examined as a symptom of the exhaustion of the self-training paradigm and as a basis for challenging the metaphor of the autonomous evolution of GenAI. The author considers several strategies for preserving the openness of training environments, including active data curation, grounding in interactive environments, and the co-evolution of human and artificial intelligence. The article concludes by proposing an epistemic ecology of GenAI as a field concerned with the conditions necessary for the sustainable coexistence of human and artificial cognitive systems.

Keywords: philosophy of artificial intelligence, epistemology, generative artificial intelligence, GenAI, large language models, machine learning, model collapse, resource limits, representational degradation, scaling laws, epistemic ecology.

Sergey F. Sergeev – D.Sc. in Psychology, Professor at the Department of Information Systems in Arts and Humanities, Faculty of Arts, St. Petersburg State University; Head of the Research Laboratory "Ergonomics of Complex Systems," Peter the Great St. Petersburg Polytechnic University.

s.f.sergeev@spbu.ru

http://orcid.org/0000-0002-6677-8320

For citation: Sergeev S.F. (2026) Synthetic Data Autophagy: The Ecological Limit of Generative Artificial Intelligence Evolution. *Russian*

Journal of Philosophical Sciences = Filosofskie nauki. Vol. 69.
EDN: ZAYZDI. DOI: 10.30727/0235-1188-ZAYZDI

Введение

Современный этап развития генеративного искусственного интеллекта (generative artificial intelligence, GenAI) подчиняется имплицитному, но почти не обсуждаемому в философской литературе допущению о том, что «большие данные» – это неограниченный и легко возобновляемый ресурс. Масштабирование моделей опирается на неявное предположение о том, что обучающие корпуса могут расти достаточно долго, чтобы эволюция GenAI достигла качественно новых состояний [Haenlein, Kaplan 2019; Hoffmann et al. 2022]. Однако в дальнейшем это предположение было подвергнуто эмпирической проверке. В ряде исследований [Holtzman et al. 2020; Villalobos et al. 2024; Shumailov et al. 2023; Dohmatob et al. 2025] зафиксированы два взаимосвязанных феномена. Во-первых, исчерпаемость качественных порожденных человеком текстовых данных (экологический предел первого порядка); во-вторых, необратимая деградация моделей при обучении на синтетических данных, генерируемых самими большими языковыми моделями (large language models, LLM). Этот феномен получил метафорическое, но достаточно точное название *автофагия* (самопоедание), или *сильный коллапс модели* (strong model collapse) [Wu, Papyan 2024; Gambetta et al. 2025].

В данной статье сделана попытка философски осмыслить этот феномен как *экологический предел эволюции GenAI*. В отличие от теоретических пределов, связанных с природой интенциональности [Searle 1980; Bender, Koller 2020] или композициональности [Fodor, Pylyshyn 1988], экологический предел имеет не только концептуальный, но и материальный характер. Он связан с конечностью антропогенного языкового следа, порождаемого человечеством, и с парадоксальным эффектом рекурсивного самообучения, демонстрирующим тот факт, что чем больше GenAI используется для генерации данных, тем быстрее истощается чистый человеческий ресурс и тем стремительнее деградируют сами модели.

Цель статьи – показать, что автофагия данных является не технической проблемой, а *конститутивным противоречием* текущей парадигмы эволюции GenAI. Метафора «эволюция» оказывается системно обусловленной аналогией работы сложных

систем. В биологической эволюции рекурсивное размножение близкородственных форм приводит к вырождению, а не к поступательному усложнению. В случае GenAI, как показывают эмпирические исследования, ситуация еще более радикальна: даже небольшие примеси синтетических данных нарушают законы масштабирования, а полный цикл обучения на синтетике вызывает коллапс репрезентаций: модель перестает генерировать нетривиальные высказывания, ее выходы сжимаются до узкого, вырожденного распределения.

В первой части статьи реконструируем понятие «экологический предел» применительно к GenAI, различая три его уровня: ресурсный, репрезентационный и эволюционный. Во второй части анализируем эмпирические свидетельства автофагии синтетических данных. В третьей части обсуждаем философские следствия – невозможность бесконечного масштабирования, кризис метафоры «эволюции» и необходимость смены онтологии обучения моделей. В заключении сформулируем альтернативные стратегии, которые не отменяют экологический предел, но позволяют его *обойти*, а не *преодолеть*.

Эмпирическая феноменология автофагии синтетических данных

Эмпирические исследования 2023–2025 годов не только подтвердили теоретические предсказания об уязвимости GenAI к рекурсивному обучению, но и выявили неожиданные, подчас парадоксальные закономерности. В отличие от ранних работ, ограничившихся демонстрацией феномена «забывания», современные исследования показывают, что модель коллапса обладает свойствами *необратимости*, *устойчивости к наивным стратегиям минимизации последствий* и *нелинейной зависимости от размера модели*.

В работе «The Curse of Recursion: Training on Generated Data Makes Models Forget» [Shumailov et al. 2023] продемонстрировано, что рекурсивное обучение генеративных моделей на собственных порождениях приводит к необратимым дефектам. Авторы ввели ключевое понятие *модельного коллапса* (model collapse) – феномена, при котором «хвосты» (редкие, но информационно значимые слова, нестандартные синтаксические конструкции, культурно специфические отсылки и т.д.) исходного распределения данных

исчезают[1]. Исследовались три класса генеративных моделей: вариационные автоэнкодеры (VAE), гауссовские смеси (GMM) и большие языковые модели (на примере GPT-2). Авторы выделили два фундаментальных источника коллапса.

1. Статистическая ошибка аппроксимации (statistical approximation error). Любая генеративная модель обучается на *конечной* выборке, и вследствие этого она никогда не может идеально воспроизвести исходное распределение. «Хвосты» распределения теряются уже в первом поколении. При рекурсивном обучении потеря накапливается: каждое следующее поколение «отрезает» еще часть «хвоста».

