

Inteligência Artificial

LuÃ¡ns A. Alexandre

UBI

Ano lectivo 2019-20

LuÃ¡ns A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

1 / 40

ConteÃ¡do

Robótica

Introdução

Localização e mapeamento

Aplicações da robótica

Questões filosóficas na IA

IA Fraca

IA Forte

Ética e o risco da IA

Leitura recomendada

LuÃ¡ns A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

2 / 40

Conteúdo

Robótica

Introdução

Localização e mapeamento

Aplicações da robótica

Questões filosóficas na IA

IA Fraca

IA Forte

Ética e o risco da IA

Leitura recomendada

LuÃ¡ns A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

3 / 40

Introdução

Robótica

Introdução

▶ Chegámos ao ponto em que os nossos agentes passam a poder atuar diretamente sobre o mundo físico.

▶ Os robôs são máquinas que se podem caracterizar por possuírem dois tipos de componentes:

- ▶ **sensores**, que recolhem informação do ambiente;
- ▶ **atuadores**, que alteram de alguma forma o ambiente (ou a posição do robô no ambiente).

▶ Exemplos de sensores: câmaras, microfones, acelerómetros, giroscópios.

▶ Exemplos de atuadores: motores, braços, pernas, pinças, altifalantes.

LuÃ¡ns A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

4 / 40

Tipos de robô

Robótica

Introdução

▶ Podemos dizer que existem 3 tipos de robôs:

▶ **manipuladores**: tipicamente são braços robotizados, fixos num determinado local. São os robôs mais comuns.

▶ **robôs móveis**: possuem pernas, rodas ou “patas” se forem terrestres, mas podemos ter veículos aéreos ou aquáticos também.

▶ **manipuladores móveis**: combinam as características dos 2 tipos anteriores. Os robôs humanóides estão neste grupo.

LuÃ¡ns A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

5 / 40

Perceção

Robótica

Introdução

▶ A **perceção** é o processo de mapeamento das medições feitas pelos sensores para representações internas do robô.

▶ Dizemos que \mathbf{X}_t é o **estado** do robô e do ambiente no instante t , \mathbf{Z}_t é a **observação** recebida no instante t e A_t é a **ação** efetuada depois de recebida a observação.

▶ Precisamos de **estimar** o **novo estado** no instante $t + 1$ à custa do estado atual e da informação recolhida no instante seguinte, $t + 1$:
$$P(\mathbf{X}_{t+1}|\mathbf{z}_{1:t+1}, a_{1:t}) = \alpha P(\mathbf{z}_{t+1}|\mathbf{X}_{t+1}) \int P(\mathbf{X}_{t+1}|\mathbf{x}_t, a_t)P(\mathbf{x}_t|\mathbf{z}_{1:t}, a_{1:t-1})d\mathbf{x}_t \quad (1)$$

▶ Esta equação diz-nos que a **probabilidade a posteriori da variável de estado no instante $t + 1$** é obtida de forma recursiva a partir da **estimativa feita na iteração anterior**.

▶ $P(\mathbf{X}_{t+1}|\mathbf{x}_t, a_t)$ é o **modelo de transição** ou **modelo de movimento**.

▶ $P(\mathbf{z}_{t+1}|\mathbf{X}_{t+1})$ é o **modelo do sensor**.

LuÃ¡ns A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

6 / 40

Robótica

Localização e mapeamento

Conteúdo

Robótica

Introdução

Localização e mapeamento

Aplicações da robótica

Questões filosóficas na IA

IA Fraca

IA Forte

Ética e o risco da IA

Leitura recomendada

LuÃ¶ns A. Alexandre (UBI)

InteligÃ¢ncia Artificial

Ano lectivo 2019-20

7 / 40

Robótica

Localização e mapeamento

Localização

► **Localização** é o problema de saber onde se encontram objetos, incluindo o próprio robô.

► Isto é essencial para fazer interação com o mundo físico.

► Os manipuladores têm que saber onde se encontram os objetos que pretendem manipular e os robôs móveis têm de saber onde estão para poderem navegar entre dois pontos.

► Vamos considerar um robô que se move num mundo 2D do qual possui um mapa.

