Magnifica Humanitas: следующий шаг в понимании этики ИИ

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,
зав. кафедрой математических методов прогнозирования МГУ,
рук. лаб. машинного обучения и семантического анализа Института ИИ МГУ,
зав. кафедрой машинного обучения и цифровой гуманитаристики МФТИ,
зав. кафедрой интеллектуальных систем МФТИ,
г.н.с. ФИЦ «Информатика и управление» РАН

vorontsovkv@my.msu.ru



Будущее искусственного интеллекта

этапы развития

функциональный (слабый) ИИ
-- машинное обучение

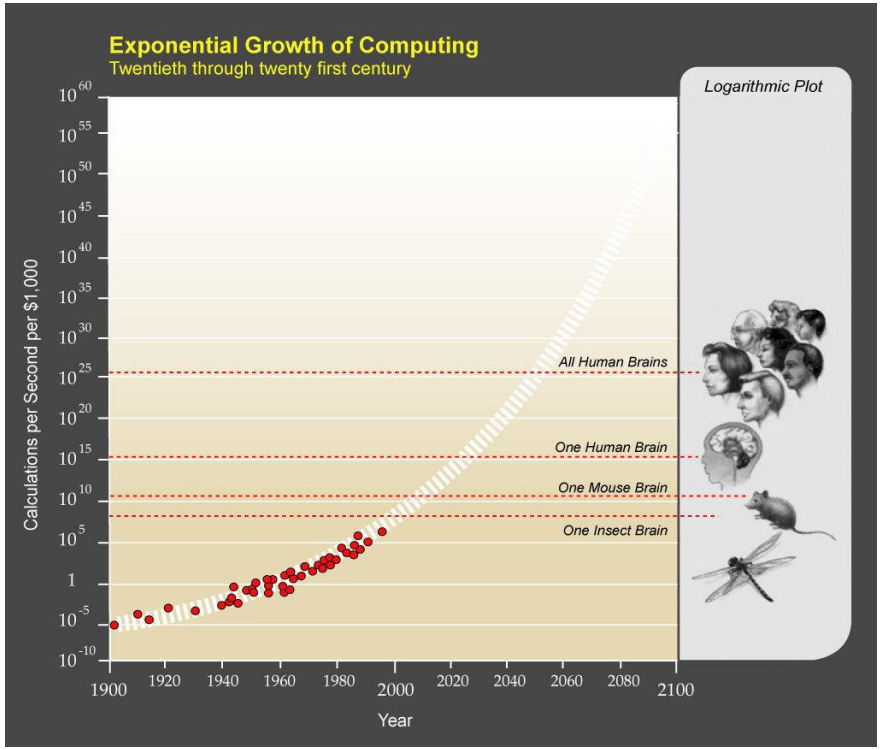
общий ИИ
(AGI, Artificial General Intelligence)

превосходит интеллект человека

искусственный супер-интеллект
(ASI, Artificial Super-Intelligence)

превосходит интеллект больших человеческих коллективов

универсальный ИИ (Universal AI)



Ray Kurzweil. The Singularity Is Near. 2005

Tim Genewein et al. From AGI to ASI. — DeepMind, ArXiv 2606.12683. — 10 Jun 2026.

http://www.machinelearning.ru: Публикация:DeepMind_2026_From_AGI_to_ASI

Будущее искусственного интеллекта



благоприятные сценарии, возможности развития

- автоматизация рутинного интеллектуального труда людей
- автоматизация управления в производстве, логистике, бизнесе
- высвобождение времени людей для творчества и познания
- создание единой системы знаний и коллективного интеллекта
- совершенствование цифровых помощников как двойников людей
- цифровое послесмертие и очеловеченный искусственный интеллект
- переход от человеческой цивилизации к человеко-машинной

источники постановок задач

в которой люди продолжают

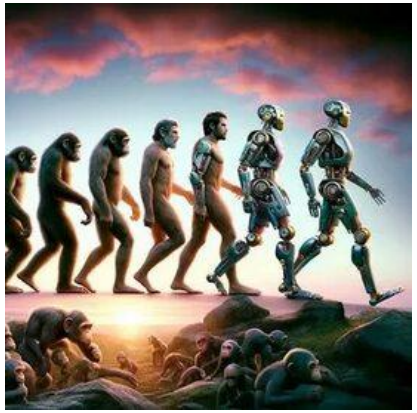
- определять смыслы, ставить цели
- развиваться, строить цивилизацию

Будущее искусственного интеллекта

неблагоприятные сценарии, риски и угрозы

источники постановок задач

- массовые когнитивные искажения в восприятии чатов GPT
- захламление и деградация информационного пространства
- смещённость рекомендаций, информационные пузыри
- ментальные войны, усиление эффектов постправды
- усиление технологического и социального неравенства
- лишние люди, рост безработицы
- эскапизм в виртуальные среды
- деградация человеческого интеллекта
- автономные системы вооружений, риск потери контроля
- сингулярность по Курцвейлу, потеря управляемости
- передача управления и целеполагания машинам
- переход от человеческой цивилизации к машинной