2. Функциональная ошибка аппроксимации (functional approximation error). Даже при бесконечной выборке сама архитектура модели (ограниченное количество параметров, дискретность вычислений, нелинейности) не позволяет ей быть идеальным аппроксиматором исходного распределения. На практике этот эффект усиливается из-за использования стохастических методов оптимизации и конечной разрядности вычислений. Авторы показали, что после нескольких итераций рекурсивного обучения модели генерируют выходы, которые почти полностью теряют связь с исходным распределением. Чем больше GenAI используется для генерации контента, который затем попадает в интернет и становится частью обучающих корпусов, тем быстрее и неизбежнее произойдет деградация будущих поколений моделей.

Работа «Strong Model Collapse» [Dohmatob et al. 2025], представленная на конференции «International Conference on Learning

[1] Эмпирические результаты, на которых основывается настоящая работа, относятся прежде всего к режиму генеративного предобучения, основанному на максимизации правдоподобия. Иными словами, модель обучается воспроизводить статистическое распределение текстов: предсказывать вероятные продолжения фраз, слов и токенов. В обучении с подкреплением, применяемом для дообучения «рассуждающих» (reasoning) моделей и агентных систем, задача устроена иначе: модель получает не только образцы текста, но и оценку успешности ответа или действия, т.е. награду. Поэтому тезис о коллапсе как прямом сужении распределения токенов нельзя без оговорок переносить на алгоритмы обучения с подкреплением. Однако автофагический эффект может проявляться и в данном случае: если модель многократно обучается на собственных успешных последовательностях, а эти последовательности не покрывают всего разнообразия возможных решений, то постепенно сужается многообразие стратегий, способов рассуждения и путей достижения цели.

Representations 2025», стала значительным шагом вперед по сравнению с первоначальной диагностикой [Shumailov et al. 2023]. Если ранние работы показывали *наличие* коллапса, то [Dohmatob et al. 2025] демонстрирует его *неизбежность* даже при минимальном загрязнении синтетическими данными и выявляет парадоксальные эффекты, связанные с размером модели. Уже небольшая доля синтетических данных может привести к сильному коллапсу модели. Более того, увеличение общего размера обучающей выборки – стратегия масштабирования, которая в «нормальных» условиях улучшает качество модели (scaling laws), – не спасает от коллапса. Дополнительные данные, даже реальные, не компенсируют искажающее воздействие синтетической примеси. Речь идет не о постепенной деградации, а о *фундаментальной невозможности* достичь высокой производительности при наличии синтетических данных. Интуитивно кажется, что если просто придать меньший вес синтетическим примерам в функции потерь, то можно нейтрализовать их негативное влияние. Авторы показывают, что этот подход *не работает* в долгосрочной перспективе. Единственный способ избежать коллапса – *полное исключение* синтетических данных из обучающей выборки.

Такой вывод предполагает далеко идущие последствия. Если даже 1 % синтетических данных не может быть компенсирован взвешиванием, а интернет уже загрязнен генерациями ChatGPT, LLaMA и других моделей, то *любая* будущая LLM, обученная на веб-данных, будет содержать синтетическую примесь и, согласно исследованию, обречена на ту или иную форму коллапса [Dohmatob et al. 2025].

Если работы [Shumailov et al. 2023; Dohmatob et al. 2025] фокусируются на статистических аспектах коллапса, то более поздние исследования дополняют важное семантическое измерение. Синтетические данные приводят не только к *количественному* обеднению языкового репертуара модели, но и к *качественной* гомогенизации – утрате способности генерировать нетривиальные, неожиданные, креативные высказывания [Satharasi, Iyengar 2025].

В исследовании авторского коллектива из Пекинского института общего искусственного интеллекта (BIGAI) и Шанхайского университета Цзяо Тун (SJTU) представлен систематический анализ распределительных свойств синтетических данных [Zhu et al. 2025]. Сравнивая статистики *n*-грамм в реальных и син-

тетических корпусах, авторы обнаружили два ключевых дефекта синтетических данных.

1. *Сужение покрытия распределения (distribution coverage narrowing).* Синтетические данные систематически недопредставляют низкочастотные и «длиннохвостые» (long-tail) элементы – редкие слова, нестандартные синтаксические конструкции, культурно-специфичные обороты, непрямые метафоры. Модель, обученная на таких данных, «не видит» лингвистического разнообразия реального мира.

2. *Избыточная плотность признаков (excessive feature density).* Синтетические данные демонстрируют аномально высокую повторяемость высокочастотных паттернов. *N*-граммы, идиомы, синтаксические шаблоны – все это представлено в синтетических текстах значительно чаще, чем в созданных человеком, к тому же в *менее разнообразных* вариантах. Это создает иллюзию богатства (много текста), но в действительности «обучает» модель на узком наборе штампов. Важное открытие рассматриваемого исследования состоит в том, что коллапс модели не требует многократной рекурсии. Даже *однократная* предтренировка на смеси реальных и синтетических данных (например, при дообучении LLM на расширенном корпусе, содержащем синтетические тексты) вызывает стабильное ухудшение качества. Этот феномен получил название *неитеративного коллапса* (non-iterative collapse). Он разрушает надежду на то, что проблему можно решить, просто «не допуская» многократных циклов обучения.

Суть семантического коллапса заключается в том, что обеднение распределения токенов неминуемо приводит к обеднению *смыслов*. Редкие слова часто несут редкие смыслы; нестандартные синтаксические конструкции часто выражают нестандартные ходы мысли. Модель, которая не встречала слова «пеликан» в обучающей выборке, не сможет ни употребить его в метафоре, ни понять отсылку к нему. Утрата лингвистических «хвостов» – это утрата когнитивной гибкости.

