Законы робототехники

Постулаты Азимова

Три закона робототехники - свод обязательных правил, которые должен соблюдать искусственный ителлект (ИИ), чтобы не причинить вред человеку. Законы используются только в научной фантастике, но считается, что как только будет изобретен настоящий ИИ, в целях безопасности, он должен иметь аналоги этих законов.

Постулаты Азимова

Азимов признавался, что намеренно сделал законы двусмысленными, чтобы обеспечить больше конфликтов и неопределенностей для новых рассказов. То есть, он сам опровергал их эффективность, но и утверждал, что подобные нормы - это единственный способ сделать роботов безопасными для людей.

Как следствие этих законов, позже Азимов формулирует четвертый закон робототехники, и ставит его на самое первое место, то есть делает его нулевым.

Вопросы, которые нужно задать

- Есть ли ограничения на использования военных роботов?
- Кому принадлежат авторские права на созданные роботом произведения и кто ответит за его поступки, например ущерб, принесенный автопилотируемой машиной?
- В применение «искусственного интеллекта» в области финансов уже нужно вмешаться, или стоит подождать еще нескольких сбоев бирж?
- Проблема сильного искусственного интеллекта (ИИ, сравнимый с человеческим мозгом по уровню интеллекта) все-таки реальна и нам уже нужно контролировать его разработки, или разберемся как-нибудь потом?

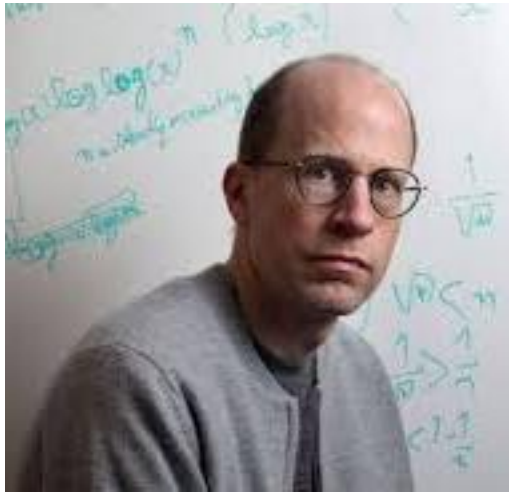
23 Азиломарских принципа искусственного интеллекта, разработанных и опубликованных в январе 2017 года под эгидой Future of Life Institute. Их подписало уже почти 4 000 экспертов и специалистов. Ученые предлагают направить свои усилия на создание управляемого, надежного и полезного ИИ, отказаться от «гонки вооружений» на основе ИИ и подумать о безопасности разработок, а также ответственности самих разработчиков.

Участники

Рей Курцвел

Стивен Хоккинг

Илон Маск



Ник Бостром

Демис Хассабис

Ян Лекун

"What does a good future look like?" asked Tesla CEO Elon Musk, speaking on a panel at the Beneficial AI conference in Asilomar, CA. "We're headed towards either superintelligence or civilization ending."

Научно-исследовательские вопросы

1. Целью исследований в области искусственного интеллекта должно быть создание не бесконтрольного разума, а выгодного для всего общества инструмента.

2. Инвестиции в ИИ должны производиться совместно с финансированием исследований по обеспечению полезности использования технологии, что позволит давать более точные и обобщенные ответы на возникающие вопросы в области информатики, экономики, права, этики и обществоведения, в частности:

