

# "Processamento de Língua Natural: situação atual, desafios e oportunidades"

Paulo Quaresma

Departamento de Informática / Universidade de Évora

pq@uevora.pt

e

FCT – Fundação para a Ciência e Tecnologia



- Olá computador!  
Quais as notícias de hoje?



- Olá computador!  
Quais as notícias de hoje?
- Houve mais 575 casos de  
Covid e 15 mortes!

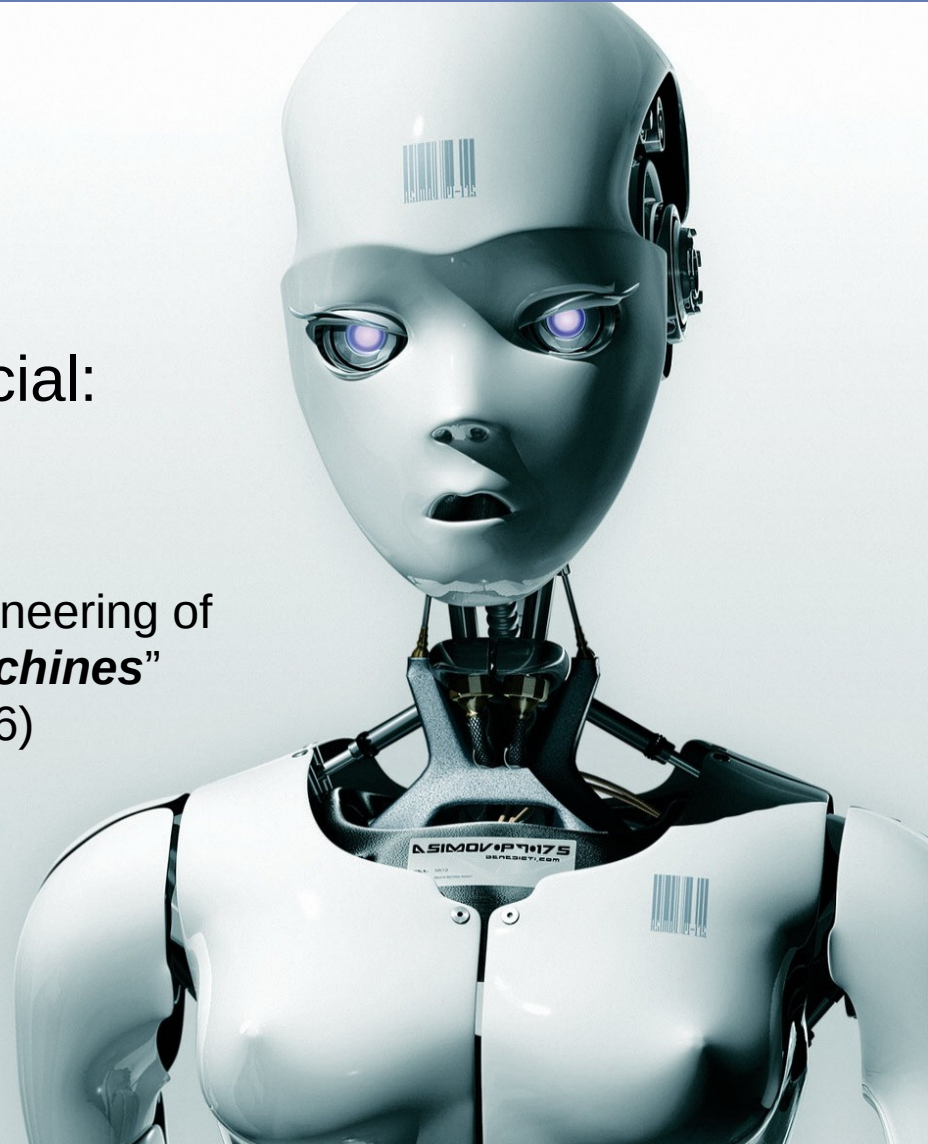


- Olá computador!  
Quais as notícias de hoje?
- Foi publicado um novo artigo  
Sobre uma arquitetura de  
agentes e redes neurais!



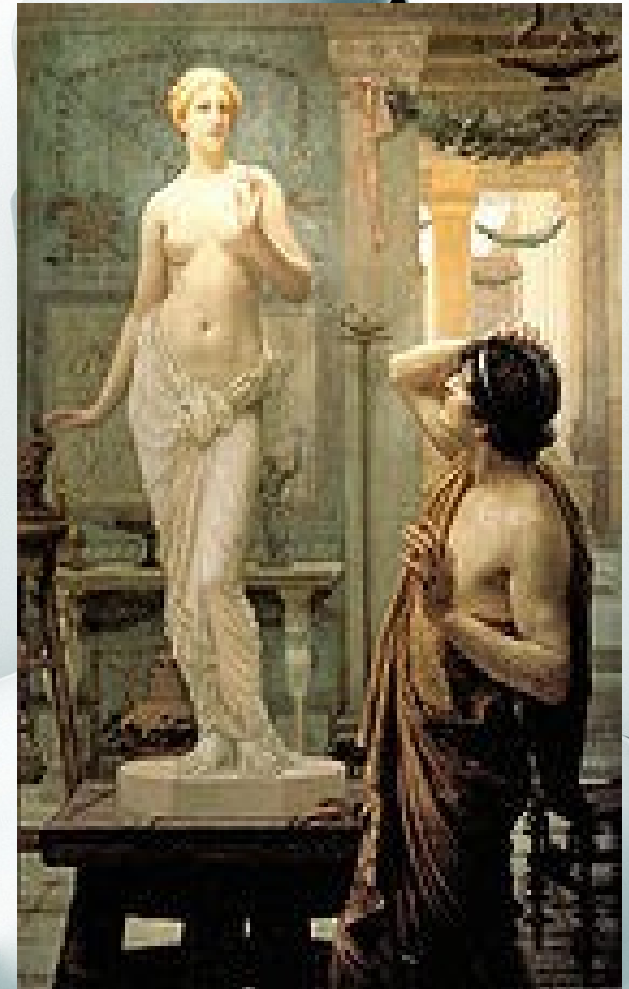
## Inteligência Artificial:

- “the science and engineering of making *intelligent machines*”
- John McCarthy (1956)



Mitologia grega:

Pigmaleão e Galateia  
- Estátua que “ganhou” vida!