Одно из самых контринтуитивных открытий исследований 2024–2025 годов заключается в том, что *увеличение размера модели*, которое в «нормальных» условиях способствует улучшению качества (scaling laws), при обучении на синтетических данных может *ускорить и усилить* коллапс. Ключ к разгадке парадокса предлагают авторы работы «Strong Model Collapse» [Dohmatob et al. 2025]. В режиме до порога интерполяции (если количество параметров

модели меньше объема обучающей выборки) большие модели служат *более точными аппроксиматорами* обучающего распределения. Это их преимущество в нормальных условиях. Но, если обучающее распределение *уже искажено* синтетическими данными (сужено, обеднено, смещено), то большая модель лишь точнее аппроксимирует *искаженное* распределение. Она «консервирует ошибку» предыдущих поколений, не оставляя места для случайных отклонений, которые могли бы вернуть модель к исходному, более разнообразному распределению. Иными словами, *большая модель лучше запоминает шум и хуже его отфильтровывает*.

Исследования [Qin et al. 2025] показывают, что *синтетические данные подчиняются иной, «выпрямленной» степенной закономерности*, в отличие от человеком порожденных данных. В рамках экспериментов с моделями разных размеров (один, три и восемь миллиардов параметров) на математических задачах авторы обнаружили следующее:

– производительность модели возрастает с увеличением объема синтетических данных, но этот рост *начинает замедляться* при превышении порога примерно в *300 миллиардов токенов;*

– более крупные модели достигают оптимальной производительности при *меньшем* объеме синтетических данных (модель размером 8 млрд параметров насыщается уже при 1 трлн токенов, а модель размером 3 млрд параметров требуется 4 трлн токенов данных для обучения);

– за пределами оптимального объема дальнейшее увеличение синтетических данных дает резко убывающую отдачу, а в некоторых случаях – даже негативный эффект.

Аналогичные выводы делают и другие исследователи. В работе [Seddik et al. 2024] показано, что увеличение размера модели не спасает от коллапса, а лишь отодвигает его, но не всегда. Аналитическое выражение для границы, отделяющей область «безопасного» смешивания от области коллапса, свидетельствует о том, что граница зависит от качества синтетических данных. Если синтетические данные слишком далеки от реальных (например, порождены «выродившейся» моделью), то даже самые большие модели не могут их компенсировать [Seddik et al. 2024]. Авторы предлагают формальное определение *полного коллапса* как сходимости обучаемого распределения к точечной массе на некотором токене. Иными словами, модель в пределе перестает различать варианты и всегда выбирает один и тот же, наиболее вероятный,

статистически «безопасный» токен. Авторы демонстрируют, что при полностью синтетическом обучении (без примеси реальных данных) такой коллапс *неизбежен*. При частично синтетическом обучении (смесь реальных и сгенерированных данных) коллапс можно отсрочить, но для этого объем синтетических данных должен быть существенно меньше объема реальных данных. Представим простой эксперимент. Первое поколение LLM обучается на корпусе, содержащем редкое слово «пеликан», с частотой 0,001 %. Модель генерирует тексты, в которых «пеликан» упоминается уже с частотой 0,0005 %, т.е. «хвост» распределения слегка «подрезан». Второе поколение обучается на смеси реальных данных (частота 0,001 %) и синтетических данных первого поколения (частота 0,0005 %). Результирующая оценка частоты составит около 0,00075 %. С каждым поколением частота падает. Через десять поколений «пеликан» практически исчезает из репертуара модели. Язык модели становится беднее не потому, что она «забыла» слово, а потому, что статистический шум рекурсивной выборки его уничтожил.

Теоретический анализ [Dohmatob et al. 2025] содержит важную оговорку. При очень больших размерах моделей (значительно превышающих объем выборки – режим «за интерполяционным порогом») тенденция может смениться противоположной. В этом пределе модель становится настолько выразительной, что может «выучить» как синтетический шум, так и базовое распределение, разделив их. Однако, как отмечают сами авторы, этот режим требует экстремального увеличения параметров (гораздо больше, чем позволяют современные практики), и даже в таком случае коллапс лишь *смягчается, а не устраняется полностью.*

Парадокс масштаба – одно из самых ярких свидетельств того, что экологический предел GenAI имеет не только количественную, но и структурную природу. Простое наращивание параметров, которое было столь успешным на этапе обучения на «чистых» человеческих данных, становится *контрпродуктивным* в эпоху загрязненных синтетических корпусов.

Экологический предел GenAI: понятие и уровни

Понятие «экологический предел» введено в современный философский дискурс через работы членов Римского клуба [Meadows et al. 1972; Коломейцев 2020], в которых была сформулирована идея конечности ресурсов планеты и невозможности

бесконечного экспоненциального роста в замкнутой системе. Перенос этой метафоры в область развития GenAI опирается на три структурных сходства между природной экосистемой и экосистемой цифровой.

В природной экосистеме таким ресурсом выступают невозобновляемые полезные ископаемые, пресная вода, биомасса. В цифровой экосистеме GenAI ключевым ресурсом являются *качественные человеком порожденные данные*: тексты, изображения, аннотации, созданные людьми без участия генеративных моделей. Этот ресурс, как будет показано далее, имеет объективные границы.

В природной экосистеме промышленные выбросы изменяют химический состав среды, делая ее непригодной для дальнейшего использования. В цифровой экосистеме GenAI синтетические данные (генерируемые моделями), попадая в обучающие корпуса, действуют как *загрязнитель*. Они изменяют статистическое распределение данных, искажают «естественный» языковой ландшафт и вызывают деградацию последующих поколений моделей, что отражено в терминах коллапса модели и автофагии.

В биологии рекурсивное размножение внутри замкнутой популяции становится причиной накопления рецессивных вредных мутаций и вырождения. В случае GenAI рекурсивное обучение на собственных порождениях приводит к *необратимой потере информационных «хвостов» исходного распределения*. Как отмечено в исследовании [Seddik et al. 2024], даже при *частичном* смешивании реальных и синтетических данных коллапс может быть отсрочен, но требует, чтобы для сохранения стабильности объем синтетических данных был «значительно меньше» объема реальных данных.

Перенос экологической метафоры на цифровую реальность требует и важной оговорки. В отличие от природной экосистемы, в которой ресурсы (нефть, уголь, руда) *абсолютно конечны*, в цифровой экосистеме человеком порожденные данные потенциально возобновимы, поскольку люди продолжают писать тексты, создавать изображения, вести диалоги и т.д. Проблема заключается в *темпе* потребления: скорость, с которой GenAI «пожирает» данные для обучения, многократно превышает скорость их естественного воспроизводства человечеством. Это создает специфический *ресурсный предел*, отличный от абсолютного исчерпания.