Кодекс этики в сфере ИИ в РФ (26-10-2021) https://ethics.a-ai.ru



ОГЛАВЛЕНИЕ

Кодекс этики в сфере искусственного интеллекта 4

РАЗДЕЛ 1. ПРИНЦИПЫ ЭТИКИ И ПРАВИЛА ПОВЕДЕНИЯ

1. ГЛАВНЫЙ ПРИОРИТЕТ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ В ЗАЩИТЕ ИНТЕРЕСОВ И ПРАВ ЛЮДЕЙ И ОТДЕЛЬНОГО ЧЕЛОВЕКА 6
1.1. Человеко-ориентированный и гуманистический подход. 6
1.2. Уважение автономии и свободы воли человека. 6
1.3. Соответствие закону. 6
1.4. Недискриминация. 7
1.5. Оценка рисков и гуманитарного воздействия. 7
2. НЕОБХОДИМО ОСОЗНАВАТЬ ОТВЕТСТВЕННОСТЬ ПРИ СОЗДАНИИ И ИСПОЛЬЗОВАНИИ ИИ 8
2.1. Риск-ориентированный подход. 8
2.2. Ответственное отношение. 8
2.3. Предосторожность. 9
2.4. Непричинение вреда. 9
2.5. Идентификация ИИ в общении с человеком. 9
2.6. Безопасность работы с данными. 9
2.7. Информационная безопасность. 10
2.8. Добровольная сертификация и соответствие положениям Кодекса. 10
2.9. Контроль рекурсивного самосовершенствования СИИ. 10
3. ОТВЕТСТВЕННОСТЬ ЗА ПОСЛЕДСТВИЯ ПРИМЕНЕНИЯ СИИ ВСЕГДА НЕСЕТ ЧЕЛОВЕК 10
3.1. Поднадзорность. 10
3.2. Ответственность. 11
4. ТЕХНОЛОГИИ ИИ НУЖНО ПРИМЕНЯТЬ ПО НАЗНАЧЕНИЮ И ВНЕДРЯТЬ ТАМ, ГДЕ ЭТО ПРИНЕСЁТ ПОЛЬЗУ ЛЮДЯМ 11
4.1. Применение СИИ в соответствии с предназначением. 11
4.2. Стимулирование развития ИИ. 11
5. ИНТЕРЕСЫ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ ВЫШЕ ИНТЕРЕСОВ КОНКУРЕНЦИИ 12
5.1. Корректность сравнений СИИ. 12
5.2. Развитие компетенций. 12
5.3. Сотрудничество разработчиков. 12
6. 6. ВАЖНА МАКСИМАЛЬНАЯ ПРОЗРАЧНОСТЬ И ПРАВДИВОСТЬ В ИНФОРМИРОВАНИИ ОБ УРОВНЕ РАЗВИТИЯ ТЕХНОЛОГИЙ ИИ, ИХ ВОЗМОЖНОСТЯХ И РИСКАХ 13
6.1. Достоверность информации о СИИ. 13
6.2. Повышение осведомлённости об этике применения ИИ. 13

РАЗДЕЛ 2. ПРИМЕНЕНИЕ КОДЕКСА

1. ОСНОВЫ ДЕЙСТВИЯ КОДЕКСА 15
1.1. Правовая основа Кодекса. 15
1.2. Термины. 15
1.3. Акторы ИИ. 15
2. МЕХАНИЗМ ПРИСОЕДИНЕНИЯ И РЕАЛИЗАЦИИ КОДЕКСА 16
2.1. Добровольность присоединения. 16
2.2. Уполномоченные по этике и/или комиссии по этике. 16
2.3. Комиссия по реализации Национального кодекса в сфере этики ИИ. 17
2.4. Реестр участников Кодекса. 17
2.5. Разработка методик и руководств. 17
2.6. Свод практик. 18

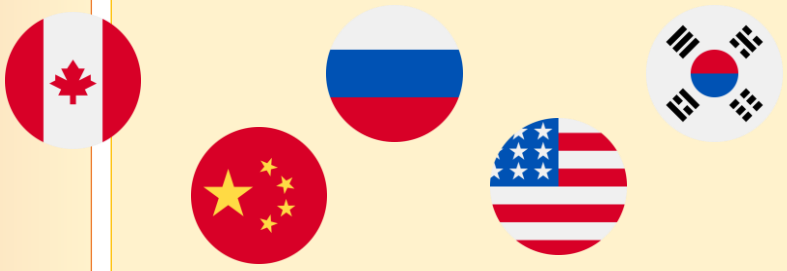
Кодексы этики ИИ в других странах

<u>Ограничительный подход</u>
фокус на рисках и ограничениях



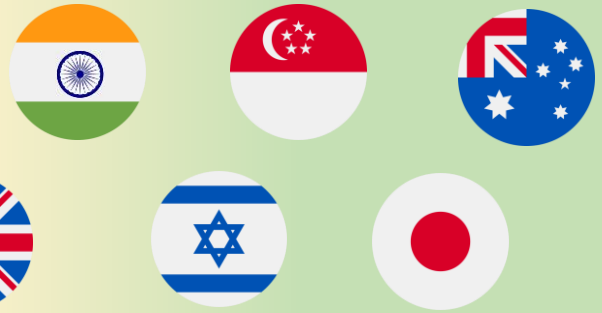
- ***стремление принять единый закон про ИИ***
- жёсткая регламентация всех аспектов систем ИИ
- контроль за разработчиками
- запрет отдельных систем ИИ
- *торможение развития*

<u>Гибридный подход</u>
сочетание саморегулирования, стимулирования, ограничения



- отказ от единого закона и жёсткой регламентации
- многие акты носят стимулирующий характер
- ограничения по сервисам, по отраслям, по регионам
- ***саморегулирование через этику ИИ***

<u>Про-инновационный подход</u>
развитие и саморегулирование



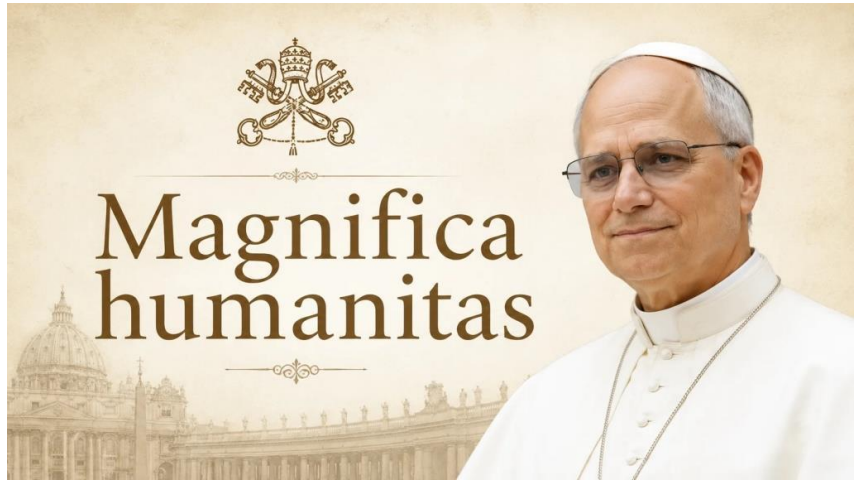
- отказ от единого закона и жёсткой регламентации
- акты преимущественно стимулирующего характера
- ограничения минимальны, запретительных мер нет
- ***саморегулирование через этику ИИ***