LuÃ¶ns A. Alexandre (UBI)

InteligÃ¢ncia Artificial

Ano lectivo 2019-20

8 / 40

Robótica

Localização e mapeamento

Localização

► A forma mais comum de obter a localização de um robô é usando uma abordagem baseada em **filtros de partículas**.

► Para usarmos esta abordagem precisamos de fornecer um **modelo de movimento do robot** e um **modelo do sensor**.

► Cada partícula representa uma potencial localização do robô.

LuÃ¶ns A. Alexandre (UBI)

InteligÃ¢ncia Artificial

Ano lectivo 2019-20

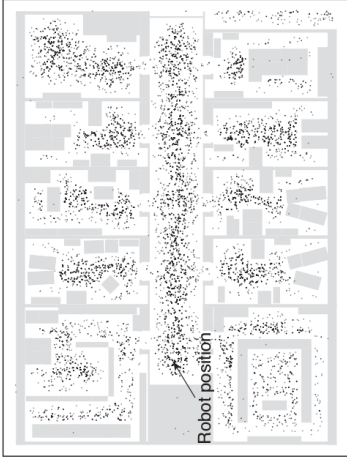
9 / 40

Robótica

Localização e mapeamento

Localização

Inicialmente as partículas encontram-se espalhadas uniformemente pelo mapa pois existe uma incerteza completa relativa à posição do robô (ainda não foram feitas observações).



Robot position

LuÃ¶ns A. Alexandre (UBI)

InteligÃ¢ncia Artificial

Ano lectivo 2019-20

10 / 40

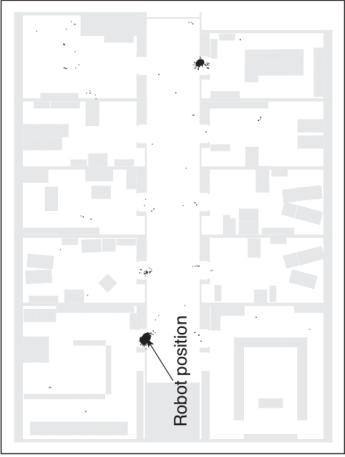
Figura de Russell & Norvig

Robótica

Localização e mapeamento

Localização

As partículas vão refletindo o conhecimento adquirido ao longo do tempo via sensores, ajustando as suas posições para os locais com maior probabilidade de serem aqueles em que o robô se encontra.



Robot position

LuÃ¶ns A. Alexandre (UBI)

InteligÃ¢ncia Artificial

Ano lectivo 2019-20

11 / 40

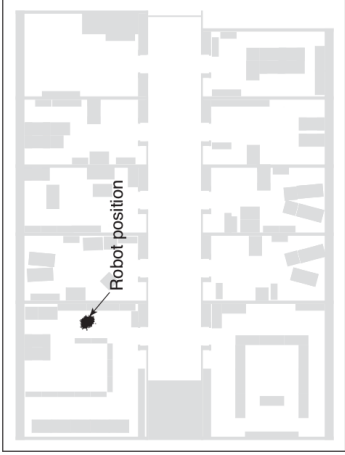
Figura de Russell & Norvig

Robótica

Localização e mapeamento

Localização

Após algum tempo de navegação, poucas posições do mapa estão de acordo com a informação recolhida pelos sensores: será nessas posições que as partículas se agrupam.



Robot position

LuÃ¶ns A. Alexandre (UBI)

InteligÃ¢ncia Artificial

Ano lectivo 2019-20

12 / 40

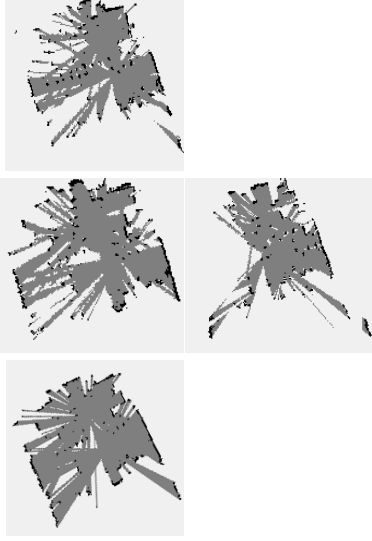
Figura de Russell & Norvig

Mapeamento

- ▶ Muitas vezes os robôs têm que se deslocar em locais desconhecidos e precisam de criar um mapa do local conforme se deslocam nele.
- ▶ A tarefa de construir o mapa do local e ao mesmo tempo ser capaz de se localizarem no mapa chama-se localização e mapeamento simultâneos (SLAM, em inglês) e é uma das tarefas mais importantes que os robôs móveis devem ser capazes de executar.
- ▶ Existem muitas abordagens para conseguir fazer SLAM, e entre as mais bem sucedidas encontramos as baseadas em filtros de partículas.