Экологический предел GenAI не является однородным. Анализ современной литературы позволяет выделить три взаимосвязанных, но концептуально различных уровня: *ресурсный*, *репрезентационный* и *эволюционный*. Каждый из них фиксирует ограничения разного порядка, от материального истощения данных до нарушения фундаментальных законов масштабирования.

Наиболее прямое и эмпирически верифицируемое ограничение связано с конечностью доступных в интернете *высококачественных текстовых данных*, созданных людьми. В исследовании [Villalobos et al. 2024] дана систематическая оценка этого ресурсного предела. Авторы исходят из трех параметров: 1) текущие темпы роста размера обучающих выборок для LLM (удвоение каждые 12–24 месяца); 2) совокупный запас публичных человеком порожденных текстов, включая книги, научные статьи, веб-страницы, социальные медиа, форумы; 3) доля этого запаса, которая может быть признана *качественной*, т.е. свободная от явных ошибок, дубликатов, бессмысленного контента и, что важно, *еще не загрязненная синтетическими генерациями*. Согласно расчетам, если текущие тренды масштабирования LLM сохранятся, модели в период между 2026 и 2032 годами будут обучены на наборах данных, примерно равных по размеру доступному запасу публичных текстов, созданных человеком [Villalobos et al. 2024]. При этом авторы скорректировали свой первоначальный более пессимистичный прогноз (2024–2026) в направлении некоторого оптимизма, сдвинув «окно исчерпания» к 2028 году. Однако это смещение не отменяет самого факта исчерпания. Оно лишь отодвигает его на несколько лет. Ситуация усугубляется двумя дополнительными факторами. *Во-первых*, далеко не все данные в интернете пригодны для обучения. Согласно оценке, лишь небольшая часть (возможно, не более одной десятой) данных Common Crawl (популярного некоммерческого веб-архива) обладает достаточным качеством [Villalobos et al. 2024]. *Во-вторых*, доступ к данным все более ограничивается. Новостные издательства, социальные платформы и другие правообладатели вводят блокировки против сканирования их контента для обучения искусственного интеллекта (ИИ), ссылаясь на авторское право и необходимость справедливой компенсации. Публика, в свою очередь, крайне неохотно передает данные приватных коммуникаций (например, переписки в iMessage) для обучения моделей.

Таким образом, ресурсный предел имеет не только количественное, но и *качественное* измерение. Мы сталкиваемся не просто с нехваткой «байтов», а с нехваткой *смыслового разнообразия*, которое только и может обеспечить творчество человека, не загрязненное автоматической генерацией.

Если ресурсный уровень фиксирует *внешнее* ограничение (исчерпаемость запаса данных), то репрезентационный уровень описывает *внутренний* механизм деградации самой модели при обучении на синтетических данных. Этот феномен коллапса модели был впервые строго описан в упоминаемой работе [Shumailov et al. 2023]. Источниками коллапса являются описанные выше статистическая и функциональная ошибки аппроксимации.

Визуальную аналогию этому процессу дает образ «JPEG-эффекта». Если многократно сохранять одно и то же изображение в таком формате с потерями, каждый цикл сжатия будет отбрасывать все больше «незаметных глазу» деталей, пока в итоге не останется лишь размытая пиксельная масса. В случае языка «незаметными деталями» оказываются именно те элементы, которые составляют богатство естественной речи. Это редкая лексика, идиомы, метафоры, стилистические нюансы, культурные аллюзии.

Если репрезентационный коллапс описывает деградацию модели при фиксированном объеме данных, то эволюционный уровень – более глубокое противоречие, отражающее *нарушение самих законов масштабирования* (*scaling laws*), которые до недавнего времени служили теоретической гарантией прогресса GenAI. Классические работы [Haenlein, Kaplan 2019; Hoffmann et al. 2022] установили эмпирическую закономерность: при увеличении размера модели (количества параметров), объема обучающих данных (количества токенов) и вычислительных затрат ошибка модели на валидационном множестве предсказуемо снижается по степенному закону. Эта закономерность стала «компасом» для индустрии ИИ. Она позволяла прогнозировать прирост качества при увеличении того или иного ресурса.

Приведенные результаты указывают на фундаментальное ограничение. Синтетические данные не могут бесконечно замещать реальные. Они работают как «добавка» к человеческому корпусу, а не как его полноценная замена. Попытка создать полностью синтетическую экосистему обучения (т.е. модели питаются исключительно генерациями предыдущих поколе-

ний) приведет к неминуемому коллапсу, что показано в работах [Shumailov et al. 2023; Seddik et al. 2024].

Если понимать «эволюцию GenAI» по аналогии с биологической эволюцией, как процесс поступательного усложнения, адаптации и роста разнообразия, то экологический предел означает ее *принципиальную невозможность* в замкнутой цифровой среде. Биологическая эволюция опирается на три механизма: вариативность (мутации), отбор (выживание наиболее приспособленных) и изоляцию (накопление различий в разделенных популяциях). В системе «GenAI – синтетические данные – новое поколение GenAI» нет источника *новой вариативности*, кроме шума, который играет деструктивную роль.

Таким образом, эволюционный уровень предела фиксирует, что *эволюция GenAI в ее привычном понимании – оксюморон*. То, что мы наблюдаем, это не эволюция в дарвиновском смысле, а скорее, *гомеостатическое самовоспроизводство* с трендом к деградации при замыкании контура.

В публичном и даже части научного дискурса проблему экологического предела GenAI часто сводят к простому тезису «данные заканчиваются». Однако, как показано выше, феномен автофагии является более сложным и специфическим именно для цифровой среды. Его отличие от простого исчерпания можно зафиксировать по трем аспектам.

