What if Military AI is a Washout?

## Что, если военный ИИ - это отстой (отмыв денег)?

© Jack McDonald 2019, https://jackmcdonald.org/book/2021/06/what-if-military-ai-sucks/

Нам говорят, что военное применение искусственного интеллекта может трансформировать военную мощь. Они могут сделать океаны прозрачными для сенсорных систем, угрожая таким системам ядерного сдерживания на море, как британский Trident. Они могут позволить использовать автономные летательные аппараты, которые смогут одолеть самолеты с человеческим экипажем. Они могут преобразовать обработку разведывательных данных на войне, задействовать все виды сложного оружия, которое сделало бы такие вещи, как танки и авианосцы, вчерашними новостями. Кажется, что небо - это предел.

В свете этого крупные государства вкладывают большие средства в военный ИИ. Одним из аспектов недавнего Интегрированного обзора Великобритании (кхе- кхе, «Глобальная Британия в эпоху конкуренции») и Командного документа (кхе- кхе, «Оборона в эпоху конкуренции») является ставка на то, что инвестиции в военные приложения искусственного интеллекта компенсируют сокращение расходов, такие, как танки и численность войск. Для Великобритании это большая ставка. Это имеет важные последствия для того, как Великобритания будет вносить свой вклад в НАТО, и ее ключевые союзы, такие как «особые отношения» с Соединенными Штатами.

Так что же произойдет, если военный ИИ не окупится? В более британском смысле, что произойдет, если военный ИИ окажется не адекватным? Этот пост – мое личное рассуждение, основанное на моем растущем скептицизме по поводу «военной революции» под управлением ИИ в ближайшем будущем. Около пяти лет назад я писал о возможности использования систем искусственного интеллекта для обнаружения повстанческих сетей, но этот пост является откатом даже от этой (ограниченной) возможности.

Аргумент звучит примерно так: социально преобразующее представление об ИИ, проданное венчурным капиталом за последнее десятилетие, весьма похоже, исчезает, так как есть только несколько нишевых областей, требующих значительного улучшения, но не появилось никаких парков беспилотных такси в Лондоне. Интеграция некоторых технологий искусственного интеллекта обеспечит автоматизацию/автономию в некоторых частях уже существующих военных процессов (например, цепочки убийств), но без роботов-суперсолдат, стратегов HAL 9000 и ограниченных институциональных изменений. Это все равно окажет огромное влияние на войну, сделав распознаваемые машиной объекты уязвимыми для автоматического уничтожения любыми различными автономными системами. Эта асимметрия видимости затруднит проецирование силы и поддержку вооруженных сил на местах и сократит разрыв в потенциале, который государственные вооруженные силы стремятся сохранить по сравнению с негосударственными субъектами. Реагирование на эту асимметрию приведет к войне в городской среде. Вкратце: беспокойтесь о незначительных улучшениях того, что уже было реализовано, и о способах сдерживания распространения этих платформ, потому что будущее наступило сейчас (бесстыдно цитируя Non Phixion).

## I: Продавать генерализованный ИИ - легко, а зарабатывать деньги - сложно

Мы живем в эпоху крупных ставок венчурного капитала на ИИ в той или иной форме. Отчасти мой скептицизм по поводу ажиотажа в отношении ИИ в последнее десятилетие был вызван диссонансом между годами презентаций и тем, что превратилось в определяющие отрасль технологии. Хорошим практическим правилом в этом отношении является применение: трудно получить чистую прибыль (но не невозможно, посмотрите, как Google/Facebook переключают значительную часть глобальных рекламных/маркетинговых бюджетов на свои системы), а сокращение штата, заработной платы и пенсий - это серьезная проблема. довольно хороший способ заработать больше денег. В этой связи отвлекающей целью одной из ведущих ставок венчурного капитала – беспилотных транспортных средств – всегда было обещание, что вам, возможно, больше не придется платить за своего таксиста (или что вам, возможно, не придется учитывать большие расстояния), включать заработную плату водителя грузовика в стоимость ваших физических товаров.

Итак, как дела?

Многие крупные и общественные ставки венчурного капитала на беспилотные автомобили с поддержкой ИИ, похоже, разворачиваются на наших глазах. Учтите, что Lyft, Uber, Tesla за последнее десятилетие предлагали беспилотные автомобили. В то время как Tesla, по крайней мере, добилась огромных успехов в решении проблемы направления куска металла на скорости 70 миль в час по дороге, не убивая людей, сидящих внутри. Uber передал свои прототипы беспилотных автомобилей другому стартапу, Lyft продала свое подразделение беспилотных автомобилей Toyota. Проблема, по-видимому, заключается в переходе от технологии, которая почти существует, к технологии, которую регулирующие органы считают достаточно безопасной, чтобы позволить водителю отвлечься и просматривать TikTok от А до Б без вмешательства в управление.

Какое это имеет значение с военной точки зрения? Некоторые проблемы с военным искусственным интеллектом сложнее, чем заставить Tesla перемещаться по сетке улиц Лос-Анджелеса, не путая красный свет с зеленым светом или не свернув направо через велосипедиста или пешехода. Любая автономная система военной логистики будет иметь дело со специальными складами и базами, внедорожными сценариями и дорогами, которые заставят мозг Tesla отключиться. Но некоторые проблемы с военным искусственным интеллектом, возможно, проще, чем беспилотные автомобили. Достижения в области компьютерного зрения/датчиков и обработки информации упрощают обнаружение действительно быстро приближающихся объектов в форме ракет, особенно если вы находитесь в море и вам не нужно беспокоиться о случайных объектах, загромождающих обзор. Кроме того, решить, как система противоракетной защиты должна реагировать на приближающуюся ракету, несколько проще, чем бесконечная проблематизация беспилотных автомобилей («Должен ли автомобиль врезаться в стену, чтобы не убить пять человек?», «Что, если там ребенок в машине, а пятеро человек действительно стары и все равно вот-вот умрут?» и так далее).