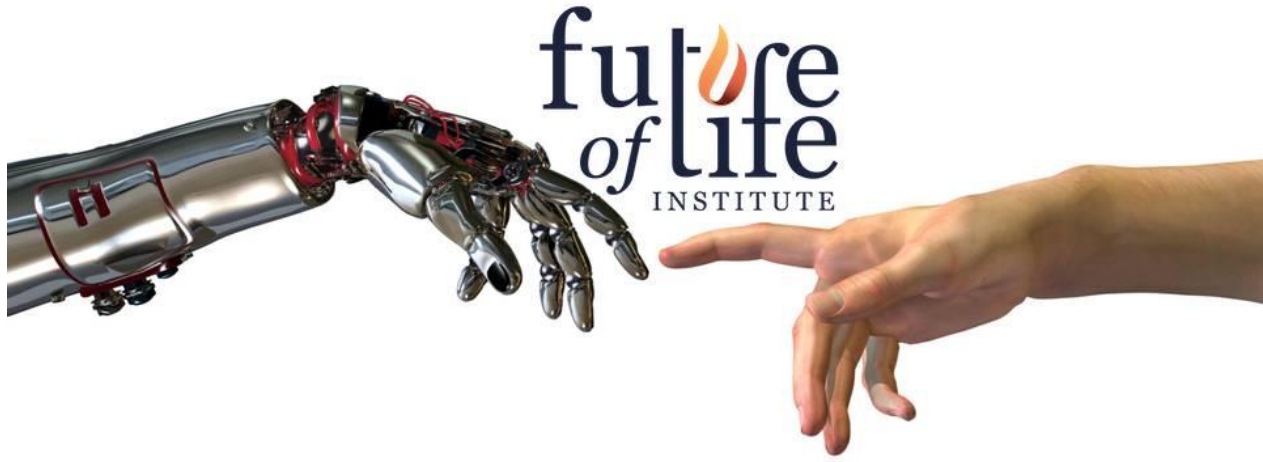
- Как обеспечить высокую надежность систем ИИ, чтобы они выполняли поставленные задачи без сбоев и угрозы быть взломанными?
- Каким образом при помощи автоматизации мы можем прийти к необходимой цели, поддерживая человеческие ресурсы и заинтересованность?
- Как грамотно обновить правовую систему, обеспечив справедливость по современным канонам и учтя появление ИИ?
- Какой правовой и этический статус получит ИИ?

Научно-исследовательские вопросы

3. Обеспечение конструктивного и равного взаимодействия исследователей и политиков.

4. В обществе исследователей должна соблюдаться полная открытость и прозрачность.

5. Исследователи ИИ должны всячески сотрудничать между собой во избежание нарушения стандартов безопасности.



Нравственные вопросы

6. Системы искусственного интеллекта должны быть безопасными и надежными на протяжении всего эксплуатационного срока.

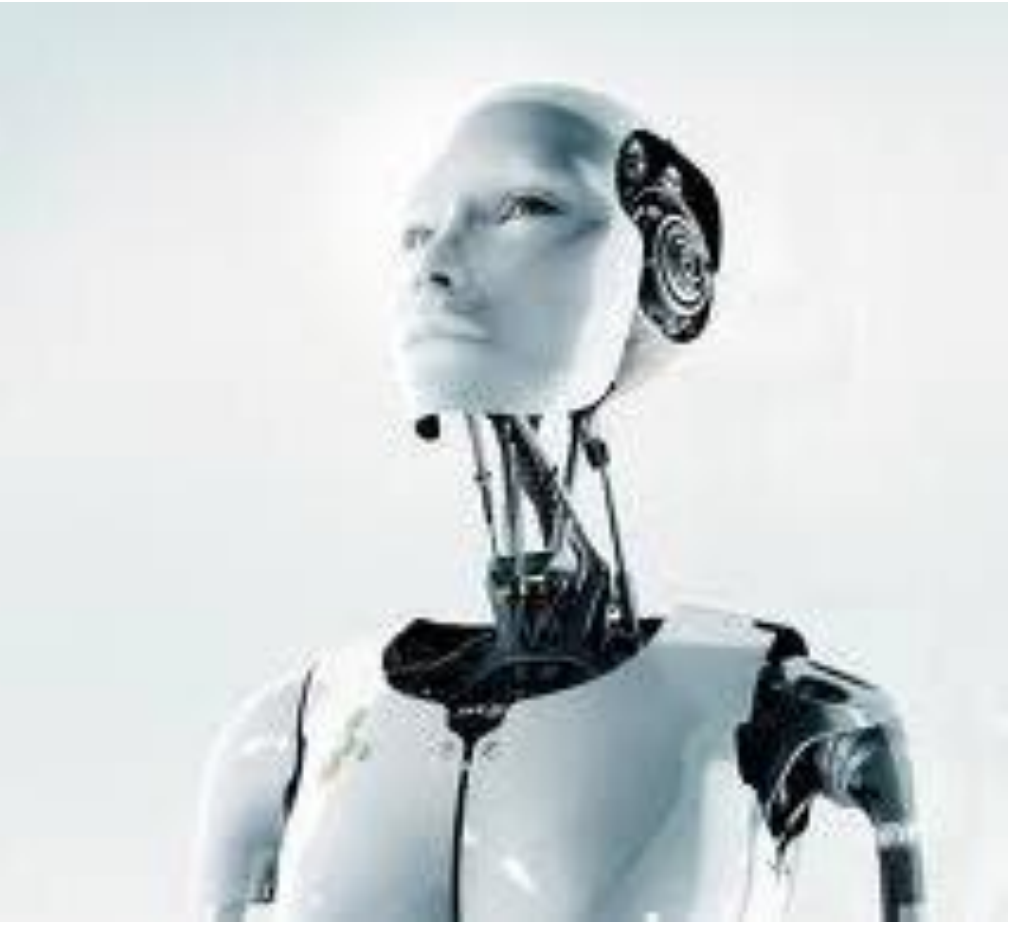
7. Если система ИИ причинит вред, должна быть возможность выяснения причины.