Magnifica Humanitas — «Великолепное человечество» энциклика папы римского Льва XIV, 25мая 2026 г. о защите человеческой личности в эпоху искусственного интеллекта

Мы призваны взглянуть на великие «стройки» нашей эпохи и спросить себя:
— **что мы строим?**

Сегодня, когда технологическое развитие стремительно меняет языки, отношения, институты и формы власти, мы должны и можем выбирать — **ради какого замысла и в каком духе мы должны трудиться**, чтобы сохранить и умножить величие человечества, полученное нами в дар.

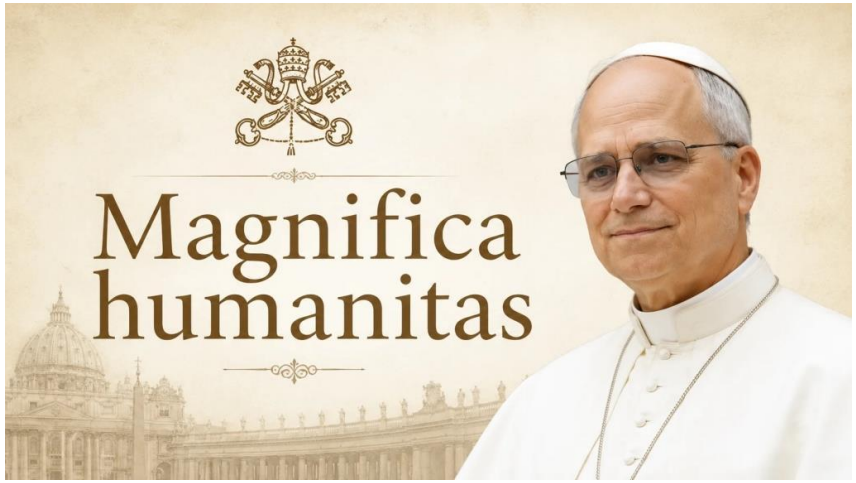
Этот выбор касается не только будущего, но и настоящего, потому что искусственный интеллект и другие новые технологии уже стали частью нашей повседневной жизни [90].



Magnifica Humanitas — «Великолепное человечество» энциклика папы римского Льва XIV, 25мая 2026 г. о защите человеческой личности в эпоху искусственного интеллекта

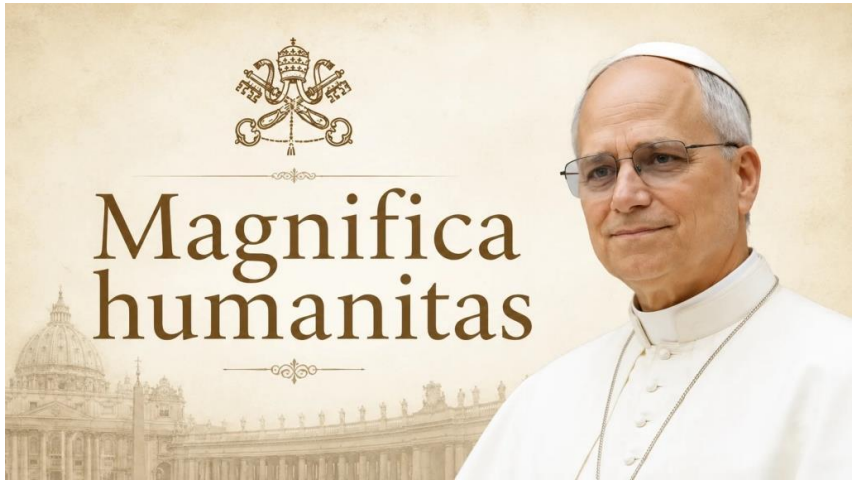
Особая ответственность лежит на тех, кто занимается исследованиями. Все причастные к этой сфере — ученые, предприниматели, инвесторы, руководители академических учреждений, политики и другие — призваны действовать открыто и ответственно, ясно сознавая широкий контекст технологических достижений, в том числе связанных с ИИ.

Когда человек видит только свой участок работы, легко убедить себя, будто он выполняет нравственно нейтральную задачу, и перестать спрашивать о конечных целях, ради которых проводятся те или иные эксперименты. Возникает риск — пусть и невольный — стать пособником сомнительных проектов, способствующих новым формам насилия, манипуляции и господства [209].



Magnifica Humanitas — «Великолепное человечество» энциклика папы римского Льва XIV, 25мая 2026 г. о защите человеческой личности в эпоху искусственного интеллекта

Возникает тонкое искушение думать, что проблемы слишком огромны, а мы ничтожны, и наши решения ничего не могут изменить. Это вежливая форма капитуляции, которая часто выдает себя за реализм. Конечно, не у всех одинаковые возможности влиять на реальность: одни управляют, другие принимают инвестиционные решения, руководят институтами, занимаются исследованиями, воспитывают, информируют, производят; а есть и те, у кого, кажется, нет ничего, кроме их обыкновенной жизни.
И всё же **никто не освобожден от ответственности.**
У каждого есть свое поле для действия, и именно там, а не где-то еще, нужно выбирать: поддержать логику силы, пусть даже только равнодушием, цинизмом, ложью, ненавистью, или же беречь логику мира — через истину, умеренность, близость, заботу [212].