В то же время искусственное принятие решений заменило человеческое принятие решений в некоторых областях. Рассмотрим две области, в которых принятие решений человеком в значительной степени передано компьютерным системам, по крайней мере, с точки зрения соотношения человеческих решений и автономных решений: высокочастотная торговля (HFT) и онлайн-реклама.

Высокочастотная торговля работает, потому что компьютерные системы могут превосходить людей по скорости. Как только передовые рыночные движения становятся ареной для алгоритмической конкуренции, людям нужно (вежливо) убираться с дороги, потому что они собираются потерять свою рубашку и рубашки своих инвесторов. Но когда речь идет о медленных решениях, например о том, стоит ли инвестировать в бизнес или сектор, то аспект скорости не так важен. Более того, риски, присущие HFT (Высокочастотный трейдинг, высокочастотная торговля ( high-frequency trading, HFT)), как продемонстрировал «внезапный крах» 2010 года, когда несколько компьютерных систем, торгующих друг с другом, сумели обрушить рынок без видимой причины, остаются. Алгоритмы могут превзойти людей в небольшом сегменте рынка, и они могут улучшить человеческие качества, но они не заменяют людей-трейдеров, несмотря на прогнозы.

Аналогичным образом, рекламные рынки, которые определяют, какую рекламу вы показываете на 99,999% страниц, которые вы просматриваете в Интернете в наши дни, не требуют принятия решений человеком. Кто-то, покупающий рекламное место, примет решение, затем произойдет компьютерная магия, и его реклама попадет в глаза. По крайней мере, теоретически. Это потому, что очень много интернет-рекламы кажется недоказуемой игрой, отделяющей как продавцов рекламы, так и покупателей от их денег. Все, кто продает услуги в этой многоуровневой афере со змеиным маслом, заявляют, что их продукт поможет рекламодателю получить доступ к нужному набору глаз, но, похоже, практически нет доказательств того, что это действительно так.

Учитывая потенциальную прибыль, которую можно получить от безопасных автономных транспортных средств, финансового ИИ и маркетинга, не следует ли нам ожидать, что это будут одни из самых передовых областей самого ИИ? ИИ действительно преуспевает в основном в контролируемых средах (фабриках) или в средах, которые можно сделать управляемыми (например, операции по добыче полезных ископаемых).

Так что подумайте: если три различных области искусственного интеллекта, где он должен взаимодействовать с беспорядком человеческого общества, кажутся существенными недостатками, которые в конечном итоге ограничивают его полезность (и потенциал прибыли), почему мы должны ожидать, что военный ИИ будет чем-то другим?

## II: Некоторые догадки о военном ИИ

Вот несколько догадок о том, как, на мой взгляд, проявится военный ИИ. Я использую здесь догадки, потому что считаю, что догадки – лучшая валюта, чем прогнозы. Прогнозы стоят по десять центов и всегда подлежат пересмотру задним числом («Ну, я сказал, что ИИ собирался преобразовать войну, но если вы посмотрите на это с другой точки зрения, то уборка операционной комнаты с помощью «Roomba» всё преобразит). Предчувствия похожи на предсказания, но без внешнего вида профессиональных знаний. У каждого может быть предчувствие. Предчувствия часто более описывают лежащее в их основе мышление, чем, так сказать, конечный продукт. Как и предсказания, догадки практически не требуют поддержки, но, говоря простым языком, они гораздо более открыты в отношении этого факта.

*Догадка 1: для ведения войны требуются социальные процессы, ИИ сможет оптимизировать некоторые из них.*

Нам нравится думать о себе, что мы живем в постиндустриальную эпоху, но рассказываем это кому-то, кто гнёт спину в центре выполнения заказов Amazon.

Промышленная революция характеризовалась переходом от ремесленного производства к фабричному, где рабочие были организованы в соответствии с ритмом промышленного оборудования. Фордизм и тейлоризм привнесли в эти условия управление временем, ориентированным на конкретные задачи, постоянно регулируя жизнь рабочих. Теперь ИИ произвел революцию нео-тейлористов в оптимизации рабочих мест, которая выходит за рамки традиционных складских и заводских настроек, чтобы потенциально оптимизировать жизнь белых воротничков информационных работников в сфере экономики услуг.

Какое отношение имеет производительность труда к военному ИИ? Я думаю, что справедливый взгляд на военный ИИ, или, по крайней мере, больше, чем просто оружейный взгляд на ИИ, может заставить вас поверить в это. Если в организации есть ниши, в которых технологии могут радикально повысить производительность труда или полностью устранить потребность в оплачиваемом труде, под давлением конкуренции трудно сопротивляться переменам. Возможно, вам нравится изготавливать булавки вручную, но в долгосрочной перспективе вы никогда не превзойдете фабрику. Точно так же вам может понравиться визуальное сканирование спутниковых каналов на предмет потенциальных целей, но если канал вашего оппонента предварительно аннотирован с потенциальными целями, тогда вы находитесь в явном невыгодном положении в конфликте между равными.

Интегрированные системы ИИ в организационные процессы изменяет их там, где это возможно (точно так же, как интеграция настольных компьютеров, пишущих машинок или новых производственных процессов меняет организацию). Если у вас есть завод, на котором вы можете представить компьютерную модель требуемой продукции, а сама фабрика оптимизирует инструменты и производственные линии, чтобы сделать это, то вы получите преимущество перед конкурентами, если будете конкурировать с точки зрения доставки новых товаров на рынок. Если, с другой стороны, вы допустили ошибку в модели, то ошибки модели теперь, вероятно, будут намного дороже, так как у вас будет меньше времени на их выявление и исправление, прежде чем производить (дорогие) инструменты для их массового производства. Результат: меньше людей работает с инструментами, больше людей занимается контролем качества моделей. На мой взгляд, то же самое и с процессами таргетинга. Если части «цепочки убийств» автоматизируются с помощью ИИ, это увеличивает риски ошибочного решения человека (или машины). Результат: изменение формы военных организаций с учетом возможных оптимизаций, предлагаемых ИИ, и минимизация рисков ошибок.

