Искусственный интеллект как смертный грех человечества

# Artificial Intelligence as Humanity’s Deadly Sin

Доклад на панельной сессии: Искусственный интеллект и геополитика

Report at the panel session: Artificial intelligence and geopolitics

А.Н. Козырев

A. N. Kozyrev

Искусственный интеллект (далее – ИИ) стал важным фактором геополитики благодаря в основном двум важным обстоятельствам. Во-первых, сейчас как никогда раньше реальна возможность использовать ИИ в военной сфере, в том числе, в качестве оружия. Задачи распознавания шумов, производимых винтами подводных лодок, и другие инженерные по сути задачи военного назначения решались и раньше, но сегодня речь идет о создании роботов, которые будут убивать людей. Во-вторых, стремительно растут вложения в ИИ гражданского назначения, прежде всего, вложения частных средств в развитие коммерческих приложений ИИ.

Artificial intelligence (AI) has become an important factor in geopolitics due to mainly two important circumstances. First, now more than ever before a real opportunity to use AI as a weapon in the military sphere. The problems of detecting the noise produced by the screws of submarines and other engineering tasks in fact for military purposes have been solved before. But today we are talking about creating robots that will kill people. Second, investments in civil applications are growing rapidly, primarily private investments in the development of commercial AI applications.

Тенденции развития ИИ и ожидания в оптимистическом ключе рассмотрены в докладе Top AI Trends To Watch In 2018, подготовленном CD Insights. Вместе с тем, далеко не все тут однозначно. Влияние вложений в ИИ на экономический рост в настоящее время скорее отрицательно, то есть возникает несоответствие между ожиданиями и статистикой. Таков главный вывод отчета Национального бюро экономических исследований, озаглавленного «Искусственный интеллект и современной парадокс производительности: столкновение ожиданий и статистики» (NBER, 2017). Еще более ярко разрыв между реальными достижениями ИИ и ожиданиями и показан в публикации двух авторов (Marcus & Davis, 2018), один из которых психолог, второй – специалист по информационным технологиям.

Trends in the AI development and some optimistic expectations are considered in the report “Top Trends to Watch in 2018” that prepared by CD Insights. However, not everything is clear. The impact of AI investments on economic growth is rather negative at present. There is a mismatch between expectations and statistics. This is the main conclusion of the National Bureau of economic research report entitled "Artificial intelligence and the modern paradox of productivity: meeting expectations and statistics" (NBER, 2017). The gap between the real achievements of AI and expectations is even more pronounced and is shown in the publication of two authors (Marcus & Davis, 2018), one of whom is a psychologist and the second is a specialist in information technology.

Настоящий доклад посвящен другим не менее важным и гораздо более тревожным аспектам развития технологий ИИ. Речь идет о влиянии ИИ на людей и нашу цивилизацию в целом. Обращение к понятию смертного греха в названии доклада – не попытка эпатажа, а всего лишь следование традиции, заложенной Конрадом Лоренцом – величайшим естествоиспытателем и философом 20-века. В названии одной из его книг (Lorenz, 1973) – «Восемь смертных грехов цивилизованного человечества» ассоциируется с библейским текстом о семи смертных грехах. Но речь идет не о грехах отдельного человека, а об опасных тенденциях в политике развитых стран и развития человечества в целом, способных привести человечество к гибели.

This report focuses on another important and much more worrying aspect of AI technology development. We are talking about the impact of AI on humans and our civilization in general. The reference to the concept of deadly sin in the title of the report is not an attempt to outrageous but merely following the tradition laid down by Conrad Lorenz who is the greatest naturalist and Philosopher of the 20th century. The title of his book (Lorenz, 1973) “Eight Civilized Man’s Deadly Sins” is associated with the biblical text about seven deadly sins. But we are not talking about the sins of the individual, we are talking about dangerous trends in the policies of developed countries and the development of humanity at large which can lead humanity to death.

Общее у восьми смертных грехов по Лоренцу то, что все они связаны с нарушением отрицательных обратных связей в природе и обществе как результатом целенаправленных действий людей с благими намерениями. В частности, речь идет об уничтожении отрицательных обратных связей, обеспечивающих гомеостаз в биологической системе, частью которой является человечество, но не только о них. Отрицательные обратные связи обычно воспринимаются как неудобства или даже зло. Обрыв таких связей в каждом отдельном случае легко оправдать, поскольку он вызван либо необходимостью решения каких-то серьезных, в том числе, геополитических проблем, либо соображениями гуманности, либо желанием повысить качество жизни именно в том смысле, как это понимается здесь и сейчас – на момент принятия конкретного решения. Но каждый из таких обрывов обратной связи имеет отдаленные последствия, причем не всегда предсказуемые и, возможно, фатальные для человечества в конечном итоге.