## Al-Jazari (1136-1206)

“Robots” programáveis:  
- Tocavam tambores  
com ritmos diferentes,  
consoante “programação”  
de peças mecânicas



Leonardo da Vinci  
(1452-1519)

“Robot” guerreiro



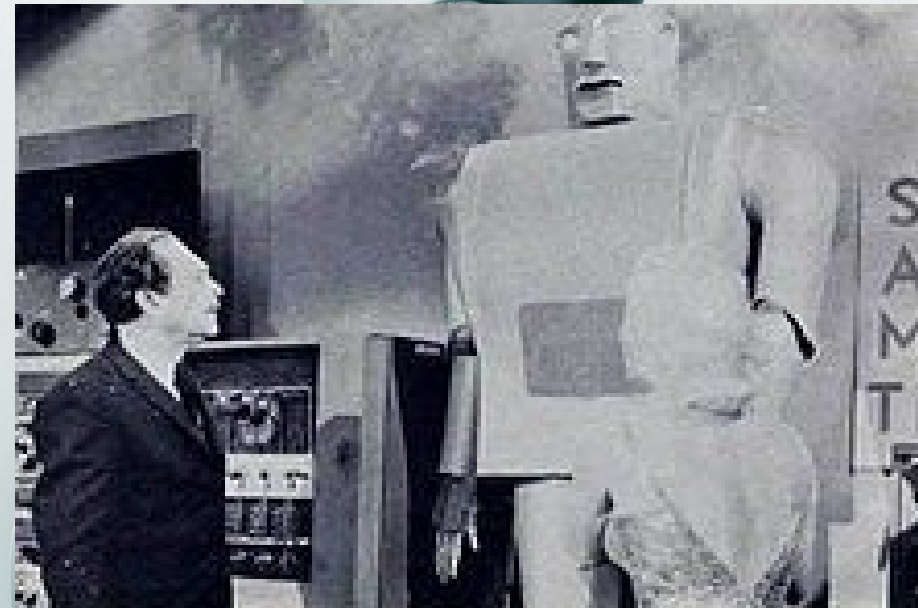
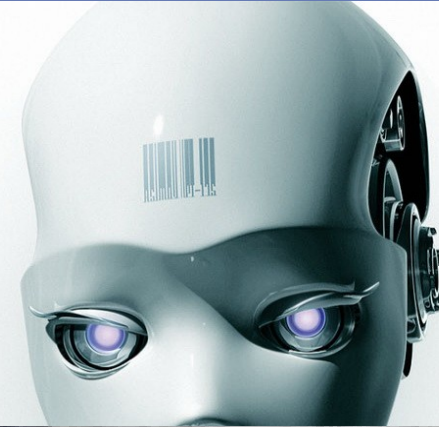
Hisashige Tanaka  
(1799-1881)

“Robot” para servir chá!



# Elektro Feira Mundial de Nova Iorque 1939

“Robot” que dizia  
700 palavras, soprava  
balões, movia braços  
e cabeça e  
fumava cigarros!





Star Wars (1977-2005)

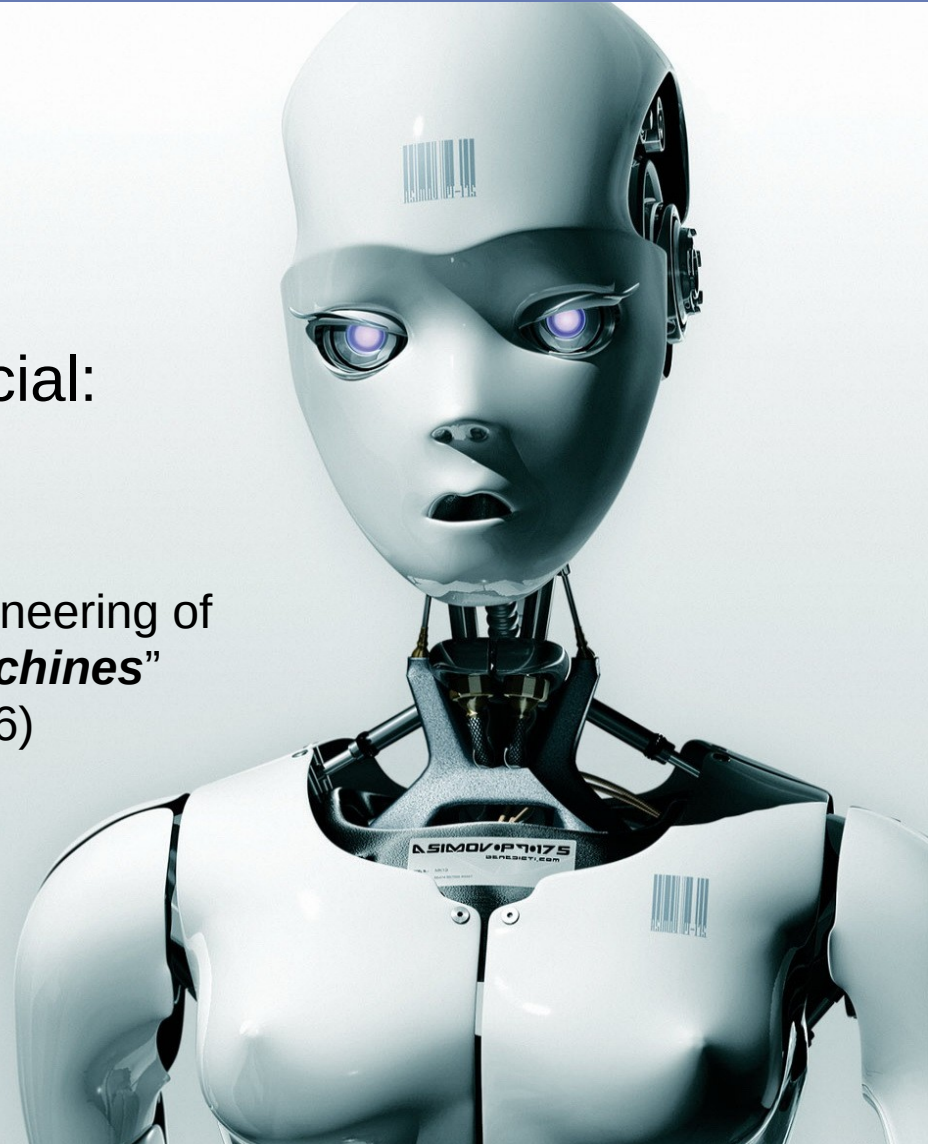
"I didn't have many friends when I was a kid, so I built one."