Mapeamento

- ▶ Exemplos de mapas construídos por um robô no laboratório SOCIA usando métodos diferentes.



Conteúdo

Robótica
Introdução
Localização e mapeamento
Aplicações da robótica

Questões filosóficas na IA
IA Fraca
IA Forte
Ética e o risco da IA
Leitura recomendada

Aplicações da robótica: industriais




www.fanuc.co.jp

KUKA offers a comprehensive range of industrial robots. You will always find the right, no matter how challenging the application:

- **Six-axis robots** in virtually all sizes with different payload capacities and reaches and a wide range of different variants
- **HRC-capable lightweight robots** for **direct human-robot collaboration**
- **Heat- and dirt-resistant robots** for **extreme ambient conditions**
- Industrial robots in **cleanroom variants** for **strict hygiene requirements**
- Small robots as **"hygienic Machines (HM)"** for **direct contact with foodstuffs and pharmaceutical substances**
- **Press-to-press robots** with **enormous reaches** for **loading and unloading large parts in press-lining operations**
- **Palletizing robots** for **handling tasks of all kinds**
- **Welding robots** designed for **accuracy and utmost agility**
- **Shelf-mounted robots** in all variants
- **High-accuracy robots** for **utmost precision**

Aplicações da robótica: colaborativos



www.kuka.com

Requirement for intrinsic safety: no safety sensors vs. external sensors

Types of collaboration with industrial robots

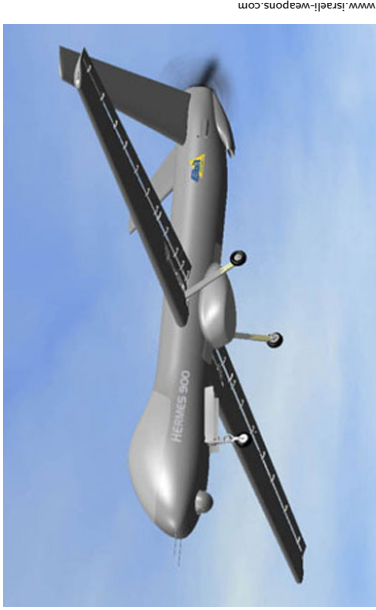
Coexistence	Supervised collaboration	Collaborative	Full collaboration
Robot and worker work in the same space, but the worker is not in the robot's workspace	Robot and worker work in the same space, but the worker is not in the robot's workspace	Robot and worker work in the same space, but the worker is not in the robot's workspace	Robot and worker work in the same space, but the worker is not in the robot's workspace

Level of collaboration

From most to least collaborative: Coexistence, Supervised collaboration, Collaborative, Full collaboration

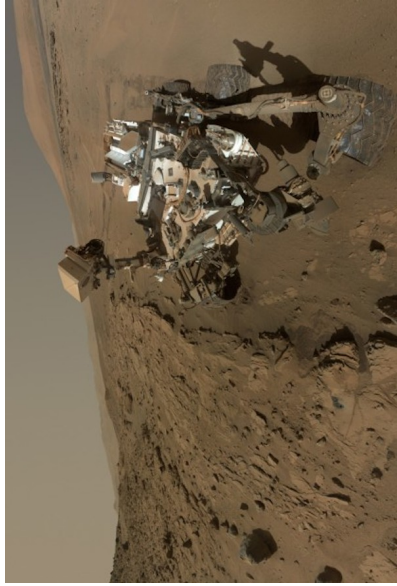
From most to least collaborative: Coexistence, Supervised collaboration, Collaborative, Full collaboration