What is united the eight deadly sins from Lorentz's book is that they are all associated with the violation of negative feedbacks in nature and society that are results of the purposeful actions of people with good intentions. We talk about the destruction of negative feedbacks that provide homeostasis in the biological system as so humanity is a part of it. But we are talking not only about such feedbacks. Negative feedbacks are usually perceived as inconveniences or even evil. The breakup of such ties in each case is easy to justify since it is caused either by the need to solve some serious, including geopolitical problems, or by considerations of humanity, or by the desire to improve the quality of life in the sense as it is understood here and now at the time of making a specific decision. But each of these feedback breaks has long-term consequences not always predictable and perhaps fatal to humanity in the end.

Каждому из названных Лоренцем смертных грехов соответствует отдельная глава в его книге. Всего в ней девять глав. Первая глава посвящена гомеостазу, обратным связям и другим общим вопросам, развиваемым в последующих восьми главах применительно к каждому из явно названых восьми грехов. Ниже они перечислены в порядке, выбранном Лоренцем, чтобы потом было удобно ссылаться.

Each of the deadly sins named by the Lorentz corresponds to a separate chapter in his book. It has nine chapters in total. The first chapter is devoted to homeostasis, feedbacks and other general issues which are developed in the following eight chapters in connection with one of the sins. They are listed below in the order chosen by Lorenz so that it is convenient to refer to them later.

Глава 1. Структурные свойства и нарушения функций живых систем.

Глава 2. Перенаселение.

Глава 3. Опустошение жизненного пространства.

Глава 4. Бег наперегонки с самим собой.

Глава 5. Тепловая смерть чувства.

Глава 6. Генетическое вырождение.

Глава 7. Разрыв с традицией.

Глава 8. Индоктринируемость.

Глава 9. Ядерное оружие.

Chapter 1. Structural Properties and Functional Disorders of Living Systems

Chapter 2. Overpopulation.

Chapter 3. Devastation of the Environment

Chapter 4. Man’s Race Against Himself

Chapter 5. Entropy of Feeling

Chapter 6. Genetic Decay

Chapter 7. The Breaking with Tradition.

Chapter 8. Indoctrinability

Chapter 9. Nuclear Weapons

Среди перечисленных явно восьми грехов (главы 2-9) не упомянут ИИ. Не упоминается он и в тексте книги. Однако легко заметить, что ИИ явно имеет отношение, как минимум, к трем упомянутым грехам – бег наперегонки с собой (глава 4), разрыв с традицией (Глава 7) и, как следует из публикации (RAND, 2018), ядерное оружие (глава 9). Более глубокий анализ показывает, что связей гораздо больше. Возможно, хотя и это спорно, нет связей ИИ с перенаселением Земли (Глава 2) и опустошением жизненного пространства (Глава 3). В остальном связь просматривается достаточно ясно. Постоянное общение с ботами не может не сказаться на чувствах и способности чувствовать (Глава 5), избавление от рутинных операций оборачивается деградацией в том, что не является рутиной (Глава 6). Так, постоянное пользование навигатором при вождении автомобиля приводит к неспособности ориентироваться на местности в реальной обстановке. Но самое интригующее – это связь ИИ с тем, что в книге Лоренца получило название Индоктринируемость. В случае с ИИ эта специфическая болезнь современной науки проявляется ярко, как нигде больше, если не считать экономическую науку.

AI is not mentioned clearly among the listed eight sins (chapters 2-9). It is not mentioned in the text of the book too. However, it is easy to see that the AI is clearly related to at least the three mentioned sins. These are man’s race against himself (Chapter 4), the breaking with tradition (Chapter 7) and, as follows from the publication (RAND, 2018), nuclear weapons (Chapter 9). Deeper analysis shows that there are many more connections. Perhaps, although it is debatable, there are no AI connections with the overpopulation of the Earth (Chapter 2) and the devastation of the environment (Chapter 3). In other cases, the relation is quite clear. Constant communication with bots can affect the feelings and the ability to feel (Chapter 5), getting rid of routine operations leads to degradation in what is not a routine (Chapter 6). Thus, the constant use of the navigator while driving a car leads to the inability to orientate on the ground in a real situation. But the most intriguing is the relationship of AI with the fact that in the book of Lorentz called Indoctrinability. In the case of AI, this specific disease of modern science is manifested more vividly than anywhere else.