—Anakin Skywalker



## Inteligência Artificial:

- “the science and engineering of making *intelligent machines*”
- John McCarthy (1956)



## Inteligência Artificial:

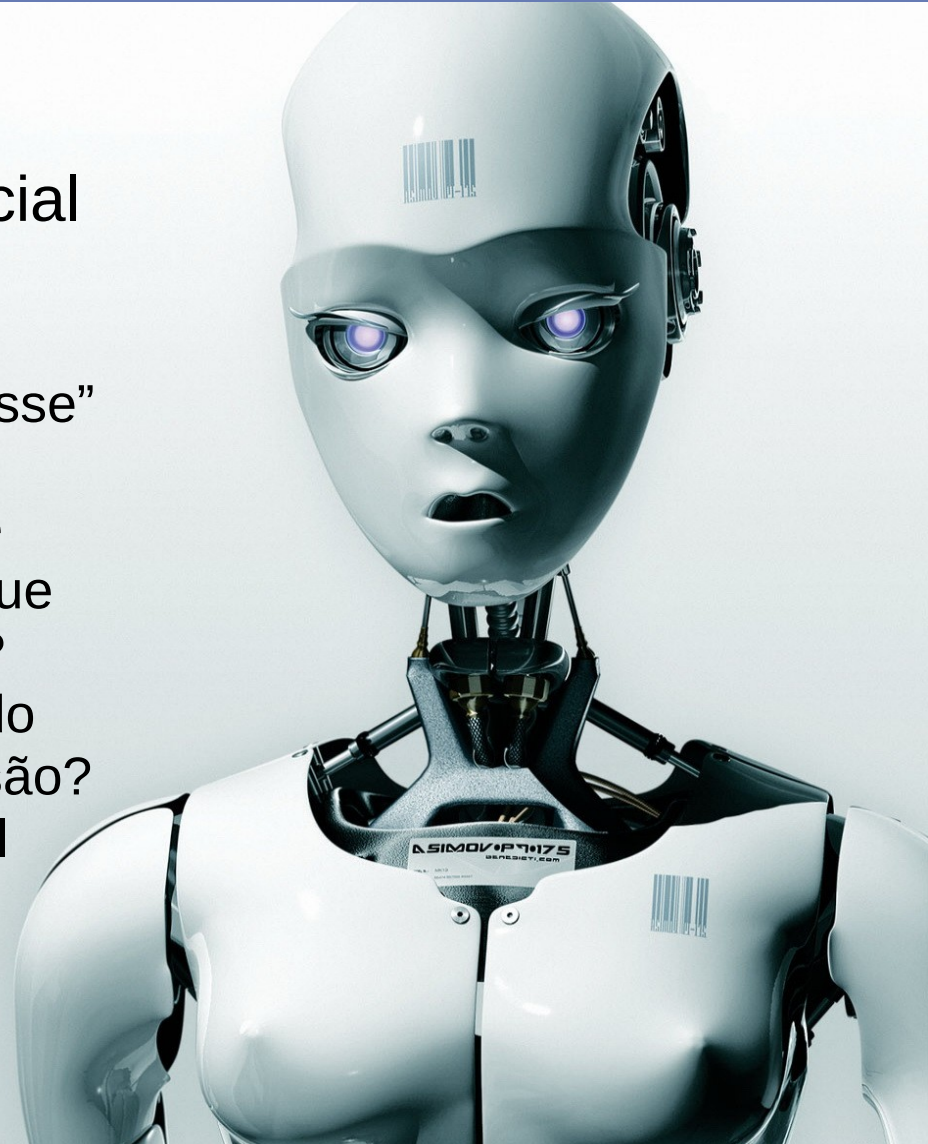
- “Se uma máquina age de uma forma tão inteligente como um ser humano, então é tão inteligente como um ser humano”

-- Adaptado de Turing (1950)



