



The Holy See

ПОСЛАНИЕ СВЯТЕЙШЕГО ОТЦА ФРАНЦИСКА НА LVII ВСЕМИРНЫЙ ДЕНЬ МИРА

1 ЯНВАРЯ 2024 ГОДА

Искусственный интеллект и мир

В начале нового года, времени благодати, которое Господь дарует каждому из нас, я хочу обратиться к Народу Божьему, нациям, главам государств и правительствам, представителям различных религий и гражданского общества, а также ко всем людям нашего времени с наилучшими пожеланиями мира.

1. *Научно-технический прогресс как путь к миру*

Священное Писание свидетельствует, что Бог наделил людей Своим Духом, дабы они обладали «мудростью, разумением и ведением и всяким искусством» (Исх 35,31). Разум есть выражение достоинства, дарованного нам Творцом, Который создал нас по Своему образу и подобию (см. Быт 1,26) и позволил свободно и сознательно отвечать на Его любовь. Наука и техника особым образом проявляют это главным образом межличностное качество человеческого интеллекта: они являются выдающимися плодами его творческого потенциала.

В пастырской конституции *Gaudium et Spes* Второй Ватиканский собор подтвердил эту истину, заявив, что «своим трудом и дарованием человек всегда старался расширить рамки собственной жизни» [1]. Когда люди «с помощью технологий» стремятся сделать землю «достойным местом обитания всей человеческой семьи» [2], они действуют согласно замыслу Бога и сотрудничают с Его волей, чтобы довести творение до конца и установить мир между народами. Также научно-технический прогресс – в той мере, в какой он способствует благоустройству человеческого общества, росту свободы и братского общения, – ведёт к совершенствованию человека и преобразованию мира.

Мы по праву радуемся и признательны за впечатляющие достижения науки и техники, благодаря которым были устранены неисчислимы бед, причинявшие человеческой жизни тяжкие страдания. В то же время научно-технический прогресс, позволяя осуществлять беспрецедентный доселе контроль над реальностью, вручает человеку огромное количество возможностей, некоторые из которых могут представлять опасность для нашего выживания и угрозу для общего дома [3].

Поразительные достижения новых информационных технологий, особенно в цифровой сфере, открывают захватывающие возможности и таят в себе серьёзные риски, значительно влияя на стремление к справедливости и гармонии между народами. Поэтому необходимо задать ряд неотложных вопросов. Каковы будут среднесрочные и долгосрочные последствия новых цифровых технологий? Какое влияние они окажут на жизнь отдельных людей и общества, на международную стабильность и мир?

2. Будущее искусственного интеллекта: между перспективами и риском

Прогресс в области информатики и развитие цифровых технологий в последние десятилетия уже привели к глубоким преобразованиям в глобальном обществе и его динамике. Новые цифровые инструменты меняют облик коммуникаций, государственного управления, образования, потребления, взаимоотношений и бесчисленных других аспектов повседневной жизни.

Кроме того, технологии, использующие множество алгоритмов, могут извлекать из цифровых следов в *Интернете* данные, которые позволяют контролировать мыслительные и межличностные привычки людей в коммерческих или политических целях, часто без их ведома, тем самым ограничивая их сознательное осуществление свободы выбора. Действительно, в таком пространстве, как *Интернет*, характеризующемся информационной перегрузкой, они могут структурировать поток данных согласно критериям отбора, не всегда осознаваемым пользователем.

Необходимо помнить, что научные исследования и технологические инновации не оторваны от реальности и «не нейтральны» [4], а подвержены культурному влиянию. Являясь целиком человеческой деятельностью, направления их развития отражают выбор, обусловленный личными, социальными и культурными ценностями той или иной эпохи. То же самое относится к их достижениям: являясь результатом специфически человеческого подхода к окружающему миру, они всегда имеют этическое измерение, тесно связанное с решениями тех, кто разрабатывает эксперименты и направляет производство к определённым целям.