Отдельным смертным грехом могло бы стать создание искусственного интеллекта, превосходящего по своей мощи человеческий интеллект. Возможные проблемы, возникающие или, точнее, ожидаемые в этой связи активно обсуждаются в некоторых аудиториях и организациях. Например, этому вопросу посвящены публикация (Yudkowsky, 2008) и библиография к ней, а также многие более поздние публикации. Но в основном внимание привлечено к реальным или мнимым успехам в области построения нейросетей и глубокого обучения.

The creation of an artificial intelligence superior in its power to the human intellect could be a separate deadly sin. Possible problems arising or, more precisely, expected in this regard are actively discussed in some audiences and organizations. For example, this question is debated in the publication (Yudkowsky, 2008) and the bibliography to it as well as many later publications. But mostly attention is drawn to real or imaginary successes in the field of neural networks and deep learning.

Также надо учесть, что в далеком уже 1972 году, когда Лоренц писал свою книгу, реальные достижения ИИ были еще слишком скромны, чтобы ставить их разрушительную силу в один ряд с ядерным оружием и генетическим вырождением. Например, робот с ИИ мог сложить башенку из 4-х кубиков, используя «руку» – манипулятор. На 5-й кубик интеллекта уже не хватало. Но в произведениях фантастов еще задолго до того (в 1950-60-х) роботы умели не только служить человеку, но и выходить из-под его контроля, и даже любить, рискуя сжечь в любовном порыве все лампы своего электронного «мозга».

It should also give regard to back in 1972, when Lorenz wrote his book, the real achievements of AI were too modest to put their destructive power on a par with nuclear weapons and genetic decay. For example, a robot with AI could put a turret of 4 cubes using a "hand" type manipulator. For the 5th cube intelligence was not enough. But in the works of science fiction long before (in the 1950s and 60s), robots were able not only to serve man, but also to get out of his control, and even to love risking burning in a love impulse all the lamps of electronic "brain".

Отчасти эти настроения разделяли и сами исследователи. Как вспоминает один из наших исследователей: «Я помню ожидания после рывка 1950–60-х. Казалось, вот-вот — еще 3 года! Еще 5 лет! — и будет создан искусственный интеллект общего назначения, способный себя осознавать, самостоятельно ставить цели, сотрудничать с человеком на равных. А потом стало ясно, что исследователи топчутся на месте, и ничего важного не происходит».

In part these sentiments were shared by the researchers themselves. Remembers one of soviet researchers: "I remember waiting after spurt 1950-60's. it Seemed, here-here another 3 years! 5 more years! We will create an artificial intelligence of general purpose, capable of self-awareness, self-set goals, to cooperate with a person on an equal footing. And then it became clear that the researchers are marking time and nothing important is happening."

Сейчас ситуация с применением ИИ во многих аспектах отличается от той, что была в шестидесятые и начале семидесятых годов 20-го века. Сегодня ИИ—своеобразный технологический агрегатор, вобравший в себя множество разных направлений, от добычи данных до робототехники. Он привлекает максимальный интерес инвесторов и бизнеса. И все же сходство в некоторых фундаментальных проблемах самого ИИ разительно. Тут самое время вспомнить про Индоктринируемость (Глава 8) – одну из болезней современной науки и научного сообщества. В разных странах эта болезнь имела тогда и имеет сегодня свои оттенки, в частности, это касается ИИ.

Now the situation with the AI application and in many aspects is different from that which was in the sixties and early seventies of the 20th century. Today AI is a kind of technological aggregator which has absorbed many different directions from data mining to robotics. It attracts the maximum interest of investors and business. And yet the similarity in some of the fundamental problems of AI itself is striking. It's time to remember about Indoctrinability (Chapter 8) as one of the diseases of modern science and the scientific community. In different countries, this disease was then and now but it has its shades. This disease concern AI researches too.