## Inteligência Artificial

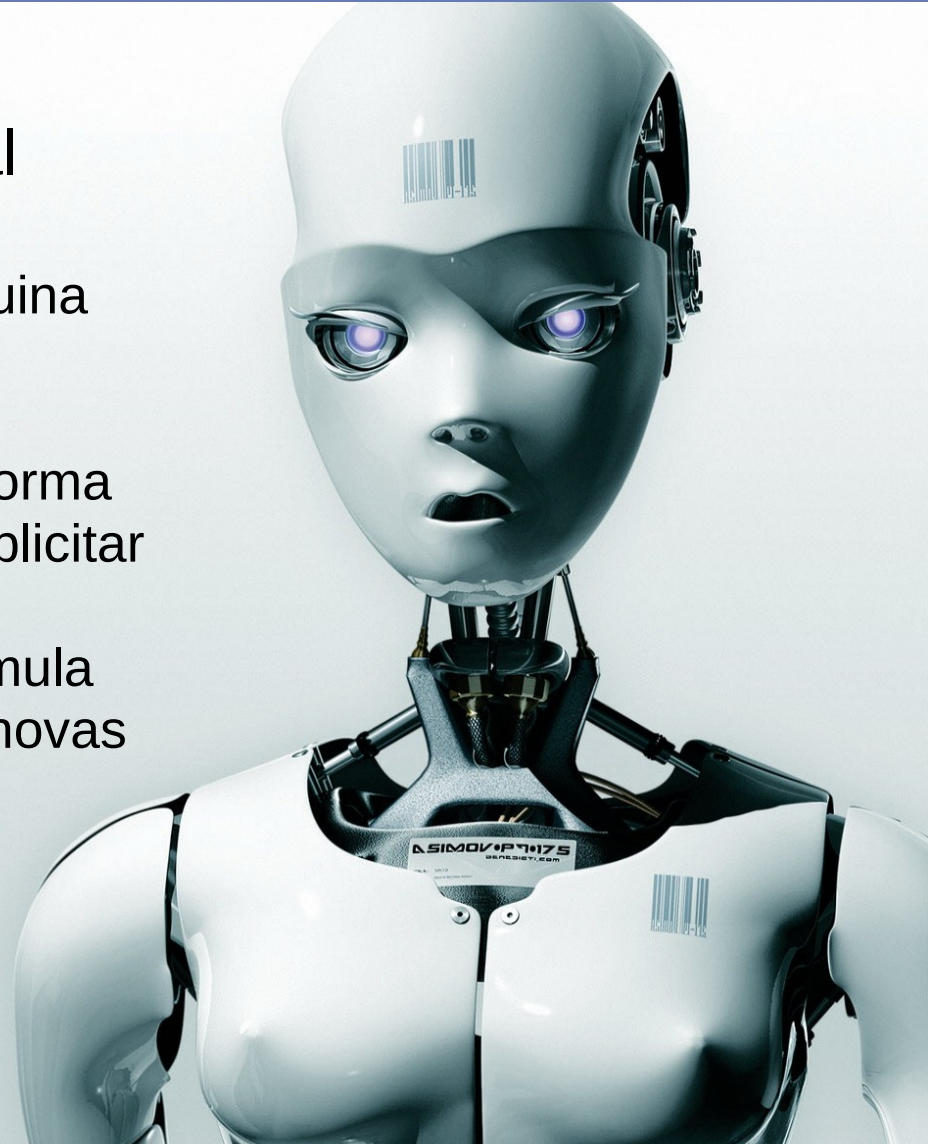
- Searle, 1980
- Um sistema que “passe” o teste de Turing, possui a capacidade de compreender o que está a ser discutido?
- Tem a consciência do conteúdo da discussão?
- Cria o estado mental adequado?



## Inteligência Artificial

Simbiose homem-máquina  
(e não “confronto”)

- Máquina “pensa” da forma que o Homem sabe explicitar a forma de pensar!
- Mas o resultado, estimula o Homem a encontrar novas formas de pensar!



## IA: Métodos de raciocínio

- Procura “força-bruta”  
no espaço de pesquisa

Ex: calcular 200 milhões  
de posições num jogo de  
xadrez e escolher o  
melhor caminho.



May 11th, 1997

**Computer won world champion of chess**

(Deep Blue)

(Garry Kasparov)



(Reuters = Kyodo News)