Это относится и к разновидностям искусственного интеллекта. На сегодняшний день в мире науки и техники не существует единого определения данного понятия. Сам термин, уже вошедший в обиход, охватывает целый ряд наук, теорий и методов, направленных на то,

чтобы машины воспроизводили или имитировали в своей работе когнитивные способности человека. Выражение «формы интеллекта» во множественном числе помогает прежде всего подчеркнуть непреодолимый разрыв, существующий между этими системами – какими бы удивительными и мощными они ни были – и человеческой личностью: в конечном итоге они «фрагментарны» в том смысле, что могут лишь имитировать или воспроизводить определённые функции человеческого интеллекта. Использование множественного числа также подчёркивает, что эти весьма разные устройства всегда нужно рассматривать как «социотехнические системы». Ведь влияние любого устройства искусственного интеллекта, независимо от лежащей в его основе технологии, зависит не только от его проектирования, но и от целей и интересов его владельцев и разработчиков, а также от ситуаций, в которых он будет применяться.

Таким образом, искусственный интеллект следует воспринимать как галактику различных реалий, и мы не можем априори считать, что его развитие внесёт благотворный вклад в будущее человечества и в мир между народами. Такой положительный результат будет возможен лишь в том случае, если мы будем действовать ответственно и уважать такие фундаментальные человеческие ценности, как «инклюзивность, прозрачность, безопасность, справедливость, конфиденциальность и надёжность» [5].

Недостаточно также полагать, что разработчики алгоритмов и цифровых технологий будут действовать этично и ответственно. Необходимо укрепить или, при необходимости, создать органы для изучения возникающих этических проблем и защиты прав тех, кто использует формы искусственного интеллекта или подвергается их влиянию [6].

Поэтому колоссальное распространение технологий должно сопровождаться надлежащей подготовкой к ответственности за их развитие. Свобода и мирное сосуществование оказываются под угрозой, когда люди поддаются соблазну эгоизма, корысти, стремления к прибыли и жадности власти. Поэтому необходимо расширить взгляд и направить научно-технические исследования на достижение мира и общего блага, в интересах целостного развития человека и общества [7].

Неотъемлемое достоинство каждого человека и братство, связывающее нас как членов единой человеческой семьи, должны лежать в основе развития новых технологий и служить неоспоримыми критериями их оценки перед использованием, чтобы цифровой прогресс происходил с соблюдением справедливости и способствовал делу мира. Технические разработки, которые не ведут к улучшению качества жизни всего человечества, а, напротив, усугубляют неравенство и конфликты, никогда не смогут считаться подлинным прогрессом [8].

Искусственный интеллект будет приобретать всё большее значение. Вызовы, которые он ставит перед собой, носят не только технический, но и антропологический,

образовательный, социальный и политический характер. Он обещает, например, освобождение от рутинной работы, более эффективное производство, облегчённую транспортировку и более динамичные рынки, а также революцию в процессах сбора, организации и проверки данных. Мы должны осознавать стремительность происходящих преобразований и управлять ими таким образом, чтобы обеспечить защиту основных прав человека, уважая институты и законы, способствующие целостному развитию человечества. Искусственный интеллект призван служить лучшему человеческому потенциалу и нашим самым высоким устремлениям, а не конкурировать с ними.

3. *Технология будущего: самообучающиеся машины*

В своих многочисленных формах искусственный интеллект, основанный на методах машинного обучения (*machine learning*), хотя и находится на этапе становления, уже вносит значительные изменения в структуру общества, оказывая глубокое влияние на культуру, социальное поведение и миростроительство.

Такие разработки, как машинное обучение (*machine learning*) или глубокое обучение (*deep learning*), поднимают вопросы, которые выходят за рамки технологий и инженерии и тесно связаны с пониманием смысла человеческой жизни, основными процессами познания и способностью разума постигать истину.

Например, умение некоторых устройств создавать синтаксически и семантически связанные тексты не является гарантией их надёжности. Считается, что они могут «галлюцинировать», то есть генерировать утверждения, которые на первый взгляд кажутся правдоподобными, но на самом деле являются необоснованными или содержат предвзятость. Создаёт серьёзную проблему использование искусственного интеллекта в кампаниях по дезинформации, распространяющих фальшивые новости и умножающих недоверие к СМИ. Конфиденциальность, владение данными и интеллектуальная собственность являются следующими областями, где технологии могут представлять потенциальную опасность. К ним добавляются и другие негативные последствия их неправильного использования, такие как дискриминация, вмешательство в избирательные процессы, слежка и контроль за населением, цифровая изоляция и обострение индивидуализма, всё больше оторванного от общества. Все эти факторы могут подпитывать конфликты и препятствовать установлению мира.