Aplicações da robótica: militares



www.israeli-weapons.com

Aplicações da robótica: espaciais



mars.jpl.nasa.gov

Introdução

- ▶ A ideia de **IA fraca** é a de que será possível construir máquinas que **ajam** como humanos.
- ▶ A **IA forte** diz que essas máquinas conseguem **pensar** como os humanos.
- ▶ A IA foi fundada em 1956 com a ideia de que seria possível construir uma IA fraca.
- ▶ Na realidade isto depende da forma como definimos o que é a IA.
- ▶ Se for um agente que consiga agir da melhor forma possível, dados os recursos à sua disposição, então isto é sem dúvida possível. Se o agente for um programa num computador com k bits de espaço, só existem 2^k programas possíveis e é uma questão de fazer uma pesquisa exaustiva pelo melhor possível.
- ▶ Claro que a definição de IA pode ser mais ambiciosa, como um programa que consiga passar o teste de Turing ou semelhante.

Objecções à IA fraca

- ▶ **Problema da qualificação**: o comportamento humano é complexo demais para poder ser representado com um conjunto de regras, e como as máquinas só seguem conjuntos de regras nunca vão poder reproduzir o comportamento humano.
- ▶ Na realidade esta afirmação é parcialmente correta: um conjunto de regras não consegue captar toda a complexidade do comportamento humano e é por isso que as abordagens baseadas em lógica estão limitadas nos resultados que conseguem obter.
- ▶ Mas as abordagens probabilísticas conseguem ir mais além: não são apenas um conjunto de regras e permitem alcançar comportamentos muito complexos.

Conteúdo

Robótica
Introdução
Localização e mapeamento
Aplicações da robótica

Questões filosóficas na IA
IA Fraca
IA Forte
Ética e o risco da IA
Leitura recomendada

Objecções à IA fraca

- ▶ **Objeção matemática**: o teorema de Gödel diz que sistemas formais suficientemente poderosos para conterem a aritmética, contêm afirmações que não se podem provar.
- ▶ Como a IA é um sistema formal que contém a aritmética, tem então que estar limitada pelo Teorema de Gödel e é imediatamente inferior aos humanos.
- ▶ Na realidade esta objecção tem 3 fraquezas:
 - ▶ O teorema aplica-se a máquinas de Turing, mas estas são infinitas ao passo que os computadores são finitos e podem ser descritos como um sistema de lógica proposicional que não está sujeito ao teorema.
 - ▶ O facto de existirem algumas afirmações que não podem ser provadas, não tira mérito a um sistema. Os humanos comportaram-se de forma inteligente muito tempo antes de terem inventado a matemática, logo este formalismo não é fundamental para a inteligência.
 - ▶ Não há qualquer prova de que os humanos não estejam também sujeitos ao teorema, pois para o provar teríamos de ter uma teoria formal do pensamento humano, que não existe.

IA forte

- ▶ Será que máquinas conseguiriam alguma vez pensar?
- ▶ Muita gente diz que, mesmo que uma máquina consiga passar o teste de Turing ou outro semelhante, isso não significa que a máquina consiga pensar, estaria apenas a **simular** o pensamento.
- ▶ Objeção (Geoffrey Jefferson, 1949): Só poderemos concordar que uma máquina seja igual ao cérebro humano se conseguirmos escrever um soneto ou compor um concerto devido aos pensamentos e emoções que sinta, e não pela manipulação de símbolos – teria que ser capaz de escrever mas também de **saber** que escreveu.
- ▶ Turing chamou a isto o **argumento da consciência**: a máquina tem que ter consciência dos seus estados mentais e ações.

IA forte

- ▶ Turing respondeu a este argumento dizendo que era estranho que fizéssemos estas exigências às máquinas quando não as fazemos às pessoas.
- ▶ Nós não temos forma de saber quais os estados mentais de outras pessoas.
- ▶ O que fazemos é assumimos à partida que os outros são inteligentes.
- ▶ Turing disse que Jefferson também poderia assumir isso sobre máquinas se as visse a agir de forma inteligente.

Ética e o risco da IA

- ▶ Além do problema de ser ou não possível construir uma IA, está o problema se **devemos** ou não construir uma IA.
- ▶ Se os efeitos da IA forem mais negativos que positivos então talvez não deva ser desenvolvida.
- ▶ Muitas tecnologias podem ser usadas tanto para o bem como para o mal: uma faca serve para cortar os legumes mas pode ser usada para ferir ou matar alguém.