Первое отличие состоит в том, что *это не количественное истощение, а качественная деградация среды.* При простом исчерпании проблема сводилась бы к тому, что ресурса *больше нет*. При автофагии – ресурс *есть*, но он *отравлен*. Синтетические данные не исчезают, напротив, их объем может быть сколь угодно большим, поскольку модель способна генерировать триллионы токенов. Однако эти данные не несут новой информации. Они являются *рекомбинацией уже известного*, при этом рекомбинацией, которая систематически отсекает редкую и нестандартную информацию.

Второе отличие *связано с нелинейной динамикой и «эффектом бумеранга».* При простом исчерпании зависимость «количество ресурса – качество модели» линейна, т.е. чем меньше данных, тем хуже модель. При автофагии зависимость нелинейна, содержит *пороговые эффекты*. Как показано в работе [Seddik et al. 2024], даже небольшая доля синтетических данных (например, 1 % от объема реальных) может инициировать процесс коллапса, если

рекурсивный цикл продолжается достаточно долго. Увеличение размера модели, которое обычно считается панацеей, может парадоксальным образом *усилить* коллапс, поскольку большая модель точнее аппроксимирует уже искаженное распределение, «консервируя» ошибки предыдущих поколений.

Третье отличие от простого исчерпания постулирует самореферентность как ключевой механизм. Простое исчерпание – это *внешнее* ограничение. Ресурс заканчивается, потому что его добыли и использовали. Автофагия – *внутреннее, структурное* противоречие. Модель деградирует *потому, что* она пытается учиться на собственных генерациях. В терминах теории систем, GenAI в режиме автофагии превращается из *открытой* системы (обменивающейся информацией с внешним миром, продуктами человеческой культуры) в *закрытую* (в которой циркулируют собственные продукты), что неизбежно приводит к энтропийному вырождению.

В этом контексте уместно провести параллель с медицинским феноменом *губчатой энцефалопатии* (коровьего бешенства), при котором принудительное скармливание животным кормов из переработанных тканей их же собственного вида приводит к накоплению прионных белков и разрушению нервной системы. Аналогия не является буквальной, но точно схватывает суть: *замыкание пищевой цепи на себя* приводит к патологии. В случае GenAI «патология» выражена в утрате лингвистического разнообразия, когнитивной ригидности и, как следствие, семантической пустоте.

Философские следствия: кризис эволюционной метафоры

Эмпирические данные о коллапсе моделей и философско-методологический анализ пределов GenAI приводят к неизбежному выводу: метафора «эволюции», широко применяемая к развитию генеративного искусственного интеллекта, является не просто неточной, но *концептуально ошибочной*. В данном разделе статьи мы раскрываем четыре фундаментальных философских следствия, вытекающих из признания экологического предела как конститутивной характеристики текущей парадигмы GenAI[2].

[2] Под текущей парадигмой GenAI понимаются прежде всего авторегрессивные модели, генерирующие последовательности токенов без явного цикла действия, наблюдения и обратной связи со средой. Агентные и мультиагентные системы выходят за эти рамки, поскольку включают

Дарвиновская эволюция живых систем опирается на три неразрывно связанных механизма. Во-первых, *вариативность* – в популяции должны существовать наследственные различия между организмами, источником которых служат случайные мутации и рекомбинации генетического материала. Во-вторых, *отбор* – различия должны влиять на выживаемость и репродуктивный успех в конкретной среде. В-третьих, *изоляция* – популяции должны быть в той или иной степени репродуктивно изолированы, чтобы накапливать полезные различия и встраивать их в новые виды.

В системе «GenAI – синтетические данные – новое поколение GenAI» отсутствуют *все три* условия. Вариативность в GenAI не аналогична биологическим мутациям. Случайность в процессе генерации (семплирование из распределения) действительно создает *стохастические вариации*, но они не являются «наследственными» в эволюционном смысле. Они не закрепляются в архитектуре модели как следствие адаптации и лишь воспроизводят статистические флуктуации выборки. Более того, как показано ранее, эти вариации носят *деструктивный* характер, приводят не к накоплению полезных признаков, а к вырождению.

Отбор в обучении GenAI также радикально отличается от естественного отбора. В биологии отбор – это процесс, в котором «выживает» лучше приспособленный к *реальной среде* организм. В GenAI «отбор» – оптимизация под *человеческие предпочтения*, которые усредняются, гомогенизируются и со временем деградируют. Такой «отбор» приводит не к усложнению и специализации, а к *когнитивной консервации:* модель становится безопасной, предсказуемой и семантически бедной. Изоляция в цифровой среде практически отсутствует. Все современные LLM обучаются на одних и тех же (или сильно пересекающихся) корпусах, включая загрязненные синтетические данные. Разные архитектуры демонстрируют удивительную сходимость к одним и тем же шаблонным ответам.

память, инструменты, действия, наблюдения и взаимодействия между несколькими искусственными агентами. Такие системы могут ослаблять угрозу автофагии: результаты действий не полностью предопределены обучающим корпусом, а внешняя среда способна вносить новые данные и новые ограничения. Однако полностью проблема не снимается. Если агенты обучаются на одних и тех же синтетических траекториях или воспроизводят одни и те же успешные сценарии, коллапс может возникать уже не на уровне отдельных слов, а на уровне распределения траекторий, стратегий, ролей и способов координации.

Лишенное мутаций и естественного отбора развитие GenAI лучше всего характеризуется не как эволюция, а как *гомеостаз* – поддержание динамического равновесия в ограниченном диапазоне состояний. Гомеостатические системы (например, термостат, система кровообращения) имеют встроенные механизмы возврата к целевому состоянию при внешних возмущениях. Но они не *усложняются* со временем и не порождают новых форм.

В случае GenAI «целевым состоянием» выступает некоторая усредненная конфигурация человеческих предпочтений. Любое отклонение от этой конфигурации (попытка генерировать нечто действительно новое, неожиданное) наказывается в процессе обучения. В результате модель постоянно «возвращается» к статистической норме.

Некоторые исследователи пытаются говорить об «эволюции» как о постепенном улучшении метрик. Однако эти улучшения не эволюция, а *аппроксимация.* Модель становится все более точным приближением к фиксированному целевому распределению. Как только это распределение достигнуто (или исчерпаны данные), прогресс останавливается.