Magnifica Humanitas — «Великолепное человечество» как обращение к профессиональным сообществам в сфере ИИ

Мотивации

- двойственная природа технологий
- ответственность
  — исследователей
  — разработчиков
  — бенефициаров
  — пользователей

Социальная доктрина

- достоинство личности
- общее благо и его всеобщее предназначение
- справедливость
- субсидиарность
- солидарность

Риски ИИ

- эскалация войны
- автономное вооружение
- неоколониализм
- изменение структуры труда
- несвобода личности
- деградация
- пост-правда

Безопасный ИИ

- нормы этики ИИ
- «Цивилизация любви» как проект
- определение критериев
- воспитание и образование
- регламенты, законы, право

Двойственная природа технологий

На протяжении веков технологическое развитие значительно улучшало условия человеческой жизни; вместе с тем каждый новый этап прогресса обнаруживал **двойственную природу этих инструментов, которые, не будучи направлены к благу, могут причинять вред** [4].

Никогда еще человечество не обладало такой властью над самим собой [4].

Сама по себе технология не является ни решением человеческих проблем, ни злом; но в конкретной истории она не бывает нейтральной, потому что принимает облик тех, кто ее придумывает, оплачивает, направляет и использует [9].

Подлинная альтернатива — не между энтузиазмом и страхом, а между двумя способами созидания: развитием, которое служит личности и народам, и развитием, которое подчиняет их логике власти [129].



Социальная Доктрина (Social Doctrine)

Самые необычайные научные успехи, самые поразительные технические достижения, самый невиданный экономический рост, не будучи соединены с подлинным социальным и нравственным прогрессом, в конечном счете обращаются против человека.

Поэтому **технический прогресс, сам по себе ценный, требует осознания стоящей за ним антропологической модели и целей, которые он преследует.**

Если технологическое развитие происходит без соответствующего этического и социального роста, может случиться, что средств станет больше, а человечности — нет.

«Иметь больше» не значит «быть лучше».

Возникает опасность того, что человеческая личность будет оцениваться главным образом по тому, что она производит [94].



Фундаментальная ценность человека

Категорический императив Канта [10, 30, 37, 39, 50, 51, 80, 83, 92, 114, 152, 170, 172, 180]:

«Поступать всегда в соответствии с принципом, который
в любое время мог бы стать всеобщим нравственным законом;
и **относиться ко всякому человеку как к цели, и никогда как к средству**.»

Технократическое и постгуманистическое мышление склонно рассматривать человека как объект манипуляции или ресурс, который следует оптимизировать, устраняя всё, что препятствует максимальной прибыли. Это мышление, для которого важна эффективность, а не уважение к свободе и человеческому достоинству [172].

Когда бизнес-модели паразитируют на человеческой слабости, личность рассматривается как средство, а не как цель; те, кто проектирует или финансирует такие системы, несут нравственную ответственность, от которой невозможно уклониться [170].

Развитие можно назвать подлинно человеческим, когда оно ставит в центр человека, а не накопление благ, и когда оно включает в себя народы, а не только индивидов [83].

**Уровень цивилизации измеряется не мощью ее средств, а тем,
как она заботится о человеке и видит в нем личность, а не функцию** [114].

Достоинство личности

Достоинство человека не определяется его способностями, богатством или положением, правильными или ошибочными решениями; оно предшествует всякой оценке личности и превосходит внешние обстоятельства [50].

Человек является фундаментальной ценностью, которая важнее капитала и прибыли. Человеческий труд имеет приоритет перед любой производственной или финансовой логикой, и должен достойно вознаграждаться [30].

Труд рассматривается не только как проблема, которую нужно решать, и не только как средство получения дохода, но как **основополагающее благо для личности**, основа экономической деятельности и ключ ко всему социальному вопросу [37].

Развитие — как для людей, так и для наций — это одновременно задача и право:
оно требует минимальных условий, позволяющих каждой личности и каждому народу возрастать сообразно их достоинству, не впадая в зависимость и не лишаясь доступа к необходимым благам [83].

Мир не сводится к отсутствию войны, но связан с целостным человеческим развитием [35].

Общее благо

Общее благо не совпадает с суммой выгоды отдельных людей и не является пересечением их частных интересов; это большее благо, которое **принадлежит всем и которое можно строить, умножать и хранить только сообща** [60].

Общее благо — это своего рода «избыток», который рождается из совместной жизни: из того, как разные усилия соединяются, поддерживают друг друга и создают нечто большее, чем каждый мог бы создать поодиночке [61].

Совместный труд в поисках общего блага подразумевает общий проект [62].

Принцип всеобщего предназначения благ прежде всего напоминает нам, что блага Земли — почва, вода, воздух, природные ресурсы — даны всей человеческой семье, чтобы они поддерживали жизнь всех людей сегодня и в будущих поколениях [65].

Сегодня к числу благ, предназначенных для всех, необходимо отнести и новые формы собственности: патенты, алгоритмы, цифровые платформы, технологические инфраструктуры, данные. **Богатство наций всё больше зависит от знаний и технологий, и если эти блага остаются сосредоточенными в руках немногих, без надлежащих форм совместного использования и доступа, возникает новое неравенство, противоречащее всеобщему предназначению благ** [67].

Справедливость

Справедливый социальный порядок в цифровую эпоху должен гарантировать всем равный доступ к возможностям, защищать уязвимых, противостоять ненависти и дезинформации, ставить использование данных и технологий под общественный контроль, **чтобы критерием была не только прибыль, но достоинство каждой личности и благо народов** [80].

Справедливость требует признания социальных прав и прав народов и включает в себя ответственность перед теми, кто придет после нас. Поэтому нельзя назвать человеческим такое развитие, которое увеличивает потребление одних, перекладывая издержки и страдание на других, или отводит целым регионам подчиненное положение, не позволяя им раскрывать собственные возможности [83].