4. *Чувство предела в технократической парадигме*

Наш мир слишком огромен, разнообразен и сложен, чтобы его можно было полностью познать и классифицировать. Человеческий разум никогда не сможет исчерпать его богатства, даже с помощью наиболее продвинутых алгоритмов. Они, по сути, не дают гарантированных прогнозов на будущее, а лишь статистические упрощения. Не всё можно

предсказать, не всё можно просчитать; в конце концов, «реальность выше идеи» [9], и какими бы выдающимися ни были наши вычислительные способности, всегда существует недоступный остаток, который ускользнёт от любых попыток количественной оценки.

Кроме того, гигантский объём данных, анализируемых искусственными интеллектами, сам по себе не является гарантией беспристрастности. Когда алгоритмы экстраполируют информацию, они всегда рискуют исказить её, воспроизводя несправедливость и предрассудки среды, в которой они зародились. Чем быстрее и сложнее они становятся, тем труднее понять, почему они выдают тот или иной результат.

«Умные» машины могут выполнять поставленные перед ними задачи с прогрессирующей эффективностью, но цель и смысл их деятельности по-прежнему будут определяться или обеспечиваться людьми, обладающими собственной системой ценностей. Риск заключается в том, что критерии, лежащие в основе определённых решений, станут менее ясными, ответственность за их принятие будет скрыта, а производители могут уклониться от обязательств работать на благо общества. В каком-то смысле этому благоприятствует технократическая система, которая объединяет экономику с технологией и отдаёт предпочтение критерию эффективности, при этом склоняясь к игнорированию всего, что не связано с её прямыми интересами [10].

Это должно заставить задуматься о том, что часто упускается из виду в современном технократическом и ориентированном на эффективность менталитете: о «чувстве предела», поскольку оно имеет решающее значение для личного и общественного развития. Человек, смертный по определению, намереваясь выйти за любые пределы с помощью технологий, рискует потерять контроль над собой в навязчивом желании контролировать всё; в поисках абсолютной свободы мы рискуем попасть в спираль «диктатуры технологии». Признание и принятие собственной ограниченности как существ – необходимое условие для достижения, а точнее, принятия самореализации как дара. Однако в идеологическом контексте технократической парадигмы, вдохновлённой пророческой презумпцией самодостаточности, неравенство может вырасти за пределы, а знания и богатство – скопиться в руках немногих, что чревато серьёзными рисками для демократических обществ и мирного сосуществования [11].

5. Насущные вопросы этики

В будущем надёжность претендента на ипотеку, пригодность человека для трудоустройства, вероятность рецидива у осуждённого, право на получение политического убежища или социальной помощи смогут определяться системами искусственного интеллекта. Отсутствие различных уровней посредничества, которое вводят эти системы, особым образом приводит к риску предвзятости и дискриминации: системные ошибки могут легко множиться, порождая не только несправедливость в отдельных случаях, но и –

в силу эффекта домино – реальные формы социального неравенства.

Кроме того, порой формы искусственного интеллекта способны повлиять на решения людей с помощью заранее определённых вариантов, связанных со стимулами и сдерживающими факторами, или действуя через систему регулирования личного выбора, основанную на организации информации. Такие формы манипулирования или социального контроля требуют пристального внимания и надзора, а также предполагают чёткую юридическую ответственность со стороны их разработчиков, заказчиков и государственных органов.

Полагаясь на автоматические процессы классификации людей, например, путём повсеместного наблюдения или через системы социального кредитования, можно также серьёзно повлиять на социальную структуру, устанавливая рейтинг среди граждан. Кроме того, эти искусственные процессы ранжирования могут привести к конфликтам власти, поскольку они касаются не только виртуальных пользователей, но и реальных людей. Фундаментальное уважение к человеческому достоинству требует, чтобы мы не отождествляли уникальность человека с набором данных. Нельзя допустить, чтобы алгоритмы определяли наше понимание прав человека, отбрасывали такие важнейшие человеческие ценности, как сострадание, милосердие и прощение, или исключали возможность того, что человек может измениться и оставить прошлое позади.