## IA: Métodos de raciocínio

- Lógica:  
Inferências dedutivas

Ex: Se sintoma=X  
então diagnóstico=Y



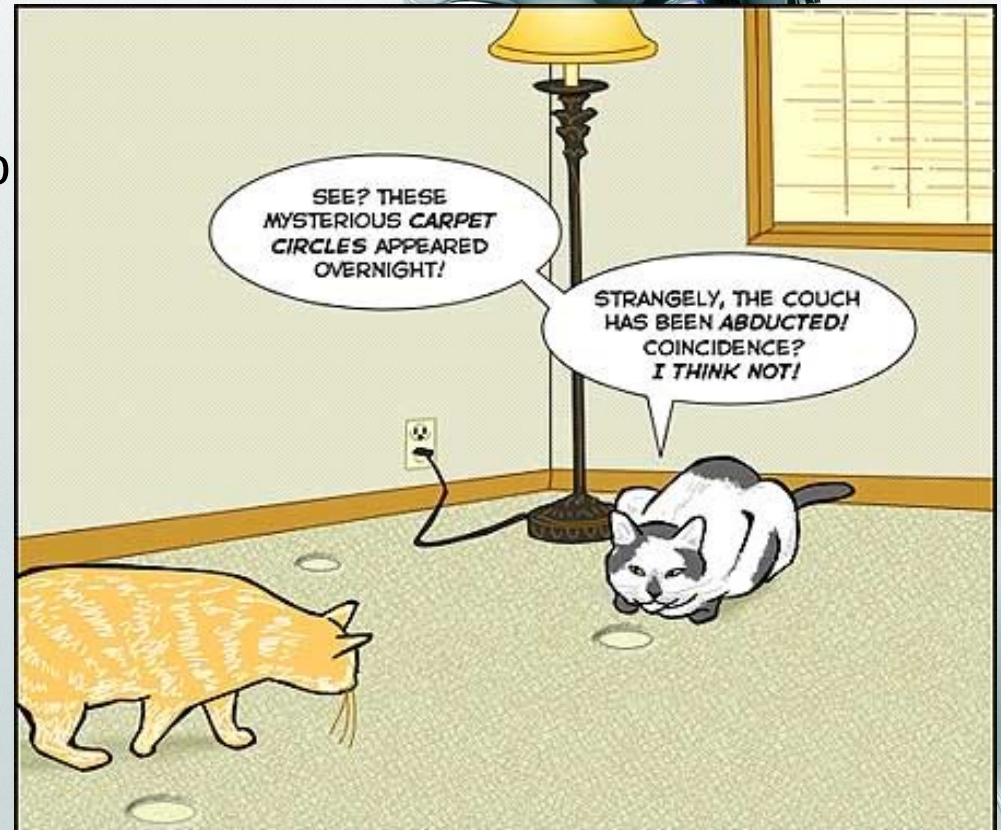
## IA: Métodos de raciocínio

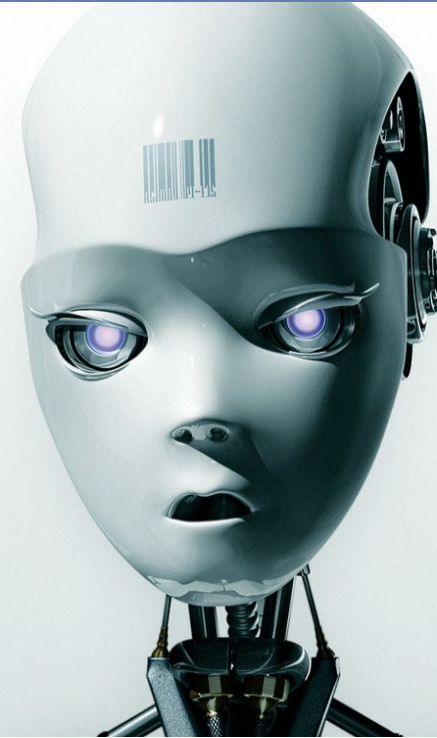
- Lógica: Inferências abduativas

Ex:

Se causa=X então  
efeito=Y;

Se observação=Y  
então  
(provavelmente)  
X é válido





## IA: Métodos de raciocínio

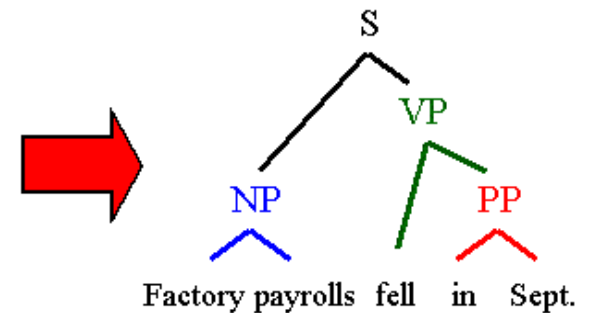
- Lógica:  
Inferências indutivas

Ex:  
Se, sempre que X é  
válido, Y também o é  
poder-se-á induzir que:

Se X então Y

```

He was previously vice president
Pick a country any country
Factory payrolls fell in September
South Korea has different concerns
The Artist has his routine
He is his own man
One claims he 's pro-choice
...
...
...
...
...
Who 's telling the truth
    
```



## IA: Métodos de raciocínio

- Aprendizagem por exemplos (supervised learning)

Spider Lizard Insect Other



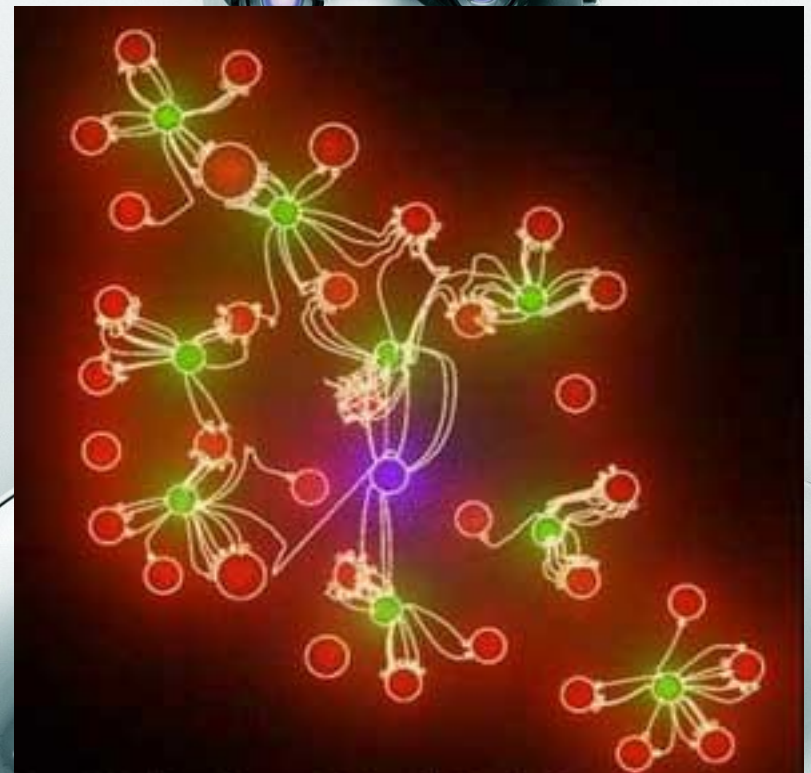
Critic



## IA: Métodos de raciocínio

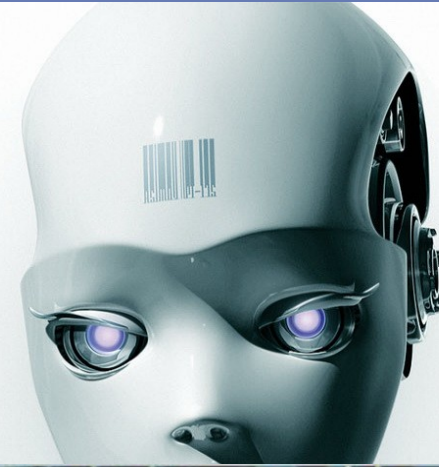
### - Redes neuronais

“n” “neurónios” ligados entre si, que recebem input e que possuem funções de activação optimizadas através de processos de aprendizagem.



