

(3) Полето на изкуствения интелект има за цел да разработи компютърни програми, които да могат да изпълняват полезни задачи като отговаряне на въпроси, разпознаване на изображения и т.н. Началото е поставено около 50-те години на миналия век. В исторически план са съществували няколко различни подхода към изкуствения интелект.

(4) През първите няколко десетилетия доминиращата парадигма е **символният ИИ**, който се фокусира върху *представянето на проблеми с помощта на твърдения* във формални езици (като езици за програмиране) и търсенето на решения чрез манипулиране на тези представяния в съответствие с фиксирани правила. Например един символен ИИ може да представи игра на шах с помощта на набор от твърдения за това къде се намират фигурите в момента и набор от твърдения за това къде е позволено да се движат фигурите. След това той може да играе шах, като търси възможни ходове, които са съвместими с всички тези твърдения. Силата на ИИ, базиран на символно търсене, беше демонстрирана от **Deep Blue** - шахматният ИИ, който победи Каспаров през 1997 г.]]

Въпреки това символните представяния, разработени от изследователите на ИИ, **се оказаха твърде прости**: има много малко явления от реалния свят, които могат лесно да се опишат с помощта на формални езици. От 90-те години на миналия век насам доминиращата парадигма в ИИ е **машинното обучение**. При машинното обучение, вместо сами да кодираме ръчно всички детайли на ИИ, ние определяме модели със **свободни параметри**, които се учат от предоставените им данни. На теория статистическите модели (като линейна регресия) също подбират параметрите на данните, които са им предоставени. Двете области обаче се различават по мащабите, в които работят: най-големите успехи на машинното обучение са постигнати чрез обучение на модели с милиарди параметри върху огромни количества данни. Това се прави с помощта на **гълбоко обучение**, което включва обучение на невронни мрежи с много слоеве. Невронните мрежи съществуват от самото начало на изкуствения интелект, но те се превърнаха в доминираща парадигма едва в началото на 2010 г.

Важно е да се отметне, че символичното програмиране не е умряло. В последните години има значителни успехи в областта на невро-символичното програмиране. Даже в СУ имаше гост-лектор от САЩ на тази тема преди година-година и половина. В университетската му страница пише за приложение на неговите невросимволични програми:

Ускоряването на научните открития е сред най-вълнуващите обещания на ИИ. Въпреки това интерпретацията и ефективността на данните са ключови съображения, поради което подходите, основани

единствено на данни, са несъвършени. Ние подхождаме към тези предизвикателства, като моделираме научните хипотези като невросимволни програми, а научното откритие - като синтез на невросимволни програми.

(5) Цялата схема от предишния слайд може да се представи по следния начин, който най-вероятно ще срещате по-често. (Обясни схемата)

(6) Невронните мрежи са в центъра на дълбокото обучение и са модел, вдъхновен от мозъка. Както всички модели за машинно обучение, те приемат входни данни и произвеждат съответните изходни данни по начин, който зависи от стойностите на техните параметри. На диаграмата всяко кръгче представлява „неврон“. В наши дни почти всички невронни мрежи са дълбоки, а някои от тях имат стотици слоеве.

(6) Какво се случва тук? Всеки неврон получава сигнали от тези в предходния слой, комбинира ги в една стойност и след това предава тази стойност на невроните в следващия слой. Както и в биологичните мозъци, сигналят, който се предава между двойка изкуствени неврони, се влияе от силата на връзката между тях - така че за всяка от линиите в диаграмата трябва да съхраним едно число, представляващо силата на връзката, известно като **тегло**.

(6) Тези тегла не се задават ръчно, а вместо това се научават чрез процес на **оптимизация**, при който се намират тегла, които правят мрежата високо ефективна по използван от нас показател. Тази метрика е известна като функция на загубите, **loss function**. Ако стигнете изходния слой и тръгнете в обратната посока да оптимизирате, тогава правите обратно разпространение, **backpropagation**.

(7) Разгледахме невронните мрежи. Нека сега обърнем внимание на още три вида да създадете ИИ: **нагзиравано, ненагзиравано и обучение с подсилване**, като всеки от тях получава различни видове данни и целеви функции. Контролираното обучение изисква набор от данни, в който всяка точка от данни има съответен **етикет/маркер**. **Целта на контролираното обучение е моделът да предскаже етикетите**, които съответстват на всяка точка от данни.). Чрез този подход можете да класифицирате разни неща или да предсказвате стойност. Исторически погледнато, контролираното обучение е най-изследваната задача в машинното обучение и техниките, разработени за решаването ѝ, са широко използвани като част от решенията на другите двете.