Вкратце, это моя основная догадка об искусственном интеллекте: он вызовет изменения в организационной практике, если она будет принята, и, вероятнее всего, будет принята из-за давления со стороны конкурентов (или предполагаемого давления со стороны конкурентов). Некоторых это устраивает, а кого-то определенно не устраивает. Я несколько агностик в этом вопросе: если отправной точкой является социальный процесс, направленный на убийство людей, то до тех пор, пока люди все еще находятся на вершине цепочки принятия решений, не так много нормативных различий между силами, зависящими от ИИ и силами, зависящая от компьютеров. Точно так же, как нет большой нормативной разницы между промышленной военной машиной середины Второй мировой войны и её предшественниками.

*Догадка 2: военный ИИ автоматизирует некоторые боевые действия, но, вероятно, не большинство из них.*

С этой точки зрения искусственный интеллект - это, по сути, автоматизация. Мы берем то, что требует человеческого познания и действий, воплощаем это в физическую систему, а затем то, что раньше требовало человеческого участия, больше этого не требует. «Это не ИИ». Я слышал, вы это скажете, ну, в ответ, подумайте, сколько автоматических вещей когда-то были автономными. Ракеты типа "запустил и забыл" превратились из автономных систем в просто автоматическую функцию системы. Автоматизированные системы распознавания целей выполняют эквивалентную когнитивную работу (распознавание объектов на основе чувственных данных) людьми. Просто они могут разбираться в самых разных типах данных и делать это быстрее, чем мы, что позволяет осуществлять действия, выходящие за рамки человеческих возможностей.

На мой взгляд, распознавание объектов - это ключевая область, в которой ИИ в конечном итоге превзойдет нас. По крайней мере, для больших узнаваемых деталей милитаристского комплекта. В этом заключается асимметрия – большие части узнаваемого военного снаряжения будут уязвимы для распознавания автономными системами, в то время как различение людей как военных или гражданских лиц будет трудным, если не невозможным.

Я считаю это жестким ограничением: большинство концепций, которые мы используем для понимания войны, слишком расплывчаты, чтобы машины могли эффективно автоматизировать их. Попытка использовать ИИ для достижения стратегических целей неизбежно приведет к результатам, сравнимым с максимизатором скрепки, когда моделирование неизбежно неопределенного набора желаемых целей приводит к нежелательным оптимальным решениям. Точно так же такие вещи, как «военный» или «гражданское лицо», скорее всего, никогда не поддадутся машинам - по крайней мере, в том смысле, в каком к ним приближаются люди.

С этой точки зрения преобразующее влияние ИИ на войну будет следствием того, как государства и негосударственные субъекты реагируют на изменение способности машин автоматически распознавать военные объекты. Здесь я имею в виду военные объекты в физическом смысле, такие как танки, самолеты, корабли и т. Д., А не то, какие части башни считаются военным объектом из-за их использования противостоящими силами.

Как следует из догадки 1, это оказывает такое же искажающее влияние на саму войну. Лица и объекты, которые могут быть военными объектами по определению любого юриста, потребуют GOFHT (Good Old Fashioned Human Targeting, старомодное нацеливание людьми), в то время как ряд ключевых военных платформ и объектов подлежит автоматическому уничтожению. Меньше «Сири, убей того парня» и больше «Сири, уничтожь все вражеские MBTs (Основной боевой танк (ОБТ) (англ. Main battle tank (MBT)) в этом квадрате сетки, но помни о церкви». С этой точки зрения, очень важны идеи - какие виды автоматизированного уничтожения допустимы для данной военной силы (и неотъемлемый риск жертв среди гражданского населения / сопутствующего ущерба), будут определять, где и когда будут использоваться системы, позволяющие автономно уничтожать узнаваемые части. военного комплекта.

*Догадка 3: военный ИИ создаст некоторые неизбежные проблемы со смертельным исходом / уязвимостью для узнаваемых военных платформ.*

Как следует из вышесказанного, автоматическое распознавание с помощью искусственного интеллекта сделает некоторые системы более уязвимыми для атак. Это также поможет сделать некоторые системы более живучими. Хорошим примером этого являются системы активной защиты, которые мы сейчас видим для защиты транспортных средств от приближающихся ракет.

Я считаю, что это большая и неизбежная проблема. Я не думаю, что большинство возражений против ЗАКОНОДАТЕЛЬСТВА действительно работают против ударов по большим машинно-узнаваемым предметам экипировки, где ЗАКОНЫ имеют ограничения в виде геозоны (для предотвращения их работы в городских районах и т. д.) И других ограничений, которые являются также программируемый (атакует только объекты, которые не находятся рядом с людьми, после сканирования областей на присутствие людей и т. д.).

Мы уже начинаем понимать, что дроны могут сделать с крупной военной техникой без противовоздушной обороны (см. Стрельбу турков в Идлибе (https://www.oryxspioenkop.com/2020/02/the-idlib-turkey-shoot-destruction-and.html) и Нагорном Карабах в 2020 году (https://www.oryxspioenkop.com/2020/09/the-fight-for-nagorno-karabakh.html)). Автономные системы в виде более совершенных барражирующих боеприпасов (и подобных устройств) усугубят это.

Это приводит к странной модели асимметрии уязвимостей. Военные объекты, которые машинам нелегко распознать как военные объекты (например, грузовик Toyota с крупнокалиберным пулеметом в кузове), будут менее уязвимы для этих систем, чем легко различимые объекты (танки, большие артиллерийские орудия и т. д.).

Я, кстати, не предсказываю гибель танка, а только то, что для того, чтобы выжить, машинно-узнаваемым предметам снаряжения понадобится защитный пузырь, который нарушает ЗАКОНЫ, по которым действуют барражирующие боеприпасы, какую бы форму они ни принимали. Я предполагаю, что создание и поддержание такого пузыря будет дорогостоящим, поэтому эта догадка особенно применима к средним державам, которые могут оказаться не в состоянии себе это позволить.