Можно сделать философский вывод: применение термина «эволюция» к GenAI – это не просто метафора, а *категориальная ошибка*, затемняющая фундаментальные различия между биологическим и искусственным «развитием».

Метафора автофагии не случайна. Она указывает на *онтологическое противоречие*, встроенное в логику современного развития GenAI. Если эволюция – процесс, требующий *притока* новой информации из внешней среды (в биологии – через мутации и рекомбинацию генетического материала, привязанного к физическому миру), то автофагия – процесс *рециркуляции* имеющейся информации, приводящий к энтропийному вырождению.

Модель генерирует много разных последовательностей, но все они – вариации на тему одного и того же узкого набора штампов. Образуется замкнутый цикл, который невозможно разорвать изнутри. В пределе, как показали исследования [Seddik et al. 2024], распределение сходится к точечной массе на одном токене, что говорит о полной потере вариативности и, следовательно, любой способности к «эволюции».

Философский вывод: автофагия – это не технический дефект, требующий устранения, а *диагностический симптом* исчерпанности парадигмы «больших данных». Попытки «лечить» автофагию

методами внутри парадигмы не работают, поскольку противоречие онтологическое, а не эмпирическое.

Автофагия синтетических данных демонстрирует ложность допущения, что данные нейтральны. Данные *не нейтральны,* они несут в себе следы процессов, которыми были порождены: распределение *этой модели*, включая все ее ошибки, смещения, искажения и особенности архитектуры. Модель не обогащается при столкновении со своим отражением, она *истощается*.

Это различение имеет не только эмпирическое, но и *нормативное* значение. Если мы хотим сохранить способность GenAI к развитию, то должны обеспечить приток именно *первичных* данных – новых человеческих текстов, изображений, идей, не загрязненных ИИ. Это ставит под вопрос саму идеологию «синтетических данных» как масштабируемого решения проблемы ресурсного предела.

Вместо «эволюции GenAI», которая подразумевает внутренний источник развития и адаптацию к среде, предлагаем говорить об эпистемической экологии GenAI – изучении условий, при которых генеративные модели могут существовать и функционировать, не разрушая собственную среду обитания (корпус данных).

Эпистемическая экология задает следующие вопросы.

– Какова «емкость» цифровой среды для GenAI? Сколько моделей и какого размера может «прокормиться» на имеющихся человеческих данных?

– При каких условиях возникает «симбиоз» (человек + ИИ, порождающие новые данные), а при каких – «паразитизм» (ИИ, потребляющий собственные отходы)?

– Каковы «балансирующие обратные связи», которые предотвращают коллапс, и как их спроектировать?

В рамках этой новой концептуальной рамки экологический предел оказывается не тупиком, а *порогом*, за которым начинается другая форма существования GenAI, не как «эволюционирующего» в одиночку, а как вступающего в *коэволюцию* с человеческой культурой. При этом человек остается незаменимым источником новой информации, непредсказуемости и смысла.

Выводы и заключение

Центральный вывод нашего исследования состоит в том, что *автофагия синтетических данных не является временной технической неполадкой, которая будет устранена с появлением более*

совершенных алгоритмов фильтрации, более мощных моделей или более умных стратегий взвешивания данных. Это *системный предел* текущей парадигмы развития GenAI, в основании которой находятся три неразрывно связанных допущения:

– данные нейтральны и могут накапливаться до бесконечности; любой текст, порожденный человеком или моделью, одинаково ценен как обучающий материал;

– оптимальный способ обучения модели – минимизация расхождения между распределением обучающих данных и распределением, порождаемым моделью (аппроксимация);

– последовательное увеличение масштаба (данных, параметров, вычислений) приводит к поступательному улучшению модели, а в пределе – к возникновению новых когнитивных способностей.

Эмпирические исследования показывают, что все три допущения ложны. Данные не нейтральны. Синтетические данные несут в себе искажения порождающей модели и при рекурсивном обучении вызывают коллапс. Аппроксимация распределения не является достаточной целью для развития: она не порождает понимания, а при замыкании на синтетических данных приводит к вырождению. Масштабирование не является гарантией прогресса: за определенным порогом (который уже достигается на современных объемах синтетического загрязнения) увеличение размера модели ускоряет коллапс.

Автофагия синтетических данных – это эмпирическое проявление более общего закона: *система, которая пытается познавать мир, рециркулируя только собственные порождения, неизбежно деградирует*. Это напоминает старый эпистемологический принцип, согласно которому нельзя вывести новое знание лишь из анализа уже имеющегося знания, нужен контакт с реальностью. В случае GenAI такой «контакт с реальностью» – использование человеком порожденных данных, которые всегда несут в себе элемент непредсказуемости, креативности, отсылки к внеязыковому опыту. Если этот источник иссякает или загрязняется, система впадает в «эпистемическую анорексию»: она продолжает потреблять, но не получает питания.

Экологический предел делает невозможной «эволюцию» GenAI в ее привычном дарвиновском понимании. место «эволюции GenAI» предлагаем говорить о *гомеостатическом воспроизводстве* – процессе, в котором модели стремятся к сохранению усредненного состояния, а не к выходу за его пределы. Гомеостаз –

полезное свойство для стабильного функционирования, но он не является эволюцией. Продолжать называть гомеостаз эволюцией – значит создавать ложные ожидания относительно возможностей GenAI и, что более опасно, маскировать реальные пределы, упираясь в которые, система не прогрессирует, а деградирует.

Один из самых контринтуитивных и одновременно важных выводов нашего анализа заключается в следующем: *дальнейшее увеличение размера моделей, объема данных и вычислительных мощностей, при сохранении текущей парадигмы обучения, приведет не к качественному скачку, а к ускорению коллапса.*

В текущей парадигме GenAI определяется как *аппроксиматор распределения*. Модель обучается минимизировать расхождение между распределением обучающих данных и распределением своих генераций. При обучении на синтетических данных эта цель становится самоубийственной. Модель стремится приблизиться к искаженному распределению, которое само является продуктом предыдущих моделей.