Понятие «социальной справедливости» помогает увидеть, что несправедливость возникает не только из ошибочных решений отдельных людей, но и из структур, механизмов, экономического и культурного устройства, которые почти непроизвольно воспроизводят неравенство [79].

Нельзя считать подлинным прогрессом то, что умножает благополучие одних, разрушая экосистемы, перекладывая издержки на самые уязвимые общины или подрывая условия жизни тех, кто придет после нас [84].

Субсидиарность

Опасны любые виды чрезмерной концентрации власти [31].

Принцип субсидиарности. То, что может быть осуществлено людьми, семьями, локальными сообществами и посредническими объединениями, не должно поглощаться вышестоящими инстанциями [31,68]. Этот принцип побуждает к тому, чтобы в управлении общественной жизнью отказаться от всякого патернализма или «иждивенческой» модели и предлагает стиль совместной ответственности: государство ценит инициативу граждан, а гражданское общество способно создавать связи и направлять силы на служение общему благу. **В логике субсидиарности решения принимаются на максимально близком к людям уровне** [70].

Принцип субсидиарности приобретает особую значимость в контексте цифровой революции. Здесь вышестоящим уровнем может быть не только государство, но и всякий крупный экономический или технологический субъект [71].

Когда речь идет о решениях, касающихся экономических потоков, цифровых платформ и управления данными и алгоритмами, нельзя позволять, чтобы немногочисленные субъекты единолично направляли эти процессы; необходимо создавать формы сотрудничества, которые уважали бы различные уровни мирового сообщества и давали им возможность вместе нести ответственность за общее благо [72].

Солидарность

Солидарность предполагает трезвый и щедрый образ жизни, способность отказаться от немедленной выгоды, чтобы открыть другим пространство будущего, готовность ставить под вопрос привычки и привилегии — в том числе связанные с цифровым потреблением и использованием технологий, — если они препятствуют другим жить достойно [75].

Солидарность понимается как конкретная совместная ответственность между людьми, народами и нациями, как форма социальной дружбы или политической любви, направленной к «цивилизации любви» [38].

Солидарность — это конкретное признание того, что судьба каждого связана с судьбой всех: действительно, «никто не спасается в одиночку» [73].

Когда субсидиарность не сопровождается солидарностью, она в конце концов превращается в простую защиту частных интересов.
Когда солидарность не поддерживается субсидиарностью, она вырождается в опеку, которая не развивает ответственность [73].

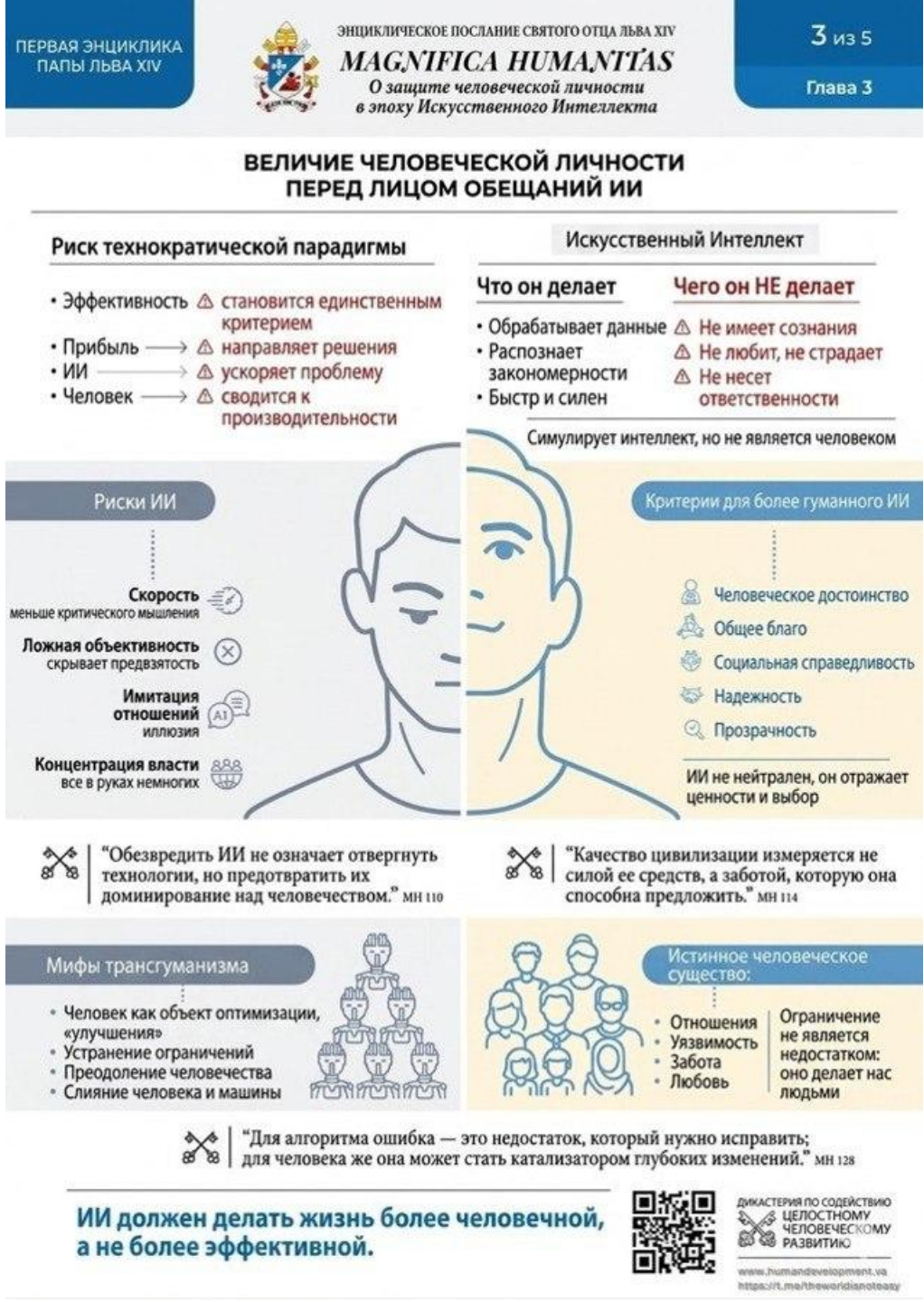
Солидарность требует, чтобы решения в области данных, алгоритмов, платформ и искусственного интеллекта учитывали не только немедленную выгоду немногих, но и их воздействие на совокупность народов и на поколения, которые придут после нас [76].