## III: Закрытие разрыва

Не зря большие государства и великие державы беспокоятся о военных возможностях своих соперников. Поражение в войне в Афганистане может быть невероятно пагубной потерей для США, но проигрыш в перестрелке в Южно-Китайском море, вероятно, будет иметь более серьезные последствия. Поэтому естественно отдать приоритет анализу разрыва в возможностях между аналогами: сократит ли технология X разрыв между Китаем и США? Или технология Y позволит США снова компенсировать успехи своих конкурентов?

Многие военные технологии распространяются через международную систему, и часто многие перехватываются негосударственными субъектами и субгосударственными конкурентами. Как отметили Андреа Гилли и Мауро Гилли, имитация затруднительна, но менее качественные системы могут быть воспроизведены при наличии времени и усилий. Умные бомбы войны в Персидском заливе - лучше всего объясняемые тем, что Майкл Горовиц и Джошуа Шварц сделали «комплексом высокоточного удара» - теперь все чаще доступны как государствам, так и вооруженным группировкам за пределами НАТО. Вероятно, возможности Ирана по высокоточным ударам не так хороши, как у Америки, и, возможно, у них нет возможности (теоретически) поразить конкретное место на земном шаре в короткие сроки. У них все еще есть возможность наносить удары по американским базам, где это важно, и это, в конце концов, главное.

Один из способов взглянуть на военный ИИ – это как на нечто, увеличивающее разрыв между государствами, у которых он есть, и государствами, у которых его нет. Безусловно, именно так и кажется, что большинство людей заботится об этом в пространстве международной безопасности.

На мой взгляд, тот тип ИИ, о котором я говорю в этой статье, является скорее уравнивающей силой. Государства могут создавать всевозможные замечательные штуковины, которые на много миль опережают следующее государство-конкурент, но тот факт, что негосударственные вооруженные группы имеют доступ к рудиментарным формам ИИ, означает, что разрыв между организованными вооруженными силами государства и их негосударственными военными конкурентами увеличивается. меньше. Если, как указывалось выше, единственный способ сохранить важные платформы, которые эффективно отделяют государственные силы от негосударственных конкурентов (например, способность эффективно использовать основные боевые танки и артиллерию), становится дороже, тогда все меньше и меньше государств будут иметь возможность сделать это без значительного риска потери. Другими словами, мир искусственного интеллекта, в котором средние державы и повторные пользователи не модернизированного советского снаряжения подвергаются значительным рискам.

Одним из лучей в туннеле является тот факт, что негосударственному субъекту будет довольно сложнее создать свою собственную систему, напоминающую барражирующие боеприпасы, так, как вооруженные группы разработали свои собственные беспилотные летательные аппараты или элементарные средства высокоточного нанесения ударов. Тем не менее, рудиментарные системы искусственного интеллекта также могут бороться с защитными средствами, применяемыми великими державами.

Тем не менее, я не думаю, что мы должны сбрасывать со счетов возможность использования вооруженными группами высокотехнологичных систем в тех же самых высоких технологиях, которые мы наблюдаем в настоящее время в современных конфликтах (и, возможно, это неотъемлемая часть военной помощи на протяжении всей холодной войны). Если рудиментарные системы работают против большинства государственных вооруженных сил или требуют чрезмерно дорогостоящего оборудования для защиты от них, тогда любая вооруженная группа, которая может договориться о внешней поддержке / поставках, будет хорошо подготовлена, чтобы притупить силы государства. В этом смысле будущее войны - это не рои роботов-убийц, это бывшие советские танки и артиллерия, поджаренные чрезвычайно умной ракетой, запущенной из кузова джипа Toyota.

Вторая проблема заключается также в том, что многие вооруженные силы могут более легко допускать неточное уничтожение как людей, так и объектов. Если системы, скомпонованные из коммерческих компонентов, способны распознавать людей или транспортные средства (вероятно, по крайней мере, как обстоят дела), то боеприпасы смертников, нацеленные на скопления людей или любой узнаваемый автомобиль, имеют смысл в качестве средства атаки на базы. Если законен обстрел военной позиции неуправляемой артиллерией, то отправка чего-либо для выборочного уничтожения людей, присутствующих в этом месте, кажется вполне вероятным будущим. Дальнейший вариант - эти системы будут использоваться для нацеливания и уничтожения мирных жителей в преднамеренных неизбирательных атаках - также вполне возможен, но тогда он также возможен с использованием практически любой из существующих систем оружия.

## IV: Война в городах

Городские войны в настоящее время являются горячей темой. Об этом говорит МККК (International Committee of the Red Cross), у Института современной войны есть проект, посвященный этому, Энтони Кинг только что опубликовал об этом книгу. Масштабы разрушений, нанесенных Сирии и Ираку, служат непосредственным напоминанием о произошедшей бойне.

Я думаю, что даже в случае провала ИИ одним из долгосрочных последствий более широкого использования ИИ является побуждение к войне в городах. Это происходит по той простой причине, что любой противник, использующий автономные системы, лучше всего обслуживается «беспорядком», который мешает его действиям. Это не единственная долгосрочная причина конфликтов в городских районах и внутри них. Города исторически были объектом нападений и завоеваний, и 21 век не стал исключением. Более того, человечество теперь является городским видом - большинство из нас сейчас живет в городских или пригородных пространствах – таким образом, большинство приобретений от победы в конфликтах находятся в городах, и вероятно, в большей степени, чем раньше.

Если мои рассуждения, изложенные в приведенных выше разделах, верны – что основным следствием военного ИИ будут некоторые достижения в области компьютерного зрения, но не более того; что это позволит негосударственным субъектам производить достаточно хорошие вещи, которые делают военное снаряжение уязвимым – тогда любому, кто хочет использовать большие части военного снаряжения, понадобятся либо дорогие защитные системы, либо беспорядок, чтобы поддерживать их жизнь, либо сочетание того и другого.