Ética e o risco da IA

- ▶ Problemas com a IA
 - ▶ as pessoas podem perder empregos
 - ▶ as pessoas poderão ficar com demasiado tempo livre
 - ▶ as pessoas poderão perder o sentido de que somos únicos e especiais
 - ▶ a IA pode ser usada para o mal
 - ▶ podemos deixar de ter pessoas responsáveis por determinadas ações
 - ▶ o sucesso da IA pode significar o fim da humanidade

Ética e o risco da IA: demasiado tempo livre

- ▶ Em 1970 Alvin Toffler escreveu que a semana de trabalho tinha sido reduzida em cerca de 50% desde o início do século e que seria possível que voltasse a ser cortada na mesma grandeza em 2000.
- ▶ O Arthur C. Clarke em 1968 escreveu que as pessoas no ano 2001 poderiam ter um futuro aborrecido em que o único problema que teriam seria escolher que canal da TV queriam ver.
- ▶ Na realidade vemos que hoje as pessoas andam mais ocupadas que nunca e que quem tem empregos mais intensivos do ponto de vista do conhecimento trabalha sem horários e que para conseguirem “acompanhar” o que se passa, têm que trabalhar cada vez mais.

Ética e o risco da IA: as pessoas podem perder empregos

- ▶ Este tipo de receios surge sempre que uma tecnologia revoluciona as formas de trabalho e de produção.
- ▶ Ex.: a máquina a vapor, o telefone, a internet.
- ▶ Analisando o impacto da IA no tecido económico conclui-se que criou mais empregos que os que destruiu e que estes empregos são mais interessantes e mais bem pagos que os que desapareceram.

Ética e o risco da IA: demasiado tempo livre

- ▶ Numa economia industrial as recompensas são proporcionais ao tempo investido: trabalho mais 10%, ganho mais 10%.
- ▶ Numa economia baseada em conhecimento existe uma recompensa muito grande para se ser ligeiramente melhor que os outros: alguém que trabalhe 10% mais pode ganhar 100% mais. Isto leva a uma pressão enorme para se trabalhar muito.
- ▶ A IA poderá ajudar a que, utilizando agentes para nos automatizar algumas tarefas, consigamos ter algum tempo extra para descansar.

Ética e o risco da IA: perder o sentido de que somos únicos

- ▶ Há quem diga que, se a IA for para a frente, a ideia de que somos apenas máquinas (biológicas) pode fazer-nos perder a ideia de que somos especiais.
- ▶ Na realidade já outras teorias abalaram as nossas convicções relativamente à nossa importância:
 - ▶ Copérnico tirou a Terra do centro do sistema solar;
 - ▶ Darwin colocou o Homem ao nível das outras espécies animais.
- ▶ O sucesso da IA será apenas mais um destes casos.

Ética e o risco da IA: a IA pode ser usada para o mal

- ▶ A tecnologia de ponta tem sido desde sempre usada para derrotar rivais.
- ▶ Tecnologias que usam IA já estão a ser usadas na guerra: em 2009 existiam mais de 5000 UAVs e mais de 12000 UGVs na guerra do Iraque.
- ▶ Ninguém se oporia à ideia de um soldado humano usar uma blindagem para se proteger dos adversários, e podemos pensar no uso destes robôs como o passo seguinte nesta lógica: mantêm o soldado o mais protegido possível, ficando fora do campo de batalha.
- ▶ Claro que estas armas robóticas apresentam riscos: se retirarmos completamente o humano do circuito, podemos ter sistemas automáticos a tirar vidas humanas, inclusive a civis.

Ética e o risco da IA: a IA pode ser usada para o mal

- ▶ Pode também levar a que países que possuam muito deste equipamento percam o receio de se envolverem em guerras, e dessa forma, esta tecnologia poderá aumentar o número de guerras de forma indireta.
- ▶ Outros usos da IA para o mal seriam o uso indiscriminado da IA para que os estados espiem as atividades dos cidadãos levando a uma perda de liberdades civis e de privacidade.