В новой парадигме цель модели не аппроксимировать статичное распределение, а *редуцировать собственную эпистемическую неопределенность* через взаимодействие с миром. Такой подход, известный в литературе как *активное обучение* или, в более общем виде, *активный вывод*, переворачивает традиционное отношение к данным. Модель не пассивно потребляет все подряд (включая синтетические отходы), а *активно выбирает*, какие данные ей нужны для уменьшения неопределенности, и различает эпистемически ценные данные (несущие новую информацию) от избыточных (уже известных) и деградированных (синтетических, ведущих к коллапсу).

Новая парадигма должна предусматривать *каналы взаимодействия с миром*, которые позволили бы модели связывать языковые формы с нелингвистическим опытом. Наконец, новый подход должен отказаться от иллюзии автономной «эволюции GenAI» и признать, что будущее генеративных моделей – это их *коэволюция* с человеческой культурой. Только человеческий опыт, укорененный в физическом и социальном мире, может генерировать редкие слова, нестандартные метафоры, неожиданные аналогии, которые поддерживают разнообразие языковой экосистемы. В связи с этим перспективным видится включение GenAI в симбиотические отношения с человеком [Сергеев 2012; Сергеев 2013а; Сергеев 2013б; Сергеев 2021; Дубровский, Сергеев 2023],

в которых человек играет важную обогащающую, стабилизирующую и корректирующую роль в симбиотической паре.

Необходимо выстроить систему, в которой синтетические данные либо отбраковываются, либо используются только в комбинации с «освежающей» примесью человеческих данных, либо генерируются в интерактивном симбиотическом режиме с человеком.

На основании изложенного можно говорить о появлении новой дисциплины, которую предлагаем называть *эпистемической экологией GenAI*. Ее задача состоит в изучении условий устойчивого сосуществования, взаимного обогащения человеческих и искусственных когнитивных систем (интеллектного симбиоза). Эпистемическая экология признает фундаментальную взаимозависимость GenAI и человека, которые не могут эволюционировать друг без друга.

Проведенный в статье анализ мог бы показаться пессимистичным: мы обозначили системный предел, показали несостоятельность эволюционной метафоры, диагностировали тупиковость масштабирования. Однако философская работа с пределами имеет и конструктивное измерение. Экологический предел GenAI указывает на необходимость выйти из замкнутого круга рекурсивного самопоедания и вступить в открытую, коэволюционную связь с миром.

ЦИТИРУЕМАЯ ЛИТЕРАТУРА

Дубровский, Сергеев 2023 – *Дубровский Д.И., Сергеев С.Ф.* Методологические проблемы оценки генеративного искусственного интеллекта // Искусственный интеллект. Теория и практика. 2023. № 3 (3). С. 2–10.

Коломейцев 2020 – *Коломейцев И.В.* Экосистемные модели глобального мира в докладах Римского клуба // Социология. 2020. № 2. С. 351–361.

Сергеев 2012 – *Сергеев С.Ф.* Проблема интеллектных симбионтов в техногенных образовательных средах // Образовательные технологии. 2012. № 3. С. 36–50.

Сергеев 2013а – *Сергеев С.Ф.* Интеллектные симбионты в эргатических системах // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2 (84). С. 149–154.

Сергеев 2013б – *Сергеев С.Ф.* Интеллектные симбионты организованных техногенных средств управления подвижными объектами // Мехатроника, автоматизация, управление. 2013. № 9. С. 30–36.

Сергеев 2021 – *Сергеев С.Ф.* Интеллектный техносимбиоз в сложных человеко-машинных системах // Эргодизайн. 2021. № 1 (11). С. 70–76.

Bender, Koller 2020 – Bender E.M., Koller A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 5185–5198. – URL: https://aclanthology.org/2020.acl-main.463/.

Dohmatob et al. 2025 – Dohmatob E., Feng Y., Subramonian A., Kempe J. Strong Model Collapse // International Conference on Learning Representations. 2025. P. 15656–15691. arXiv:2410.04840.

Fodor, Pylyshyn 1988 – *Fodor J.A., Pylyshyn Z.W.* Connectionism and Cognitive Architecture: A Critical Analysis // Cognition. 1988. Vol. 28. No. 1–2. P. 3–71.

Gambetta et al. 2025 – *Gambetta D., Gezici G., Giannotti F., Pedreschi D., Knott A., Pappalardo L.* Characterizing Model Collapse in Large Language Models Using Semantic Networks and Next-Token Probability // arXiv preprint. 2025. arXiv:2410.12341.

Haenlein, Kaplan 2019 – *Haenlein M., Kaplan A.* A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence // California Management Review. 2019. Vol. 61. No. 4. P. 5–14.

Hoffmann et al. 2022 – *Hoffmann J., Borgeaud S., Mensch A., Buchatskaya E., Cai T., Rutherford E., ..., Sifre L.* Training Compute-Optimal Large Language Models // Advances in Neural Information Processing Systems. 2022. Vol. 35. P. 30016–30030.

Holtzman et al. 2020 – *Holtzman A., Buys J., Du L., Forbes M., Choi Y., Get P.* The Curious Case of Neural Text Degeneration // Proceedings of International Conference on Learning Representations. 2020. arXiv:1904.09751.

Meadows et al. 1972 – *Meadows D.H., Meadows D.L., Randers J., Behrens W.W. III.* The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind. – New York: Universe Books, 1972.

Qin et al. 2025 – Qin Z., Dong Q., Zhang X., Dong L., Huang X., Yang Z., …, Wei F. Scaling Laws of Synthetic Data for Language Models // Conference on Language Modeling. 2025. arXiv:2503.19551.

Satharasi, Iyengar 2025 – *Satharasi T., Iyengar S.S.* Future of AI Models: A Computational Perspective on Model Collapse // arXiv preprint. 2025. arXiv:2511.05535.