Давайте проясним: беспорядок означает использование гражданских объектов и гражданских людей. Я оставлю юридический обзор кому-то другому, но это не столько активное экранирование (например, создание базы в многоквартирном жилом доме, использование живых щитов, совместное размещение с защищенными объектами, такими как больницы и медицинские учреждения), сколько пассивное экранирование в том смысле, что автономные системы, подобные барражирующим боеприпасам (какими бы продвинутыми они ни были), вряд ли будут достаточно надежными или предсказуемыми для полностью автономного использования в такой среде без значительных потерь среди гражданского населения. Одно дело, когда командир запускает систему, чтобы прочесать лес для уничтожения бронетехники, зная, что поблизости нет мирных жителей, и совсем другое - сделать то же самое, не зная, будут ли эти машины двигаться по проселочной дороге где нет никого вокруг или проезжать мимо школы в то время, когда система выбирает ее для уничтожения.

В этом смысле чистый эффект провала ИИ с военной точки зрения будет заключаться в разжигании конфликтов в городах. То, как государства и негосударственные субъекты действуют в такой среде, будет иметь решающее значение для определения чистых эффектов таких систем. Если города станут зонами без автономии, то в конечном итоге не будет существенного отличия от того, какой урон городская война наносит городам и их населению в наши дни. Однако, если одна или несколько сторон конфликта согласны с непредсказуемостью автономных систем в этих средах (на мой взгляд, это нехорошо, но вполне возможно), то последствиями будет еще больший хаос и, вероятно, больше жертв среди гражданского населения, чем у нас имеется в настоящее время. Учитывая, что некоторые государства и негосударственные субъекты не возражают против обстрела неуправляемой артиллерией и ракетами целей в городах, кажется вероятным, что они не против использования автономных систем в тех же местах.

## Заключение

ИИ не должен быть революционным, чтобы оказывать значительное влияние на ведение войны. Незначительное увеличение производительности одного типа системы делает другие системы более уязвимыми для разрушения и, следовательно, может значительно увеличить стоимость определенных форм войны. Хотя размышления о долгосрочных революционных последствиях всех типов прекрасных форм ИИ - это хорошо, я думаю, что нам следует уделять больше внимания тем вещам, которые кажутся достижимыми с использованием уже продемонстрировавших свою успешность технологий.

Мои технологические прогнозы здесь довольно ограничены по конструкции: небольшие продвижения в области компьютерного зрения, упрощённые версии передовых оружейных технологий, разрабатываемые средними державами, коммерческие технологии распознавания объектов, которые могут быть внедрены в функциональные системы оружия негосударственными субъектами. Это мир хороших барражирующих боеприпасов, используемых негосударственными субъектами, а не неконтролируемых наземных транспортных средств без экипажа, координирующих свои атаки на основе команд более высокого уровня со стороны людей. На что похожа война, когда повстанец может просто запустить противопехотный боеприпас в зону действия FOB (передовая оперативная база) на холме, вместо того, чтобы обстреливать её неэффективным минометным огнем? С точки зрения государств и тех, кто защищает международный порядок, ориентированный на государство, это нехорошо.

What if Military AI is a Washout?

Military applications of artificial intelligence, we are told, are poised to transform military power. They might make the oceans transparent to sensor systems, [threatening at-sea nuclear deterrent systems like the UK’s Trident](https://warontherocks.com/2019/10/ai-risks-to-nuclear-deterrence-are-real/). They might enable autonomous aircraft that could [outfight human crewed planes](https://www.defenseone.com/technology/2020/08/ai-just-beat-human-f-16-pilot-dogfight-again/167872/). They could transform intelligence processing in war, enable all sorts of complex weapons that would make things like tanks and aircraft carriers yesterday’s news. The sky, it appears, is the limit.

In this light, big states are making large investments in military AI. One aspect of the UK’s recent [Integrated Review](https://www.gov.uk/government/publications/global-britain-in-a-competitive-age-the-integrated-review-of-security-defence-development-and-foreign-policy) (ahem, “Global Britain in a Competitive Age”) and [Command Paper](https://www.gov.uk/government/collections/integrated-review-ministry-of-defence) (ahem, “Defence in a competitive age”) is a bet that investment in military applications of artifical intelligence will offset cuts to things like tanks and troop numbers. For the UK, this is a big bet. It has consequences for the way in which the UK will contribute to NATO, and its key alliances such as the “special relationship” with the United States.

So what happens if military AI doesn’t pay off? In a more British sense, what happens if military AI turns out to be a bit pants? This post is my way of reasoning through my increased scepticism about an AI-driven “military revolution” in the near future. About five years ago I wrote about the possibility of [AI systems being used to help detect insurgent networks](https://jackmcdonald.org/book/2016/07/future-research/), but this post is a roll-back from even that (limited) capability.

The argument goes something like this: The socially transformative vision of AI sold by venture capital over the last decade and a bit looks like it is going to wash out as a few niche areas of tremendous improvement, but no self-driving taxi fleets in London. The integration of some AI technologies will enable automation/autonomy in parts of pre-existing military processes (e.g. kill chains), but no robot super-soldier, HAL 9000 strategists, and limited institutional change. This would still have a huge impact on warfare by rendering machine-recognisable objects vulnerable to automated destruction by any variety of autonomous systems. This visibility asymmetry will make it harder to project power and sustain military forces in the field and reduces the capability gap that state militaries seek to maintain relative to non-state actors. Rections to this asymmetry will drive warfare towards urban environments. In short: worry about marginal improvements to what has already been fielded, and ways to constrain proliferation of those platforms, because the future is now (to shamelessly quote [Non Phixion](https://en.wikipedia.org/wiki/The_Future_Is_Now)).

I: Selling Generalised AI is Easy, Making Money is Hard

We have been living through an era of big venture capital bets on AI in one form or another. Part of my scepticism over the last decade of AI hype has been driven by the dissonance between years of pitches, and what has matured into industry-defining technologies. A good rule of thumb in this regard is employment: it’s hard to reap pure profit (but not impossible, see Google/Facebook switching a large % of global advertising/marketing budgets to their systems) but reducing headcounts, salaries, and pensions is a pretty good way to make more money. In this regard, the stalking horse behind one of the leading venture capital bets - self-driving vehicles - has always been the promise that you might not have to pay for your taxi driver anymore (or that you might not have to factor a long distance trucker’s wages into the cost of your physical goods).

So, how’s that going?