У нас в 1960-70-х (в СССР) тематика ИИ была окружена особым ореолом, в котором явно присутствовал своего рода декаданс. Среди интеллигенции считалось, что кибернетику у нас «разгромили» по политическим мотивам, снабдив непристойным ярлыком –«Продажная девка империализма». Так ли было на самом деле? Вопрос более, чем спорный, поскольку на связанные с обороной и космосом направления кибернетики тратились очень значительные средства. Велись и вполне мирные исследования. Однако в глазах широкой публики и гуманитарной интеллигенции транслируемая «сарафанным радио» передача про «девку» вполне объясняла причины довольно скромных успехов кибернетики и, прежде всего, ИИ в гражданском секторе, который у всех на виду.

We have in the 1960-70s (in the USSR) the subject AI was surrounded by a special aura which was clearly present a kind of decadence. Among the intelligentsia, it was believed that Cybernetics as science have been "crushed" for political reasons under the obscene label "whore of imperialism". Was that really the case? The question is more than controversial because to Cybernetics for the defense and space-related areas have been spent very significant funds. Quite peaceful researches were carried out too. However, in the eyes of the General public and the humanitarian intelligentsia, the broadcast of the "word of mouth" about the "whore" explained the reasons for the rather modest success of Cybernetics and, above all, AI in the civil sector, which is in plain sight.

Но то же самое было за океаном, только без идеологического оттенка. Кибернетику «громили» представители точных наук, видевшие ее эклектичность и несоответствие достижений обещаниям. В частности, знаменитый доклад математика сэра Джеймса Лайтхилла, подготовленный по заказу парламента в 1973 году (Lighthill, 1973), привел к почти полному демонтажу исследований ИИ в Англии. Примечательно, что доклад осуждался публично, дискуссия по нему транслировалась BBC, а запись сохранена. Ее можно посмотреть на YouTube. Суть доклада, если ее формулировать одной фразой, состояла в том, что не существует такой дисциплины как ИИ. Все решаемые ИИ реальные задачи могут быть решены в других дисциплинах. А в объединяющей их части реальных достижений фактически нет.

But the same situation had a place behind the ocean and without the ideological tint. Representatives of the exact Sciences saw eclecticism of Cybernetics and inconsistency of achievements to promises and they "smashed" this branch of science. The famous report of mathematician sir James Lighthill that commissioned by Parliament in 1973 (Lighthill, 1973) led to an almost complete dismantling of AI studies in England. It is noteworthy that the report was discussed in public, the discussion was broadcast by the BBC, and the record is preserved. You can watch it on YouTube. The essence of the report, if it is formulated in one sentence, was that there is no such discipline as AI. All real problems solved by AI can be solved in other disciplines. There are no real achievements in the uniting part of the actually.

Сейчас происходит нечто подобное, но все строится вокруг денег. Разрыв между реальными достижениями и фантазиями на эту тему и стал общей травмой. Более того, она стала злокачественной, когда на смену писателям-фантастам пришли маркетологи. Только исследования, которые приносят быстрые деньги, имеют шанс для инвестиций. "Много денег вкладывается в системы, способные в режиме реального времени подменять голоса и подражать политикам и просчитывать поведение избирателей". Но это не научная проблема.

Something similar is happening now, but everything is now built around money. The gap between real achievements and fantasies on the topic and became a generic trauma. Moreover, it became malignant when science fiction-writers had been replaced by marketers. Only researches that brings quick money have a chance for investment. "A lot of money is invested in systems capable of real-time replacement of votes and mimics of politicians and to calculate the behavior of voters". But it is not scientific problem.

Сенсации при ближайшем рассмотрении оборачиваются не вполне добросовестной подачей материала. Типичный пример – подача в некоторых источниках информации о победе AlphaZero над Stockfish в матче из 100 партий как сенсации мирового уровня. При попытке объективного рассмотрения выясняется, что «для выравнивания условий» у Stockfish были отключены базы данных по дебютам и эндшпилям при том, что именно они у Stockfish были «фишкой», определявшей силу игры. Вместе с тем AlphaZero использовал принципиально новый тензорный процессор, а Stockfish – стандартный центральный процессор, что давало AlphaZero дополнительное преимущество. Иначе говоря, здесь скорее можно говорить об успехе в разработке новых процессоров, а не интеллекта. В разработке нейросетей успех, несомненно, тоже есть, но он отнюдь не сенсационный. И снова речь отнюдь не об универсальном интеллекте, а о новом подходе к решению конкретной задаче.