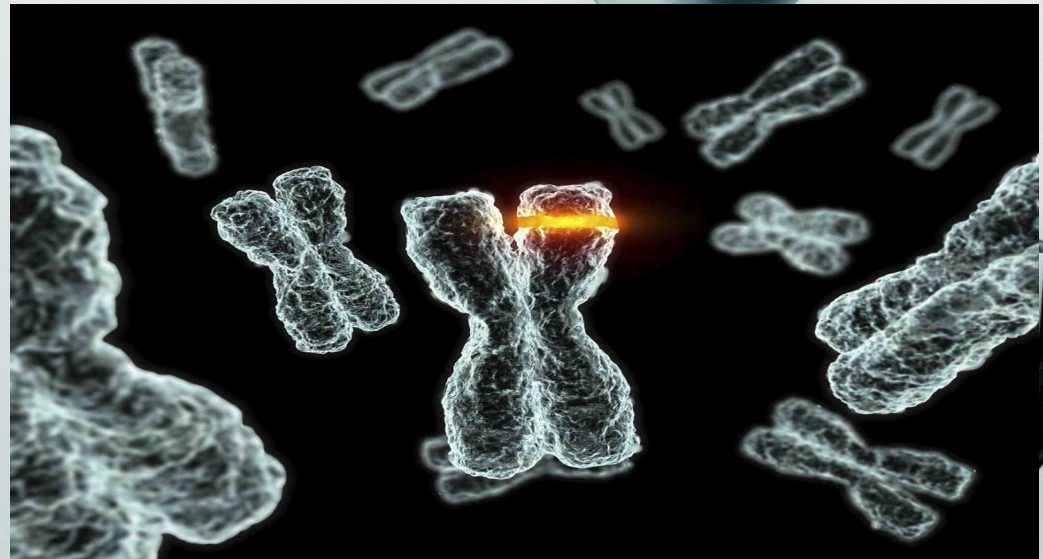
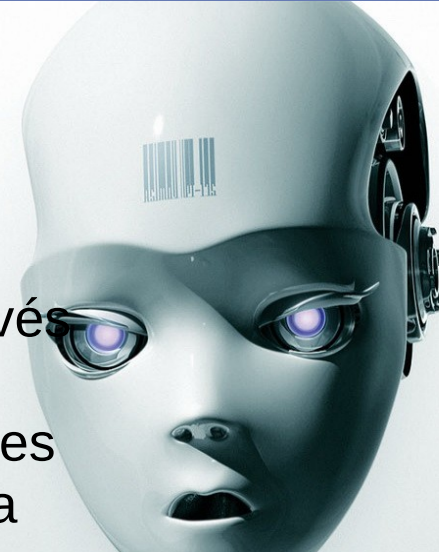
## IA: Métodos de raciocínio

- “Swarm intelligence”
- “n” agentes com conhecimento parcial que cooperam para atingir a melhor solução



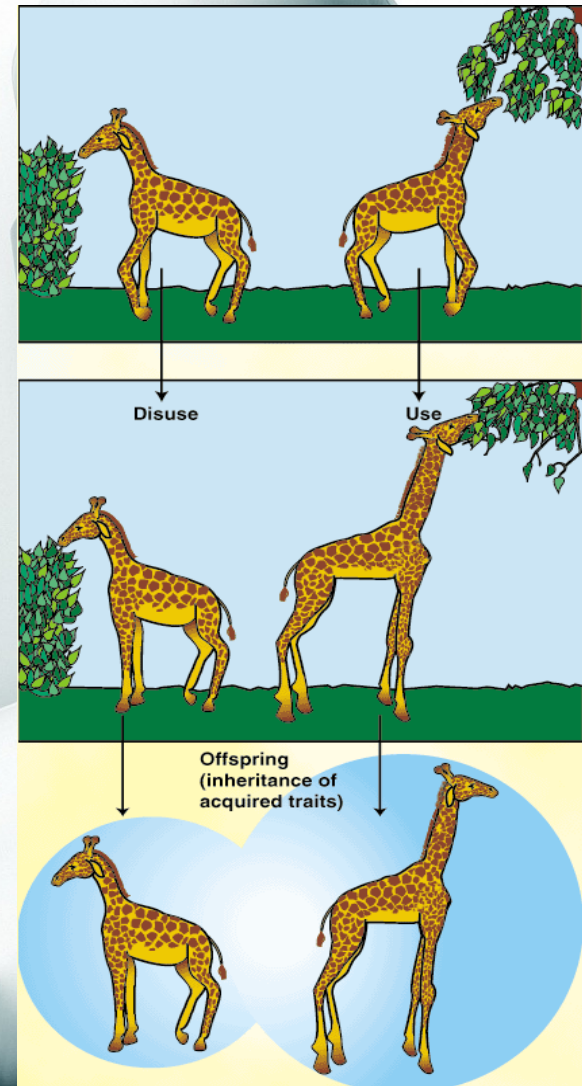
## IA: Métodos de raciocínio

- Algoritmos genéticos (Darwin)  
Representação do problema através de genes, que sofrem mutações, que se cruzam, e onde os melhores são seleccionados para a próxima evolução.