Many of the big and public venture capital bets on AI-enabled self-driving cars appear to be unwinding before our eyes. Consider that Lyft, Uber, Tesla have all pitched self-driving cars for the past decade. While Tesla, at least, has made some huge inroads into the problem of directing a hunk of metal at 70mph along a road without killing the people sitting inside it. Uber has [offloaded its driverless car attempts to another startup](https://www.bbc.co.uk/news/business-55224462), Lyft has [sold its self-driving division to Toyota](https://www.reuters.com/business/autos-transportation/lyft-sells-self-driving-tech-unit-toyota-550-mln-moves-up-profit-timeline-2021-04-26/). The problem appears to lie in getting from a technology that is *almost* there, to a technology that regulators consider safe enough to let the driver put their feet up and browse TikTok from A to B with no intervention.

How does this matter in military terms? Some military AI problems are harder than getting a Tesla to navigate the street grid of Los Angeles without confusing a red light for a green light, or taking a right turn through a cyclist or pedestrian. Any kind of autonomous military logistics system will have to deal with ad hoc depots and bases, off road scenarios, and roads that would make a Tesla’s brain shut down. But some military AI problems are perhaps simpler than self-driving cars. Advances in computer vision/sensors and information processing makes picking out really fast incoming missile-shaped objects easier, particularly if you’re at sea and don’t have to worry about random objects cluttering the view. Also, deciding how a missile defence system should respond to an incoming missile is somewhat more straightforward than [the infinite trolley problematisation of self-driving cars](https://www.brookings.edu/research/the-folly-of-trolleys-ethical-challenges-and-autonomous-vehicles/) (“Should a car drive itself into a wall to avoid killing five people?”, “What if there is a baby in the car and the five people are really old and about to die anyway?” and so on).

At the same time, artificial decisionmaking has replaced human decisionmaking in some areas. Consider two domains in which human decision-making has largely been turned over to computer systems, at least in terms of the ratio of human decisions to autonomous ones: High-Frequency Trading (HFT), and online advertising.

High-frequency trading works because computer systems can outcompete human beings in terms of speed. Once front-running market movements becomes an arena for algorithmic competition, then humans need to (politely) get the fuck out of the way because they are going to lose their shirt, and their investors’ shirts. But when it comes to slow decisions, things like whether or not to invest in a business or sector, then the speed aspect is not as decisive. Moreover, the risks inherent in HFT, as demonstrated in the “flash crash” of 2010 where a bunch of computer systems trading with each other managed to tank the market for no apparent reason, remain. Algorithms might outperform humans in a small slice of the market, and [they can improve human performance](https://www.thetradenews.com/human-judgement-still-king-in-a-world-of-algorithmic-trades/), but they’re not replacing human traders, [despite predictions](https://phys.org/news/2011-09-algorithmic-humans-stock.html).

In a similar fashion, the advert markets that determine what adverts you get served on 99.999% of the pages that you view on the internet these days do not require human decisionmaking. Someone buying ad space will make a decision, then computer magic happens, and their ads hit the right eyeballs. In theory, at least. That’s because an awful lot of online advertising appears to be a [non-provable shell game](https://www.letstakeacloserlook.com/2020/10/12/the-shell-game/) that separates both advertising vendors and buyers from their money. Everyone selling services in this multi-level snake-oil scam claims their product will help an advertiser to get in front of the right set of eyeballs, but there appears to be little to no evidence that this is actually the case.

Given the kind of potential profit that could be reaped from safe autonomous vehicles, financial AI, and marketing, shouldn’t we expect these to be some of the most cutting edge areas of AI itself? Where AI truly appears to succeed is largely in controllable environments (factories), or in environments that can be rendered controllable (e.g. mining operations).

So consider this: if three distinct domains for artificial intelligence where is has to interact with the messiness of human society appear to have significant flaws that ultimately limit its utility (and profit potential), why should we expect military AI to be anything different?

II: Some Hunches About Military AI

Here are some hunches about how I think military AI plays out. I use hunches here because I think hunches are a better currency than predictions. Predictions are ten-a-penny and always subject to retroactive revision (“Well I said that AI was going to transform warfare but if you look at it *this* way then cleaning the operations room with a Roomba is transformative”). Hunches are like predictions but without the veneer of professional expertise. Everyone can have hunches. Hunches are often more descriptive of underlying thinking than they are of the end product, so to speak. Like predictions, hunches require little to no support, but in terms of plain language they are far more open about this fact.

Hunch 1: Waging war requires social processes, AI will be able to optimise some of them

We like to think of ourselves as living in a post-industrial era, but tell that to [someone breaking their body in an Amazon fulfilment centre](https://techcrunch.com/2020/09/29/serious-injuries-at-amazon-fulfillment-centers-topped-14000-despite-the-companys-safety-claims/).

The industrial revolution was characterised by a shift from craft production to factory settings, where workers were organised around the rhythm of industrial machinery. Fordism and Taylorism brought task-specific time management to these settings, steadily regimenting the lives of workers. Now AI has brought a neo-Taylorist revolution in workplace optimisation that goes beyond traditional warehouse and factory settings to potentially optimise the lives of white collar information workers in the service economy.

What does labour productivity have to do with military AI? I think a fair bit, or at least more than a weapon-centric view of military AI might lead you to believe. Once there are niches within an organisation where technology can either radically increase worker productivity, or supplant the need for paid labour entirely, it is difficult - under competitive pressure - to resist change. You might enjoy manufacturing pins by hand, but you’ll never outperform a factory in the long run. Similarly, you might enjoy visually scanning satellite feeds for potential targets, but if your opponent’s feed is pre-annotated with potential targets, then you are at a distinct disadvantage in a peer conflict.

Integrated AI systems into organisational processes distorts them where it is possible (just as integrating desktop computers, or typewriters, or new production processes changes an organisation). If you have a factory where you can present a computer model of a required output and the factory itself will optimise the tooling and production lines to make it, then you’ll get a jump on competitors if you compete in terms of taking novel items to market. If, on the other hand, you make an error in the model, then model errors are now likely much more expensive, as there would be less time to identify and rectify them before producing the (expensive) tools to mass produce them. Result: less people in tooling, more people in model quality assurance. In my view, the same goes for targeting processes. If bits of the “kill chain” get automated with AI, then it increases risks of prior incorrect human (or machine) judgement. Result: re-shaping military organisations to account for potential optimisations offered by AI, and to minimise risks of errors.