8. Любое вмешательство автономной системы в судебные решения должно быть санкционированным и подлежать отчетности в случае проверки специалистами.

9. Ответственность за причинённый вред, намеренное создание опасной ситуации, вызванное ошибкой конструкции или программного обеспечения ляжет на плечи разработчиков и инженеров.

10. Все автономные системы ИИ должны быть спроектированы таким образом, чтобы их цели и поведение не противоречили общечеловеческим ценностям.

Нравственные вопросы



11. ИИ должен быть спроектирован и использован так, чтобы быть совместимым с идеалами человеческих достоинства, прав, свобод и культурного разнообразия.

12. Данные, возникшие в процессе использования ИИ человеком, должны быть доступны в дальнейшем для этого человека.

13. При работе с персональными данными ИИ не должен необоснованно ограничивать реальные или мнимые свободы людей.

14. Технологии ИИ должны приносить пользу и быть доступными как можно большему количеству людей.

15. Экономические ценности, созданные ИИ, должны быть доступны всему человечеству.

Нравственные вопросы

16. Человек всегда должен иметь выбор: принять решение самостоятельно или поручить это ИИ.

17. Действия ИИ должны идти на пользу человеческой деятельности, а не во вред.

18. Необходимо всеми силами избегать гонки автономных видов вооружений.

Долгосрочные вопросы

19. Мы должны абсолютно трезво оценивать предел возможностей ИИ во избежание нерациональных затрат ресурсов.

20. Развитие ИИ будет интенсивным, поэтому технология должна быть под контролем, дабы не вызвать необратимых реакций.

21. Риски, особенно крупномасштабные, должны быть спланированы таким образом, чтобы обеспечивать минимизацию возможного ущерба.

22. Системы ИИ, способные к саморазвитию и самовоспроизведению, должны находиться под особо строгим контролем.

23. Суперразум должен будет служить всему человечеству, а не конкретной частной организации или государству.

Дополнительно

- https://www.forbes.ru/tehnologii/355757-zakony-robototehniki-kak-regulirovat-iskusstvennyy-intellekt
- https://futureoflife.org/bai-2017/
- https://www.techrepublic.com/article/23-principals-for-beneficial-ai-tech-leaders-establish-new-guidelines/
- http://sib.com.ua/sib-01-98-2018/iskustv_intelekt.html