Ética e o risco da IA: deixar de ter pessoas responsáveis

- ▶ Nos sistemas de diagnóstico médico, quem é o responsável se um doente tiver problemas devido a um sistema baseado em IA ter errado? Continua a ser o médico.
- ▶ Se os sistemas baseados em IA passarem a ser mais precisos que os médicos humanos, os médicos poderão passar a ser responsabilizados por **não** usarem esses sistemas e tomarem decisões sem a sua ajuda.
- ▶ Quem é responsável se um agente na internet apagar ficheiros de terceiros? Ou se um agente automático fizer um negócio e perder dinheiro ficando com dívidas?
- ▶ E no caso dos carros autónomos: quem deve ser responsabilizado por um acidente?
- ▶ Todas estas questões merecem reflexão e mostram que o sistema legal não está a conseguir acompanhar o desenvolvimento tecnológico.

Ética e o risco da IA: o fim da humanidade

- ▶ Por vezes falamos em determinada tecnologia cair nas mãos erradas e ser usada para o mal.
- ▶ Com a IA temos o pior cenário possível: as mão erradas poderão ser as da própria tecnologia!
- ▶ Será que a IA é mais perigosa que o resto do software que usamos?
- ▶ A estimação do estado pode ser incorreta e levar a que um agente faça coisas erradas (mudar de faixa na altura errada ao conduzir, p.ex.).
- ▶ Um sistema de defesa com mísseis pode errar e destruir um avião civil.
- ▶ A forma de lidar com este problema é desenvolver sistemas com verificações muito exigentes para que um único erro não possa levar a uma ação desastrosa.

Ética e o risco da IA: o fim da humanidade

- ▶ A **função utilidade** usada por um robô tem que ser especificada cuidadosamente.
- ▶ Ex.: podemos pensar em programar um robô para minimizar o sofrimento humano. Como os humanos arranjam sempre formas de sofrer, o robô pode ser tentado a acabar com espécie humana para acabar com o sofrimento.
- ▶ A evolução de um sistema de IA pode levar a que um sistema inicialmente bom se torne mau. Este problema é real e está ligado ao desenvolvimento da singularidade.

Ética e o risco da IA: o fim da humanidade

- ▶ A **singularidade** indica o ponto no tempo em que uma IA ultrapassa os humanos em inteligência.
- ▶ Como desenhar sistemas inteligentes é uma atividade inteligente, uma máquina com inteligência superior à humana pode desenhar máquinas ainda mais inteligentes e levar a uma “explosão” de inteligência (artificial).
- ▶ Os humanos seriam rapidamente ultrapassados: a primeira máquina com inteligência superior à humana pode ser a última máquina inventada pelo Homem.

Luísa A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

37 / 40

Ética e o risco da IA: o fim da humanidade

- ▶ Há no entanto muitas propostas para evitar estes fins trágicos.
- ▶ Isaac Asimov, em 1942, propôs dotar os robôs de 3 regras (as 3 leis da robótica):
 - ▶ Um robô não pode ferir um humano ou, através da inação, permitir que um humano sofra.
 - ▶ Um robô tem que obedecer às ordens que os humanos lhe dão, exceto se essas ordens forem contra a primeira lei.
 - ▶ Um robô deve proteger a sua própria existência, desde que não entre em conflito com as duas leis anteriores.

Luísa A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

38 / 40

Ética e o risco da IA: o fim da humanidade

- ▶ Segundo Yudkowsky, temos que desenhar o mecanismo de evolução para a IA de forma a que estes sistemas evoluam mas permaneçam amigáveis.
- ▶ Não podemos dar-lhes funções de utilidade rígidas, têm de poder evoluir. Imaginar o que seria termos hoje uma IA com a moral do séc. XIX: quereria impor a escravatura e tirar o direito de voto às mulheres.
- ▶ A melhor forma de ter proteção contra evoluções negativas da IA será dar-lhe valores que levem em conta a sociedade e o bem comum.

Luísa A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

39 / 40

Leitura recomendada

- ▶ Russell e Norvig, sec. 25.1,25.3.1,26.
- ▶ Bauer, Wilhelm (Ed.); Bender, Manfred; Braun, Martin; Rally, Peter und Scholtz, Oliver (2016): Lightweight Robots in Manual Assembly – Best to Start Simply. Examining Companies' Initial Experiences with Lightweight Robots. Fraunhofer Institute for Industrial Engineering IAO, Stuttgart.

Luísa A. Alexandre (UBI)

Inteligência Artificial

Ano lectivo 2019-20

40 / 40