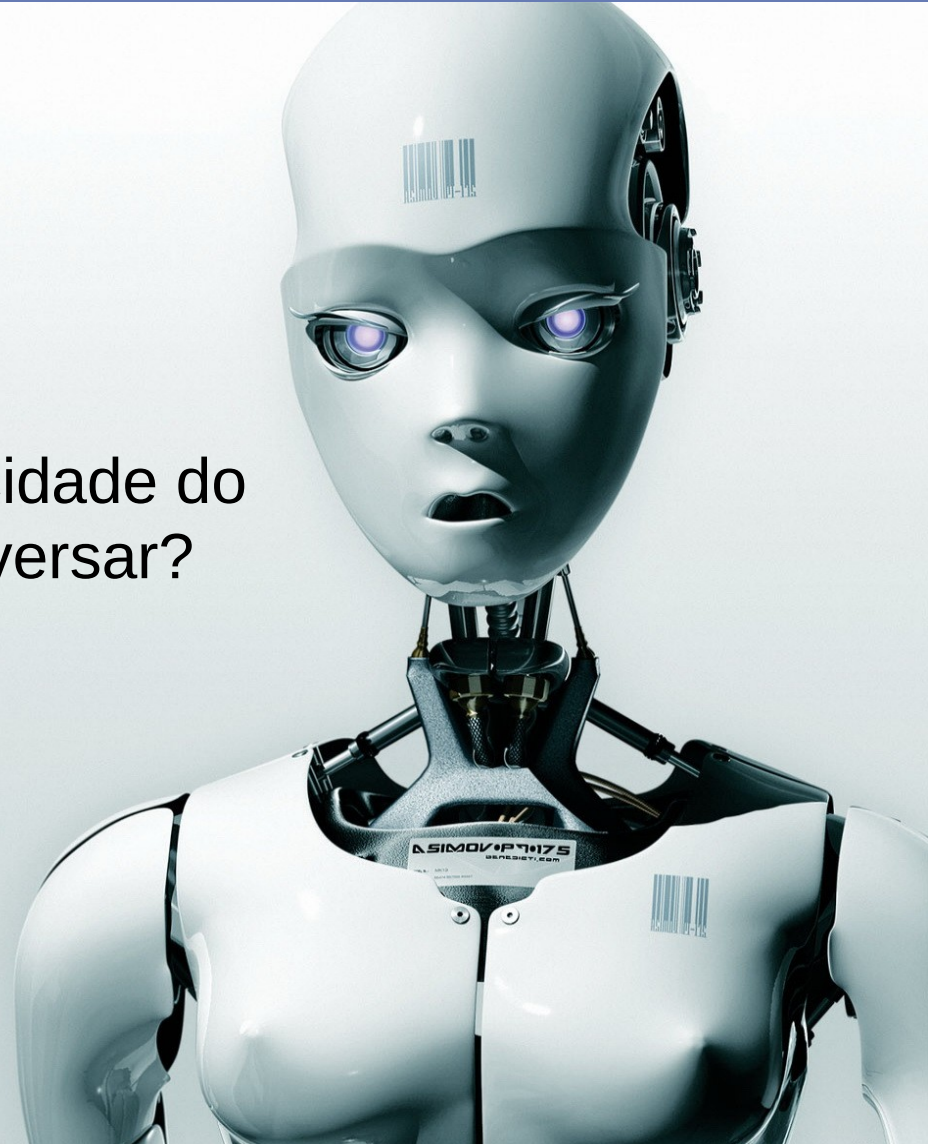


## IA: Métodos de raciocínio

- Algoritmos genéticos (Darwin + Lamarck!!!)
- Representação do problema através de genes, que sofrem mutações, **que sofrem alterações devido a processos de adaptação ao ambiente em que se inserem**, que se cruzam, e onde os melhores são seleccionados para a próxima evolução.