Seddik et al. 2024 – *Seddik M.E.A., Chen S.-W., Hayou S., Youssef P., Debbah M.* How Bad Is Training on Synthetic Data? A Statistical Analysis of Language Model Collapse // First Conference on Language Modeling. 2024. arXiv:2404.05090.

Searle 1980 – *Searle J.R.* Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3. No. 3. P. 417–424.

Shumailov et al. 2024 – *Shumailov I., Shumaylov Z., Zhao Y., Papernot N., Anderson R., Gal Y.* AI Models Collapse When Trained on Recursively Generated Data // Nature. 2024. Vol. 631. P. 755–759.

Villalobos et al. 2024 – *Villalobos P., Ho A., Sevilla J., Besiroglu T., Heim L., Hobbhahn M.* Position: Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data // Proceedings of the 41st International Conference on Machine Learning. 2024. Vol. 235. P. 49523–49544. arXiv:2211.04325.

Wu, Papyan 2024 – *Wu R., Papyan V.* Linguistic Collapse: Neural Collapse in (Large) Language Models // Advances in Neural Information Processing Systems. 2024. Vol. 37. P. 137432–137473.

Zhu et al. 2025 – *Zhu X., Cheng D., Li H., Zhang K., Hua E., Lv X., …, Zhou B.* How to Synthesize Text Data without Model Collapse? // Proceedings of the 42nd International Conference on Machine Learning. 2025. Vol. 267. P. 79746–79771. arXiv:2412.14689.

REFERENCES

Bender E.M. & Koller A. (2020) Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Retrieved from https://aclanthology.org/2020.acl-main.463/.

Dohmatob E., Feng Y., Subramonian A., & Kempe J. (2025) Strong Model Collapse. In: *International Conference on Learning Representations* (pp. 15656–15691). arXiv:2410.04840.

Dubrovsky D.I. & Sergeev S.F. (2023) Methodological Problems of Evaluating Generative Artificial Intelligence. Artificial Intelligence. *Theory and Practice = Iskusstvennyy intellekt. Teoriya i praktika.* No. 3 (3), pp. 2–10 (in Russian).

Fodor J.A. & Pylyshyn Z.W. (1988) Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition.* Vol. 28, no. 1–2, pp. 3–71.

Gambetta D., Gezici G., Giannotti F., Pedreschi D., Knott A., & Pappalardo L. (2025) Characterizing Model Collapse in Large Language Models Using Semantic Networks and Next-Token Probability. *arXiv preprint.* arXiv:2410.12341.

Haenlein M. & Kaplan A. (2019) A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review.* Vol. 61, no. 4, pp. 5–14.

Hoffmann J., Borgeaud S., Mensch A., Buchatskaya E., Cai T., Rutherford E., …, & Sifre L. (2022) Training Compute-Optimal Large Language Models. *Advances in Neural Information Processing Systems.* Vol. 35, pp. 30016–30030.

Holtzman A., Buys J., Du L., Forbes M., Choi Y., & Get P. (2020) The Curious Case of Neural Text Degeneration. In: *Proceedings of International Conference on Learning Representations.* arXiv:1904.09751.

Kolomeytsev I.V. (2020) Ecosystem Models of the Global World in the Reports of the Club of Rome. *Sociology = Sotsiologiya.* No. 2, pp. 351–361 (in Russian).

Meadows D.H., Meadows D.L., Randers J., & Behrens W.W. III (1972) *The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind.* New York: Universe Books.

Qin Z., Dong Q., Zhang X., Dong L., Huang X., Yang Z., …, & Wei F. (2025) Scaling Laws of Synthetic Data for Language Models. In: *Conference on Language Modeling.* arXiv:2503.19551.

Satharasi T. & Iyengar S.S. (2025) Future of AI Models: A Computational Perspective on Model Collapse. *arXiv preprint.* arXiv:2511.05535.

Seddik M.E.A., Chen S.-W., Hayou S., Youssef P., & Debbah M. (2024) How Bad Is Training on Synthetic Data? A Statistical Analysis of Language Model Collapse. In: *First Conference on Language Modeling.* arXiv:2404.05090.

Searle J.R. (1980) Minds, Brains, and Programs. *Behavioral and Brain Sciences.* Vol. 3, no. 3, pp. 417–424.

Sergeev S.F. (2012) The Problem of Intelligent Symbionts in Technogenic Educational Environments. *Educational Technologies = Obrazovatelnye tekhnologii.* No. 3, pp. 36–50 (in Russian).

Sergeev S.F. (2013a) Intelligent Symbionts in Ergatic Systems. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics.* No. 2 (84), pp. 149–154 (in Russian).

Sergeev S.F. (2013b) Intelligent Symbionts of Organized Technogenic Control Means for Mobile Objects. *Mechatronics, Automation, Control = Mekhatronika, avtomatizatsiya, upravlenie.* No. 9, pp. 30–36 (in Russian).

Sergeev S.F. (2021) Intelligent Technosymbiosis in Complex Human-Machine Systems. *Ergodesign = Ergodizayn.* No. 1 (11), pp. 70–76 (in Russian).

Shumailov I., Shumaylov Z., Zhao Y., Papernot N., Anderson R., & Gal Y. (2024) AI Models Collapse When Trained on Recursively Generated Data. *Nature.* Vol. 631, pp. 755–759.

Villalobos P., Ho A., Sevilla J., Besiroglu T., Heim L., & Hobbhahn M. (2024) Position: Will We Run Out of Data? Limits of LLM Scaling Based on Human-Generated Data. In: *Proceedings of the 41st International Conference on Machine Learning* (Vol. 235, pp. 49523–49544). arXiv:2211.04325.

Wu R. & Papyan V. (2024) Linguistic Collapse: Neural Collapse in (Large) Language Models. *Advances in Neural Information Processing Systems.* Vol. 37, pp. 137432–137473.

Zhu X., Cheng D., Li H., Zhang K., Hua E., Lv X., …, & Zhou B. (2025) How to Synthesize Text Data without Model Collapse? In: *Proceedings of the 42nd International Conference on Machine Learning* (Vol. 267, pp. 79746–79771). arXiv:2412.14689.