On closer examination some sensations turn out not quite good presentation of the material. A typical example is the submission by some sources of information about the victory of AlphaZero over Stockfish in a match of 100 games as a sensation of world level. When you try to do objective consideration, it appears that for the "equalization of conditions" from Stockfish has been disabled databases for openings and endgames. But the playing strength of Stockfish was defined by these databases. At the same time, Stockfish used a standard CPU, but AlphaZero used a fundamentally new tensor processor which gave AlphaZero an additional advantage. In other words, here we can talk about success in the development of new processors, not intelligence. In the development of neural networks success, of course, too, but it is not sensational. Again, it is not about universal intelligence, but about a new approach to solving a specific problem.

Тут при желании можно снова сказать, что тематику ИИ «громят» представители точных наук. Типичный пример – лекция известного физика Фримена Дайсона о человеческом мозге (Dyson, 2014). В частности, он рассказал поучительную историю с докладом его давнего, но к 2014 уже покойного сэра Джеймса Лайтхилла. Еще тогда в 1973 году все работы по ИИ (в широком смысле) четко делились на три категории, обозначенные в докладе A, B и C. Здесь категории A и C имеют четко определенные мотивы: каждая из них имеет четко определенное общее направление своих предполагаемых целей, но эти два направления совершенно разные. В обеих этих категориях в течение двадцати пяти лет (начиная со статьи Тьюринга 1947 года "интеллектуальный механизм" и кончая публикацией доклада 1973) был достигнут определенный прогресс, хотя ожидания часто не оправдывались. разочаровывались. Категорию A составляли чисто прикладные инженерные исследования типа распознавания речи, машинного перевода и некоторые другие практические задачи, решаемые обычно людьми. Категория C — все, что связано с когнитивным, нейроморфным, мозгу-подобным компьютингом. Были отмечены некоторые перспективные исследовательские работы в области нейронауки.

Here at desire it is possible again to say that the theme of AI was "attacked" by the representatives of the exact Sciences. A typical example is a lecture by the famous physicist Freeman Dyson on the human brain (Dyson, 2014). He told a teaching story about the report prepared by his long-standing friend, but by 2014 already late sir James Lighthill. Back then, in 1973, all AI researches (in the broad sense) was clearly divided into three categories, identified in the report as A, B and C. Here categories A and C have clearly defined motives: each has a clearly defined General direction of its intended goals, but these two directions are completely different. In both categories, over a period of twenty-five years (from the Turing article of 1947 "the intellectual mechanism" to the publication of the 1973 report), some progress has been made, although expectations have often not been met. Category A consisted of purely applied engineering studies such as speech recognition, machine translation, and some other practical problems usually solved by people. Category C contains all that is associated with the cognitive, neuromorphic, brain-like computing. Some promising research work in the field of neuroscience was noted.

В течение того же периода проводилась исследования еще одной категории (категория B), где цели и задачи гораздо труднее различить, но которая в значительной степени опирается на идеи как из А, так и из С и наоборот стремится влиять на них. Исследования в категории B, если приемлемые аргументы для этого могут быть согласованы, работают на основе его взаимозависимости с исследованиями в категориях A и C, чтобы обеспечить единство и согласованность всей области исследований ИИ. Вместе с тем прогресс в этой промежуточной категории В вызвали еще большее разочарование как в отношении фактически проделанной работы, так и в отношении установления веских причин для такой работы и, таким образом, для создания какой-либо единой дисциплины, охватывающей категории А и С.

During the same period, research was conducted in another category (category B), where the goals and objectives are much more difficult to distinguish, but which is strongly based on ideas from both A and C and Vice versa, to influence them. Studies in category B, if acceptable arguments for this can be agreed, work based on its interdependence with research in categories A and C to ensure the unity and consistency of the entire AI research area. However, progress in this intermediate category B has been even more disappointing, both in relation to the actual work done and in relation to the establishment of good reasons for such work and thus for the establishment of a single discipline covering categories A and C.

Далее Фримен Дайсон говорит: смотрите, прошло еще почти 50 лет, а раскладка — не изменилась. По-прежнему бурно развивается первое направление: компьютер уже неплохо распознает и переводит. На третьем направлении успехи также заметны: мы картировали мозг и стали лучше разбираться в его функциях. А на втором направлении — по-прежнему полный ноль.