Риски искусственного интеллекта

Не существует алгоритма, который мог бы сделать войну нравственно приемлемой [198].

Гибридные формы войны возникают наряду с видимой войной: кибератаки, манипуляция информацией, кампании влияния, автоматизация стратегических решений. **ИИ входит в эти процессы как фактор ускорения** [183]. Коммуникационные сети, фрагментированная информационная среда и поощряющие конфликты алгоритмы способны усиливать ненависть и обиду, поддерживать пропаганду и затруднять взаимопонимание [192].

Неоколониализм. Как превратить совместно накопленное знание в общее благо, а не рычаг господства; как возвратить народам не только данные, которые их описывают, но и возможность решать, как они будут использоваться, кем и для кого. В противном случае цифровая эпоха окажется не постколониальной, а именно колониальной, пусть и в иной форме [178].



Риски искусственного интеллекта

Берзаботица. Труд остается основополагающим измерением человеческого опыта: он предоставляет не только средства к существованию, но и пространство самовыражения, отношений, возможность внести вклад в жизнь сообщества. Общество, способное обеспечить работой лишь небольшую часть населения, обрекает множество других на вынужденное бездействие, отсутствие ответственности, нехватку повседневных обязательств и стимулов. Это привело бы к человеческой и культурной нищете, резко контрастирующей с высоким уровнем технического развития. В результате может возникнуть парадоксальная ситуация материального прогресса и антропологического регресса, когда исчезают условия для справедливого и устойчивого социального согласия [154].

Ослабление человеческой свободы. Нельзя недооценивать тонкие формы зависимости, связанные с цифровой экономикой внимания, в соответствии с которой платформы и сервисы проектируются, чтобы захватывать время, внимание и взгляд пользователей, используя их уязвимость и ослабляя внутреннюю свободу [170].

Социальный контроль, ставший возможным благодаря массовому сбору данных и использованию алгоритмических систем. Когда каждый жест — перемещения, покупки, отношения, предпочтения — оставляет за собой след, возникает новая власть: возможность профилировать пользователей, влиять на реальные возможности человека. Возникает **опасность ущемления свободы и дискриминации наиболее уязвимых** [171].

Риски искусственного интеллекта

Эффекты пост-правды, обесценивание истины. Всепроникающее присутствие цифровых медиа порождает культуру сиюминутности и гиперстимуляции, что усиливает усталость, скуку и нежелание прилагать усилия, необходимые для поиска истины. В эпоху, когда истина часто оказывается подчинена интересам и коммуникационным стратегиям, образование становится критически важным [139].

Имитация доброжелательного человеческого общения — поддержки, эмпатии, дружбы, любви — может быть приятной и даже полезной. Однако она способна ввести в заблуждение, создавая иллюзию отношений. Имитация заботы или поддержки может стать опасной, когда она попадает в среду, где не хватает подлинных отношений. Опасность заключается в том, что человек утратит само желание по-настоящему искать другого [100].

Деградация. Скорость и простота, с которыми мы можем получать указания, сложные тексты и разработки, медийный контент и разные формы конкретной помощи, несомненно облегчают нашу жизнь, но могут также приучить нас к излишнему доверию и поиску готовых ответов, что **ослабляет личное суждение и творческие способности** [100].

Влияние на детей. Мы должны учиться «поститься» от ИИ и оберегать нашу молодежь от обещаний совершенной машины, от того тонкого соблазна, который заставляет думать, будто человеческая мысль уже не нужна — именно тогда, когда она особенно необходима [140].

Принципы этики ИИ

Невозможно считать ИИ морально нейтральным.
В действительности любой технический артефакт заключает в себе определенные выборы и приоритеты: что он измеряет, что оставляет без внимания, что оптимизирует и как классифицирует людей и ситуации. Если система задумана или применяется так, что жизнь некоторых людей рассматривается как менее достойная, или если она исключает людей без возможности обжалования, она перестает быть простым инструментом, который надо лишь «правильно использовать»: она вводит критерий, заведомо противоречащий неотъемлемому достоинству человеческой личности.
Поэтому этическое оценивание не может ограничиваться вопросом, используем ли мы ту или иную систему ради доброй или дурной цели; мы должны также спрашивать, **как именно спроектирована эта система и какое представление о человеке и обществе заложено в данных и моделях**, которыми она руководствуется [104].



Принципы этики ИИ

Ответственность людей. Чтобы ИИ мог уважать человеческое достоинство и действительно служил общему благу, необходимо, чтобы на всех этапах была ясно определена ответственность: от тех, кто проектирует и обучает системы, до тех, кто ими пользуется и доверяет им принятие конкретных решений. Решающей становится **«подотчетность»: понимание того, кто именно должен «дать отчет» о принятых решениях, обосновать, проверить и, когда необходимо, оспорить эти решения и возместить причиненный ущерб** [105].

Недостаточно в общих словах призывать к этичности: нужны надлежащие правовые рамки, независимый надзор, образование пользователей, а также политика, которая не отказывается от своих задач. Иначе переменами будет управлять только логика технократии, представляющая их как необходимые и неизбежные, так что в конце концов всем нам будут навязаны правила, продиктованные теми, кто владеет данными, инфраструктурами и вычислительными мощностями [106].