В этом контексте нельзя не отметить влияние новых технологий на трудовые условия: рабочие места, которые раньше были прерогативой исключительно человеческого труда, быстро поглощаются промышленными приложениями искусственного интеллекта. В этом случае также существует значительный риск получения непропорционально больших выгод немногими за счет обнищания многих. Уважение достоинства трудящихся и важность занятости для экономического благосостояния отдельных людей, семей и обществ, гарантии занятости и справедливой заработной платы должны стать приоритетными для международного сообщества по мере того, как эти формы технологии всё глубже проникают на рабочие места.

6. Перекуём ли мы мечи на лемехи плугов?

В наши дни при взгляде на окружающий мир не избежать серьёзных этических вопросов, связанных с индустрией вооружений. Возможность проведения военных операций с помощью дистанционно управляемых систем привела к снижению восприятия наносимых ими разрушений, к снижению ответственности за их использование, способствуя ещё более холодному и отстранённому подходу к огромной трагедии войны. Серьёзную этическую озабоченность вызывают исследования новых технологий в области так называемых «смертоносных автономных систем вооружения», включая военное использование искусственного интеллекта. Автономные системы оружия никогда не смогут стать морально ответственными субъектами: уникальная человеческая способность к нравственному

суждению и принятию этических решений – это нечто большее, чем сложный набор алгоритмов, и эта способность не может быть сведена к программированию машины, которая, какой бы «умной» она ни была, всё равно остаётся машиной. По этой причине крайне важно обеспечить адекватный, полный и последовательный человеческий надзор за системами вооружений.

Также нельзя игнорировать вероятность того, что сложное оружие попадёт в чужие руки, способствуя, например, террористическим атакам или интервенциям, направленным на дестабилизацию законных государственных институтов. Одним словом, миру не нужны новые технологии, способствующие несправедливому развитию рынка оружия и торговли им, поощряющие безумие войны. При этом не только интеллект, но и само человеческое сердце рискует стать всё более «искусственным». Самые передовые технические разработки должны использоваться не для продвижения насильственного разрешения конфликтов, а чтобы прокладывать путь к миру.

В более позитивной перспективе использование искусственного интеллекта для целостного развития человечества может привести к значительным инновациям в сельском хозяйстве, образовании и культуре, к повышению уровня жизни целых стран и народов, росту человеческого братства и социальной дружбы. В конечном счёте то, как мы используем искусственный интеллект в пользу наименьших, то есть наших самых слабых и нуждающихся братьев и сестёр, является показателем нашей человечности.

Гуманное мировоззрение и стремление к лучшему будущему нашего мира приводят к необходимости междисциплинарного диалога, направленного на этическое развитие алгоритмов, – *алгорэтику*, – в котором именно ценности направляют пути новых технологий [12]. Этические соображения должны приниматься во внимание с самого начала исследований, а также на этапах тестирования, проектирования, производства, дистрибуции и маркетинга. Таков подход этики проектирования, в котором образовательные учреждения и лица, принимающие решения, должны сыграть ключевую роль.

7. Вызовы для образования

Развитие технологий, которые уважают человеческое достоинство и служат ему, имеет очевидные последствия для учебных заведений и для мира культуры. Умножая возможности коммуникации, цифровые технологии позволяют по-новому встречаться друг с другом. Однако по-прежнему необходимо размышлять о том, к каким отношениям они приводят. Молодёжь растёт в культурной среде, пронизанной технологиями, и это не может не бросать вызов методам преподавания, воспитания и обучения.

Образование в области использования форм искусственного интеллекта должно быть направлено прежде всего на развитие критического мышления. Пользователям всех

возрастов, особенно молодёжи, необходимо развивать способность к анализу при использовании данных и контента, собранных в *Интернете* или созданных системами искусственного интеллекта. Школы, университеты и научные сообщества призваны помочь студентам и специалистам осознать социальные и этические аспекты развития и использования технологий.