Then Freeman Dyson says: look, it's been almost 50 years, and the layout has not changed. The first direction is still developing rapidly: the computer already recognizes and translates well. In the third area, the success is also noticeable: we get the brain map and began to better understand its functions. And on the second direction we have still full zero.

Не меньшее разочарование на сегодняшний день вызывает рост затрат вычислительных мощностей на глубокое обучение. В исследовании (AMODEI & HERNANDEZ, 2018) приведены цифры показывающие, что с 2012 года количество вычислений, используемых в крупнейших тренировочных прогонах AI, растет экспоненциально с 3,5-месячным периодом удвоения (для сравнения, закон Мура имел 18-месячный период удвоения). При таком росте затрат развитие вычислительной техники не успевает за потребностями.

No less frustrating today is the increase in the cost of computing power for deep learning. The study (AMODEI & HERNANDEZ, 2018) presents figures showing that since 2012 the number of computations used in the largest AI training runs has been growing exponentially with a 3.5-month doubling period (for comparison, Moore's law had an 18-month doubling period). With this increase in costs, the development of computer technology does not keep up with the needs.

В итоге приходим к выводу, что опасаться появления сверхмощного искусственного интеллекта пока не следует. Человечество гораздо быстрее сумеет убить себя, продолжая следовать другим восьми обозначенным Конрадом Лоренцом греховным традициям. А потому опасность ИИ скорее косвенная, связанная с педалированием тех восьми традиций. Это касается и возможности нечаянно спровоцированной ядерной войны, и гонки человечества с самим собой, и утраты чувств, и разрушения традиций. Но главная опасность – Индоктринируемость – болезнь разума цивилизованного человечества.

As a result, we conclude that the fear of the emergence of super-artificial intelligence should not be. Humanity have good chance to kill himself much quicker continuing to follow the other eight sinful traditions that are marked by Konrad Lorenz. But the danger of AI is rather indirect it associated with the pedaling of the eight traditions. This applies to the possibility of unintentionally provoked nuclear war, and the race of humanity with itself, and the loss of feelings, and the destruction of traditions. But the main danger is Indoctrinability that is the civilized humanity's mind's disease.

AMODEI & HERNANDEZ (2018) AI and Compute By DARIO AMODEI & DANNY HERNANDEZ <https://goo.gl/CaZCZ8>

IBM (2009) The Cat is Out of the Bag: Cortical Simulations with109 Neurons, 1013 Synapses Rajagopal Ananthanarayanan1, Steven K. Esser1 Horst D. Simon2, and Dharmendra S. Modha1

NBER (2017) ARTIFICIAL INTELLIGENCE AND THE MODERN PRODUCTIVITY PARA-DOX: A CLASH OF EXPECTATIONS AND STATISTICS, by Erik Brynjolfsson, Daniel Rock, Chad Syverson, Working Paper 24001 NATIONAL BUREAU OF ECONOMIC RESEARCH 1050[[1]](#footnote-1) Massachusetts Avenue Cambridge, MA 02138, November 2017

CB Insights (2018), Top AI Trends To Watch In 2018

<https://www.nytimes.com/2018/05/18/opinion/artificial-intelligence-challenges.html>

15 Trends Shaping Tech In 2018

Dyson, F. (2014) IS LIFE ANALOG OR DIGITAL?

A.I. Is Harder Than You ThinkA.I. Is Harder Than You Think By Gary Marcus and Ernest Davis

Mr. Marcus is a professor of psychology and neural science. Mr. Davis is a professor of computer science. May 18, 2018

Lorenz, Konrad. Die acht Todsu¨nden der zivilisierten Menschheit. R. Piper & Co. Verlag, Mu¨nchen, 1973.

RAND (2018) How Might Artificial Intelligence Affect the Risk of Nuclear War? by Edward Geist, Andrew J. Lohn, Perspective EXPERT INSIGHTS ON A TIMELY POLICY ISSUE, <https://www.rand.org/pubs/perspectives/PE296.html>

AI and the ‘Augmentation’ Fallacy Philipp Gerbert May 16, 2018 <https://sloanreview.mit.edu/article/ai-and-the-augmentation-fallacy/>

Artificial Intelligence as a Positive and Negative Factor in Global Risk Eliezer Yudkowsky Machine Intelligence Research Institute, Edited by Nick Bostrom and Milan M. Ćirković, 308–345. New York: Oxford University Press.

1. http://www.nber.org/papers/w24001 [↑](#footnote-ref-1)