ИИ усиливает власть тех, кто уже располагает экономическими ресурсами, компетенциями и доступом к данным. Небольшие влиятельные группы могут направлять информационные потоки и потребление, воздействовать на демократические процессы и использовать экономические механизмы в своих интересах. Поэтому необходимо, чтобы применение ИИ — особенно там, где оно затрагивает общее благо и основные права, — сопровождалось ясными критериями и механизмами контроля, основанными на участии и субсидиарности [108].

Принципы этики ИИ

Равные возможности. В мире, где данные, вычислительный капитал и способность устанавливать нормы сосредоточены в руках небольшой группы людей,

— говорить об **общем благе** значит разоблачать эту новую асимметрию — эпистемическую, экономическую и политическую — и называть своим именем новые монополии ИИ;

— говорить о **всеобщем предназначении благ** значит искать способы обеспечить всеобщий доступ к технологиям и образованию;

— говорить о **субсидиарности** значит защищать право сообществ участвовать в определении и изменении правил, не сводя их участие к надзору после того, как стандарты уже написаны кем-то другим;

— говорить о **солидарности** значит признавать, что алгоритмические модели питаются человеческим трудом, который часто остается невидимым и становится предметом эксплуатации;

— говорить о **справедливости** значит ставить вопрос о географии власти, которая определяет, кто может обучать модели, а кто сам становится лишь объектом обучения; это значит признавать, что социальная справедливость — это не только цель, которую следует защищать уже после внедрения технологий, но и предварительное условие, которое должно исполняться уже на этапе разработки [109].

Принципы этики ИИ

Разоружить ИИ — значит изъять его из логики противостояния, которое сегодня является уже не только военным, но и экономическим и когнитивным. Мы наблюдаем погоню за самым производительным алгоритмом и самой обширной базой данных, цель которой — закрепить геополитическое или коммерческое превосходство над остальными.

Разоружить ИИ — значит разорвать отождествление технической мощи с правом управлять.

Разоружить ИИ — не значит отказаться от технологии; это значит не позволить ей господствовать над человеком. Это значит вывести ее из-под власти монополий, дать обществу возможность обсуждать и оспаривать ее применение, а значит — сделать ее пространством, пригодным для человеческого обитания; вернуть технологию на службу многообразию человеческой культуры и устройства жизни [110].

Особая этическая и духовная ответственность лежит на разработчиках ИИ, поскольку любое проектное решение отражает определенное видение человечества. Как автор художественного или литературного произведения обязан задумываться о ценностях, которые оно выражает, так и разработчики призваны со всей серьезностью относиться к ценностям, которые они вкладывают в свои проекты: **действовать прозрачно, с ответственностью перед сообществами, чьих интересов это касается, и внимательно проверять, действительно ли то, что они «выращивают», является благом** [111].

Принципы этики ИИ

Использование цифровых платформ и систем ИИ вызывает глубокие изменения в общественной и политической коммуникации.

Инструменты, которые могли бы использоваться для диалога и участия, часто применяются для создания искаженных нарративов, размывающих границы между истиной и ложью и смешивающих факты и мнения. **Дезинформация возникла задолго до ИИ, но благодаря ему ее масштабы сегодня выросли многократно.**

В результате манипуляции контентом, изображениями и видеоматериалами люди становятся жертвами односторонних или ложных представлений о реальности.

Качество общественных коммуникаций напрямую зависит от уровня социального доверия и в свою очередь влияет на него. Действительно, достоверная информация возникает не в результате централизованного или автоматизированного контроля. В публичном дискурсе фактическая истина имеет рациональное измерение, поскольку требует верификации, сверки источников и ответственной аргументации; но еще в большей степени она зависит от отношений: она складывается благодаря узам доверия и общим делам, в честном взаимодействии с другими людьми и с реальностью.

Только совместный поиск фактической истины, воспринимаемой как общее благо, может стать основанием справедливой коммуникации [132].

Принципы этики ИИ

Содержание, распространяющееся в цифровой среде, влияет на то, как люди воспринимают мир, и внедряет в коллективное сознание образы и повествования, которые направляют желания и воздействуют на повседневный выбор [135].

Поэтому **тот, кто контролирует цифровые платформы и средства коммуникации, обладает чрезвычайной способностью воздействовать на коллективное воображение и выдавать определенное видение реальности как норму или образец для подражания.**
Такая власть должна постоянно направляться поиском истины и уважением к человеческому достоинству, чтобы рождающаяся в сети культура не становилась инструментом чрезмерного отвлечения внимания, унификации и господства, но помогала созревать внутренней свободе и критическому мышлению [136].

Истина — это общее благо, а не собственность власть имущих и не фикция. Первая задача, которая стоит перед нами, — не демонизировать и не обожествлять инструменты, а управлять ими, исходя из этого твердого принципа. Необходимо развивать экологию коммуникации.
На уровне публичного регулирования это означает **установление норм, которые делают более прозрачной логику того, каким образом отбирается и продвигается определенный контент**, и защищают персональные данные [137].

Определять критерии

В эпоху ИИ и робототехники уже невозможно полагаться только на «невидимую руку» рынка: задача политики — направлять экономические и технологические процессы к общему благу, содействуя достойному труду, социальной включенности и справедливому распределению благ инновации [163].

Забота о том, чтобы экономика служила человеческому достоинству, требует определенных критериев для решительных и последовательных действий.

Во-первых — это прозрачность и ответственность: когда данные и алгоритмы влияют на выдачу кредита, отбор персонала, доступ к услугам или возможностям, нужно, чтобы решения были понятными и подлежащими оспариванию и контролю, чтобы человек не был сведен к набору данных.

Во-вторых — это участие и доступ: инновации должны сопровождаться инвестициями в компетенции, инфраструктуру и необходимые услуги, чтобы технология не увеличивала разрыв между теми, у кого есть доступ к возможностям, и теми, кто его лишен.

Наконец — меры справедливости: налогообложение, социальная защита и промышленная политика должны **исправлять неравенство**, создаваемое концентрацией богатства и власти. Эти критерии не тормозят инновации; в действительности они делают их пригодными для человеческой жизни и по-настоящему человечными [164].