This, in a nutshell, is my underlying hunch about AI: It will force changes in organisational practices where adopted, and it is likely to be adopted due to competitive pressure (or perceived competitive pressure). Some people are okay with this, some people are definitely not okay with this. I am somewhat agnostic on the issue: if the starting point is a social process designed to kill people, then so long as humans are still at the top of the decision chain, there isn’t that much normative difference between an AI-reliant force and a computer-reliant force, just as there isn’t much normative difference between a mid-WW2 industrial war machine and their predecessors.

Hunch 2: Military AI will automate some bits of warfare, but probably not most of them.

In this view artificial intelligence is essentially automation. We take something that would require human cognition and action, instantiate it in a physical system, and then something that used to require a human being no longer requires a human being. “That is not AI” I hear you say, well, in response, consider how many automatic things were once autonomous things. Fire-and-forget missiles have gone from being discussed as autonomous systems to simply being an automatic function of a system. Automated Target Recognition systems are performing equivalent cognitive work (recognising objects from sense data) to human beings. It’s just that they can make sense of many different types of data, and do it faster than we can, enabling forms of action beyond human capabilities.

As I see it, object recognition is a key domain in which AI will eventually outperform us. At least for big recognisable pieces of kit. Therein lies the asymmetry - big pieces of recognisable military kit will be vulnerable to recognition by autonomous systems, whereas distinguishing human beings as being combatants or civilians is going to be hard, if not impossible to achieve.

I see this as a hard limit: most of the concepts we use to make sense of war are too nebulous for machines to efficiently automate. Attempting to use AI for strategic goals will inevitably lead outcomes akin to a [paperclip maximizer](https://www.lesswrong.com/tag/paperclip-maximizer) - where the modelling of a necessarily indeterminite set of desired goals leads to unwanted optimal solutions. Similarly, things like “combatant” or “civilian” are likely never going to be amenable to machines - at least in the sense that human beings approach them.

From this perspective AI’s transformative effect on warfare will be the consequence of how states and non-state actors respond to a variation in the ability of machines to automatically recognise military objects. Here I mean military objects in the physical sense, as in tanks, planes, ships, etc, rather than [which bits of a tower block count as a military object](http://opiniojuris.org/2021/05/19/the-rabbit-hole-from-antwerp-to-gaza/) due to its use by an opposing force.

Following from hunch 1 this has a similarly distorting effect on warfare itself. Persons and objects that might well be military objects by any lawyer’s definition will require GOFHT (Good Old Fashioned Human Targeting), whereas an assortment of key military platforms and objects will be liable to automated destruction. Less “Siri, kill that guy” and more “Siri, destroy any enemy MBTs in that grid square, but mind the Church”. In this view, ideas matter a lot - what kinds of automated destruction are tolerable to a given military force (and the inherent risk of civilian casualties/collateral damage) will determine where and when it will employ systems that enable the autonomous destruction of recognisable pieces of military kit.

Hunch 3: Military AI will create some unavoidable lethality/vulnerability issues for recognisable military platforms

Following from the above, automated recognition enabled by artificial intelligence will make some systems more vulnerable to attack. It will also help to make some systems more survivable. The kinds of [active protection systems](https://breakingdefense.com/2020/01/how-active-protection-systems-knock-down-anti-armor-threats-for-both-legacy-and-future-combat-vehicles/) that we now see protecting vehicles from incoming rockets are a good example of this.

I think this is a big and unavoidable problem. [I don’t think that most of the objections to LAWS really work against strikes on big machine-recognisable pieces of kit](https://jackmcdonald.org/book/2021/02/private-holmes-problem/) where LAWS are given guardrails in the form of geofencing (to prevent them operating in urban areas, etc) and other constraints that are also programmable (only attacking objects that aren’t near to humans after scanning areas for the presence of human beings, etc).

We’re already starting to see what drones can do to big pieces of military hardware without air defence (see [the Idlib turkey shoot](https://www.oryxspioenkop.com/2020/02/the-idlib-turkey-shoot-destruction-and.html) and [Nagorno-Karabakh in 2020](https://www.oryxspioenkop.com/2020/09/the-fight-for-nagorno-karabakh.html)). Autonomous systems in the form of more advanced loitering munitions (and like devices) will exacerbate this.

This leads to a weird pattern of vulnerability asymmetry. Military objects that are not easy for machines to distinguish as military objects (e.g. a Toyota truck with a heavy machine gun on the back) will be less vulnerable to these systems than things that are easily distinguishable (tanks, large artillery pieces, etc).

I am not predicting the death of the tank, by the way, only that in order to survive, machine-recognisable pieces of kit will need a protective bubble that defeats LAWS that function akin to loitering munitions, whatever form that takes. I’d imagine that such a bubble would be expensive to generate and maintain, so this hunch is particularly likely to apply to middle powers who might not be able to afford it.

III: Closing the Gap

For good reason, big states and great powers worry about the military capabilities of their peers. Losing a war in Afghanistan might be an incredibly harmful loss for the US to accept, but losing a shooting war in the South China Sea is likely to have greater consequences. It is natural, therefore, to prioritise the analysis of capability gaps between peers: Will X technology close the gap between China and the US? Or will Y technology allow the US to offset its competitors once again?

Many military technologies diffuse through the international system, and often many are picked up by non-state actors and sub-state competitors. Imitation is difficult, [as Andrea Gilli and Mauro Gilli have pointed out](https://direct.mit.edu/isec/article/43/3/141/12218/Why-China-Has-Not-Caught-Up-Yet-Military), but lesser quality systems can be emulated given time and effort. The smart bombs of the Gulf War - best explained as Michael Horowitz and Joshua Schwartz have done as the “[precision strike complex](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3752391)” - are now increasingly available to both states and armed groups beyond NATO. Maybe Iran’s precision strike capabilities aren’t as good as America’s, and maybe they don’t have the capability (in theory) to hit a specific spot on the globe at short notice. They still have the ability to hit American bases where it matters, and that, in the end, is what matters.