Подготовка к использованию новых средств коммуникации должна учитывать не только дезинформацию, *фейковые новости*, но и тревожное возрождение «унаследованных от предков страхов, которые так и не были преодолены техническим прогрессом» [13]. К сожалению, нам снова приходится бороться с «соблазном построить культуру стен и возводить стены, которые препятствуют встрече с другими культурами, с другими людьми» [14] и развитию мирного и братского сосуществования.

8. Вызовы для развития международного права

Глобальный масштаб искусственного интеллекта делает очевидным, что, наряду с ответственностью суверенных государств за регулирование его использования внутри страны, международные организации могут сыграть решающую роль в достижении многосторонних соглашений и координации их применения и реализации [15]. В связи с этим я призываю сообщество наций к совместной работе по принятию обязательного к исполнению международного договора, регулирующего разработку и использование искусственного интеллекта в его многочисленных формах. Целью такого регулирования, разумеется, должно быть не только предотвращение неудачных практик, но и поощрение эффективных практик, стимулирование новых, креативных подходов и содействие личным и коллективным инициативам [16].

В поисках нормативных моделей, которые могли бы стать этическим руководством для разработчиков цифровых технологий, следует определить человеческие ценности, которые должны лежать в основе усилий общества по разработке, принятию и соблюдению необходимых законодательных рамок. Составление этических рекомендаций по созданию форм искусственного интеллекта не может обойтись без рассмотрения более глубоких вопросов о смысле человеческого существования, о защите основных прав человека, о стремлении к справедливости и миру. Этот процесс этического и правового распознавания может стать ценной возможностью для совместных размышлений о том, какую роль должны играть технологии в нашей индивидуальной и общественной жизни и как их использование может способствовать созиданию более справедливого и гуманного мира. По этой причине при обсуждении вопросов регулирования искусственного интеллекта нужно учитывать мнения всех заинтересованных сторон, включая неимущих, обездоленных и других, которые часто остаются неслышанными в глобальных процессах принятия решений.

Я надеюсь, что эти размышления побудят к тому, чтобы прогресс в развитии форм искусственного интеллекта в конечном итоге послужил делу человеческого братства и мира. За это отвечают не единицы, а вся человеческая семья. Мир, по сути, является плодом отношений, в которых признаётся и приветствуется достоинство другого человека; он – плод сотрудничества и стремления к целостному развитию всех людей и народов.

В начале нового года я молюсь, дабы стремительное развитие форм искусственного интеллекта не привело к умножению существующего неравенства и несправедливости, но помогло положить конец войнам и конфликтам и облегчить многие формы страданий, угнетающих человеческую семью. Пусть христиане, верующие различных религий, люди доброй воли сообща и в гармонии воспользуются возможностями и справятся с вызовами, которые бросает цифровая революция, и таким образом передадут грядущим поколениям более солидарный, справедливый и мирный мир.

Из Ватикана, 8 декабря 2023 года.

ФРАНЦИСК

[1] N. 33.

[2] Там же, п. 57.

[3] См. Энциклика *Laudato si'* (24 мая 2015 г.), 104.

[4] Там же, 114.

[5] Аудиенция участникам симпозиума *Minerva Dialogues* (27 марта 2023 г.).

[6] См. там же.

[7] См. Послание исполнительному президенту *World Economic Forum* в Давосе (12 января 2018 г.).

[8] См. Энциклика *Laudato si'*, 194; *Обращение к участникам семинара «Общее благо в цифровую эпоху»* (27 сентября 2019 г.).

[9] Апост. посл. *Evangelii gaudium* (24 ноября 2013 г.), 233.

[10] См. Энциклика *Laudato si'*, 54.

[11] См. *Обращение к участникам пленарного заседания Папской академии защиты жизни* (28 февраля 2020 г.).

[12] См. *там же*.

[13] Энциклика *Fratelli tutti* (3 октября 2020 г.), 27.

[14] См. *там же*.

[15] См. *там же*, 170-175.

[16] См. Энциклика *Laudato si'*, 177.