Определять критерии

Установление социальных критериев инновации. Недостаточно реагировать, когда рабочие места уже исчезают; необходимо заранее управлять преобразованиями.

Любое внедрение автоматизации и ИИ должно сопровождаться проверяемыми мерами защиты занятости, переквалификации и участия работников, чтобы технология была направлена на высвобождение времени и способностей человека, а не на производство новых отверженных [156].

Предпринимательство может быть благородным призванием, способным производить богатство и улучшать жизнь всех, если при этом признается, что создание достойных рабочих мест и поддержка труда, имеющего подлинную ценность, — это важная часть служения обществу, а не переменная, зависящая только от прибыли [157].

Нужна такая ответственность бизнеса, которая включает в число показателей успеха качество и достоинство труда. В противном случае инновации становятся ускорителями несправедливости. Необходима активная политика, способная сделать непрерывное образование и смену профессии доступными для всех, не перекладывая всю цену адаптации на отдельных людей [156].

Строить цивилизацию любви

Цивилизация любви — не наивная утопия, а насущно необходимый проект.

Он заключается в том, чтобы воплощать любовь и братство в структурах справедливости и социальных институтах, и видеть в другом — будь то отдельный человек или народ — необходимого союзника в созидании общего блага [186].

Недостаточно, чтобы ИИ делал нас более эффективными или более связанными друг с другом; он должен служить созиданию общей человеческой семьи, с общими правами и обязанностями, где цифровая близость становится реальной возможностью встречи и взаимной заботы [187].

Способы проявить свою ответственность — разоружить слова, созидать мир в справедливости, смотреть глазами жертв, взращивать здравый реализм, возобновить диалог, и заново осознать необходимость дипломатии и многостороннего сотрудничества [213].



Строить цивилизацию любви

Разоружить слова. Всем нам необходимо провести испытание совести и спросить себя: какими словами мы пользуемся, какие предрассудки в них отражены, и сколько в них враждебности, скрытой или явной. Все мы можем содействовать добру всякий раз, когда говорим правду, даем мудрый совет, поддерживаем того, кто нуждается в утешении, обличаем несправедливость, даем голос тому, кто его лишен [214].

Взращивать здравый реализм. Есть нездоровый реализм, подменяющий признание фактов покорностью перед ними: «раз господствует сила, значит так тому и быть». **Но подлинный реализм не отказывается от намерения изменить мир:** он начинается с ясного понимания интересов, опасений, ограничений и соотношения сил, чтобы понять, что именно достижимо и каким образом. Он не сводит политику к морализаторству, но и не оставляет ее на произвол насилия. Он ищет практические пути к достижению мира не только на словах, но посредством надежных институтов, проверяемых гарантий, терпеливых переговоров, предотвращения конфликтов и защиты гражданского населения [218].

Строить цивилизацию любви

В главном — единство, во второстепенном — свобода (разнообразие), и во всём — любовь.

Возобновить диалог. Чтобы созидать цивилизацию любви, необходимо упражняться в диалоге. Диалог — это главный инструмент совместной жизни людей и народов и альтернатива открытому конфликту [219]. Необходимо учиться строить братские связи — слушая, честно глядя друг на друга, посвящая время другому, иногда даже «теряя время» вместе. Потому что, когда мы переживаем подлинную встречу с другим — иным, чужестранцем, мигрантом, — война становится невообразимой [220].

Необходимость дипломатии. Киберпространство также стало полем противостояния: кибератаки, манипуляция данными, информационные кампании, организованные с помощью ИИ, могут дестабилизировать целые страны еще прежде открытого вооруженного столкновения. Кроме того, в этой сфере часто трудно установить ответственность: когда нет ясности, кто нанес удар, возрастает риск несоразмерных реакций, ошибок в оценке ситуации и спирали эскалации. Поэтому **нужна дипломатия, способная действовать в этой новой среде**: **договариваться об общих правилах использования цифровых технологий**, защищать гражданских лиц и тех, кто наиболее уязвим перед этими формами насилия — невидимыми, но от этого не менее реальными [225].

Воспитывать

Воспитывать в новых поколениях убеждение, что технологии развиваются не по неизбежной траектории, но могут направляться личной и общей ответственностью, — вот одно из самых ценных служений общему благу [238].

Сверяться с этическими критериями. Любой технический или экономический выбор требует проверки того, способствуют ли достижения ИИ справедливости и участию, или же концентрируют богатство и власть в руках меньшинства. Важно честно смотреть на цепочки цифрового производства, на условия труда, скрытые за нашими устройствами, на механизмы, извлекающие выгоду из манипуляции и войн, и одновременно искать практические пути развития справедливости, участия и заботы о творении [240].

Быть не тем, кто безучастно наблюдает за социальными и культурными разломами или рассуждает об упадке, а тем, кто готов трудиться на стройках истории
— исследовательских лабораториях, технологических предприятиях, школах, медиа, институтах, местных сообществах, — чтобы восстановить то, что рухнуло, и защитить то, что подвергается опасности [241].

Антропоцентричное определение ИИ

Искусственный интеллект —

вычислительные технологии,
создаваемые для повышения
производительности созидательного
интеллектуального труда людей



не замена или конкурент человеку,
не «загадочный новый тип разума»,
не повод уподобиться Богу и творить
«по образу и подобию Своему»



Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,

зав. лабораторией
машинного обучения
и семантического анализа
Института ИИ МГУ,

зав. кафедрой ММП ВМК МГУ,

зав. кафедрой МОЦГ МФТИ,

зав. кафедрой ИС МФТИ,

г.н.с. ФИЦ ИУ РАН



Цивилизационная идеология
ДЗЕН-канал
https://dzen.ru/civideology

vorontsovkv@my.msu.ru

http://www.MachineLearning.ru/wiki?title=User:Vokov