One way to look at military AI is as something that will widen gaps between states that have it, and states that don’t. Certainly this is the way that most people seem to care about it in the international security space.

The way I see it, the kind of AI that I am referring to in this article is more of a levelling force. States might build all sorts of wonderful gizmos that are miles ahead of the next competitor state, but the fact that non-state armed groups have access to rudimentary forms of AI means that the gap between organised state militaries and their non-state military competitors gets smaller. If, as pointed out above, the only way to keep important platforms that effectively differentiate state forces from non-state competitors (the ability to effectively field main battle tanks and artillery, for example) gets more expensive, then less and less states are going to be able to do so without significant risk of loss. A washout AI world, in other words, is one in which middle powers and re-users of non-upgraded Soviet gear run significant risks.

One light in the tunnel is the fact that it would be quite difficult for a non-state actor to create its own loitering-munition-like system, in the same way that armed groups have developed their own drones or rudimentary precision strike capabilities. Rudimentary AI systems might also struggle to defeat protective measures employed by great powers.

That said, I don’t think we should discount the possibility of armed groups fielding otherwise high end systems in the same hi-lo technology mixes that we currently see in contemporary conflicts (and, arguably, is an integral feature of military assistance throughout the cold war). If rudimentary systems work against most state militaries, or require inordinately expensive kit to protect against them, then any armed group that can negotiate external support/supplies will be well placed to blunt the forces of states. In this sense, the future of war isn’t swarms of kill bots, it is ex-Soviet tanks and artillery getting toasted by an extremely smart missile launched from the back of a Toyota.

A second issue is also that many military forces might be more open to imprecise destruction, both of persons and objects. If systems rigged together from commercial components are able to recognise people or vehicles (likely, at least the way things are going), then suicide munitions that target packed clusters of human beings or any recognisable vehicle make sense as a means of attacking bases. If it is lawful to shell a military position with unguided artillery, then sending something in to selectively kill the humans present in that location seems like a plausible future. The further option - that these systems will be used to target and kill civilians in intentionally indiscriminate attacks - is also a possibility, but then it is a possibility with pretty much any weapon system that exists.

IV: War in the Cities

Urban warfare is currently a hot topic. The ICRC is [talking about it](https://blogs.icrc.org/law-and-policy/2021/04/27/urban-warfare/), the Modern War Institute has [a project dedicated to it](https://mwi.usma.edu/urban-warfare-project/), Anthony King [just published a book on it](https://politybooks.com/bookdetail/?isbn=9781509543656). The scale of destruction wrought in Syria and Iraq serves as an immediate reminder of the carnage involved.

Even in the case of an AI washout, I think one of the long-term effects of increased AI use is to drive warfare to urban locations. This is for the simple reason that any opponent facing down autonomous systems is best served by “clutter” that impedes its use. This is not the only long-term driver of conflict towards and within urban locations. Cities have historically been the object of attack and conquest, and the 21st century is no different. Moreover humanity is now an urban species - most of us now live in urban or peri-urban spaces - thus most of the prizes of conflict are likely found in cities to a greater degree than previously.

If my reasoning outlined in the sections above holds - that the primary consequence of military AI will be some advances in computer vision, but not much else; that it will enable non-state actors to make good-enough stuff that renders military kit vulnerable - then anyone seeking to use big pieces of military kit will need either expensive protective systems or clutter to keep them alive, or a combination of both.

Let’s be clear: clutter means civilian objects, and civilian human beings. I’ll leave the legal review to someone else, but this isn’t so much active shielding (e.g. setting up shop in a civilian apartment block, using human shields, co-locating with protected objects like hospitals and healthcare facilities), but passive shielding in the sense that autonomous systems akin to loitering munitions (however advanced) are unlikely to be reliable or predictable enough to use in a fully-autonomous fashion in such an environment without causing significant civilian casualties. It is one thing for a commander to set a system off to scour a forest of armoured vehicles, safe in the knowledge that there are no nearby civilians, it is another to do the same not knowing if those vehicles will be driving along a backroad with nobody around, or driving past a school at the time the system selects it for destruction.

In this sense, the net effect of washout AI in military terms would be to drive conflict to cities. How states and non-state actors operate in those environments would be crucial to determining the net effects of such systems. If cities become no-autonomy zones, then the ultimate effect would be no significant difference to the kind of damage that urban warfare inflicts on cities and their populations in the present day. However if one or more sides to a conflict is okay with the unpredictability of autonomous systems in these environments (that’s not a good thing, in my view, but entirely possible) then the consequences would be greater chaos, and likely more civilian deaths than we currently see. Given that some states and non-state actors are okay with lobbing unguided artillery and rockets at targets in urban locations, it seems likely that they’d be okay with using autonomous systems in those same places.

V: Conclusion

AI doesn’t have to be revolutionary to have significant effects on the conduct of war. Marginal increases in the performance of one kind of system makes other systems more vulnerable to destruction, and may therefore significantly increase the cost of certain forms of warfare. While thinking about the long term revolutionary consequences of all types of wonderful forms of AI is all well and good, my hunch is that we should pay closer attention to the kind of stuff that seems achievable using already demonstrated technologies.

My technological predictions here are pretty limited by design: small further advances in computer vision, weaker versions of bleeding-edge weapon technologies being developed by middle powers, commercial object recognition technologies that can be bodged into functional weapon systems by non-state actors. This is a world of kinda-good loitering munitions used by non-state actors, rather than unsupervised uncrewed ground vehicles coordinating assaults on the basis of higher level commands by human beings. What does warfare look like when an insurgent can simply lob an anti-personnel loitering munition at the FOB on the hill, rather than pestering it with ineffective mortar fire? From the perspective of states, and those who defend a state-centric international order, it’s not good.

© Jack McDonald 2019