(7) Един от недостатъците на **нагзираваното** обучение е, че етикетването на набор от данни обикновено трябва да се извършва ръчно

от хора, което е скъпо и отнема много време. На практика за да се избегне това се ползва **ненагдувано обучение**. Стандартен пример е предсказването на следваща дума от текст. Някои впечатляващи приложения са GPT-2 и GPT-3 за езици и Dall-E за изображения.

(7)И накрая, при **обучението с подсилване** източникът на данни не е фиксирана съвкупност от данни, а по-скоро **среда**, в която изкуственият интелект извършва действия и получава наблюдения - по същество все едно играе видеоигра. След всяко действие получава и **награда**, която се използва за подсилване на поведението, което пък води до високи награди. Тъй като действията могат да имат дълготрайни последици, основната трудност при обучението чрез подсилване е да се определи **кои действия са отговорни за кои награди**. Досега най-впечатляващите демонстрации на този подход са били в обучението на агенти за игра на настолни игри - най-вече AlphaGo, AlphaStar (за Starcraft) и OpenAI Five (за Dota 2).

(13) За да пристъпим към темата за работните места, трябва да видим с какво този канديدат е **по-добър**. Ще го наемаме ли или няма да го наемаме!?

(14) По-долу изброяваме няколко предимства на изкуствения интелект, които могат да позволят на изкуствения интелект да стане не само значително по-интелигентни от всеки човек.

- 1) - Човешкият мозък има 85-100 милиарда неврона.
 - наложено от еволюционни ограничения и метаболизма
 - размерът на мозъка и коефициентът на интелигентност при хората корелират положително с коефициент на корелация около 0,35 (McDaniel 2005).
 - Алгоритмите могат да се променят по всяко време...
- 2) - Аксоните пренасят сигналите на шиповете със скорост 75 метра в секунда или по-малко (Kandel, Schwartz и Jessell 2000).
 - фиксирана от нашата физиология
 - За разлика от тях софтуерните умове могат да бъдат пренесени на по-бърз хардуер
 - **Представете си, ако архитектурите преминат към фотонни невронни мрежи!**
- 3) - **създаването на допълнителни ИИ е просто въпрос на копиране на софтуер.**

- Тъй като уменията на даден ИИ се съхраняват в цифров вид, точното му текущо състояние може да бъде копирано, **включително спомените и придобитите умения (затова може да теглим тренирани модели от интернет!)**.

- Човек, който преминава през обучение, повишава само собствената си производителност, но ИИ, който става с 10 % по-добър в печеленето на пари (на долар хардуер под наем) от други ИИ, може да бъде използван за замяна на останалите в цялата хардуерна база, което прави *всяко копие* с 10 % по-ефективно.

4) Homo economicus -> Homo Simpson

- Homo economicus: себещцентрично рационално същество, което прави това, което вярва, че ще максимизира изпълнението на цели си.

- Schneider (2010) посочва, че ние приличаме повече на Хоумър Симпсън: ние сме ирационални същества които нямат последователни, стабилни цели

Всички тези цели подсказват, че ИИ ще има по-голям когнитивен капацитет. Въпросът за мотивацията остава неизяснен (ще иска ли да го направи?).

(15) /Дискутиране на целите/.

- изключителната възможност да общувате с умен, рационален събеседник, ако вече ви е писнало от хората

- Съгласни ли сте, че интелигентността може да се използва за постигане на всяка цел? Както казва Bostrom (2012) : *повишаването на интелигентността на изкуствените интелекти няма да ги накара да променят целта си*
Всъщност ИИ ще бъдат мотивирани да запазят първоначалните си цели.

-Така стигаме до основната характеристика на риска, свързан с ИИ: **Освен ако ИИ не е специално програмиран да запази това, което хората ценят, той може да разруши тези ценни структури.**

- Както се изразява Югковски (2008), „ИИ не ви обича, нито ви мрази, но вие сте направени от атоми, които той може да използва за нещо друго“.

Какви задачи обхващат по-голямата част от работното ви време? Обърнете внимание на различните категории. Колко добре се справят системите с изкуствен интелект в момента с тези задачи?

1. А сега си представете, че е 2034 г. Въз основа на темпа на подобрение, демонстриран в ресурсите, как биха могли да изглеждат системите с ИИ след още 10 години? Колко по-добри са нашите системи за ИИ, които могат да възвеждат и възвеждат изображения, видеоклипове, аудио и текстови данни (включително код)? Какви дълги последователности от действия могат да извършват чрез интернет - а може би дори и в реалния свят?
2. Каква част от ежедневната ви работа очаквате, че сега ще бъде възможна за системите с изкуствен интелект? Каква част от работата на вашите приятели и семейство ще бъде засегната?
3. Как мислите, че това ще се отрази на вашия живот по отношение на работата, която вършите? Ако си намерите друга работа, какви професии според вас биха могли да съществуват? Ако не, с какво бихте се занимавали вместо това?