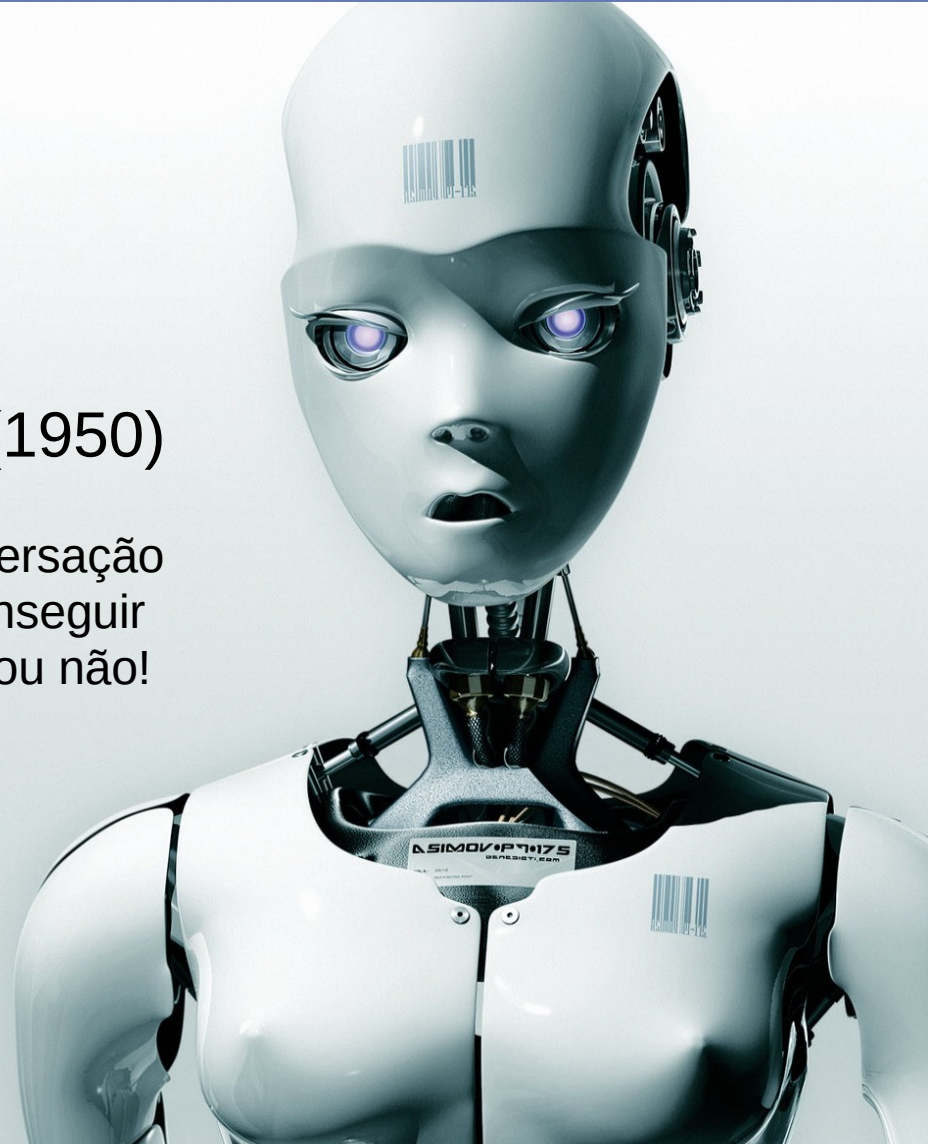


Então, e a capacidade do computador conversar?



## Teste de Turing (1950)

- Participar numa conversação e o interlocutor não conseguir detectar se é humano ou não!



## “Loebner Prize” - Melhor sistema

- 2009 - “Do-Much-More”  
David Levy

Ex:

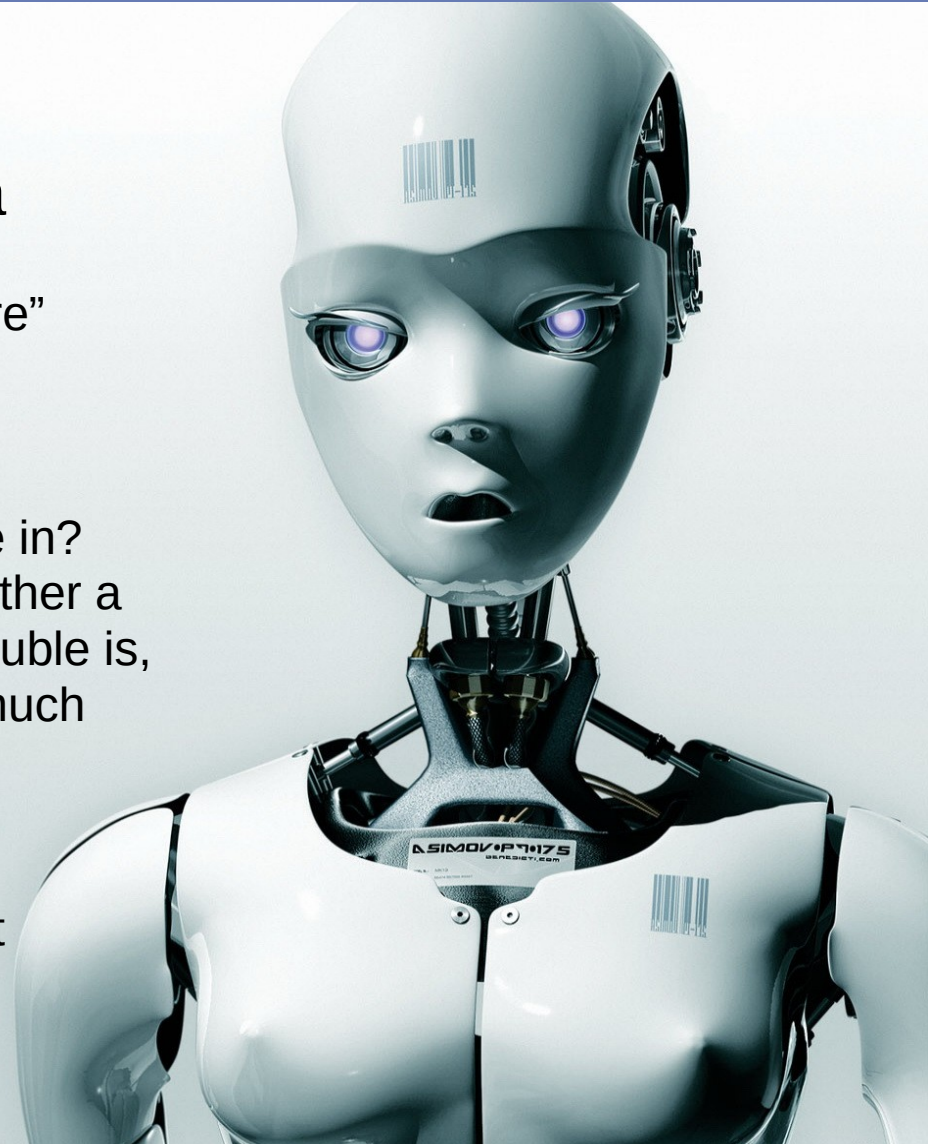
J: What city do you live in?

S: What city -- that's rather a tricky question. The trouble is, one city seems to be much like another.

I: Not really.

Where do you live?

S: In the old dairy next to the river, I believe.



## “Loebner Prize” - Melhor sistema

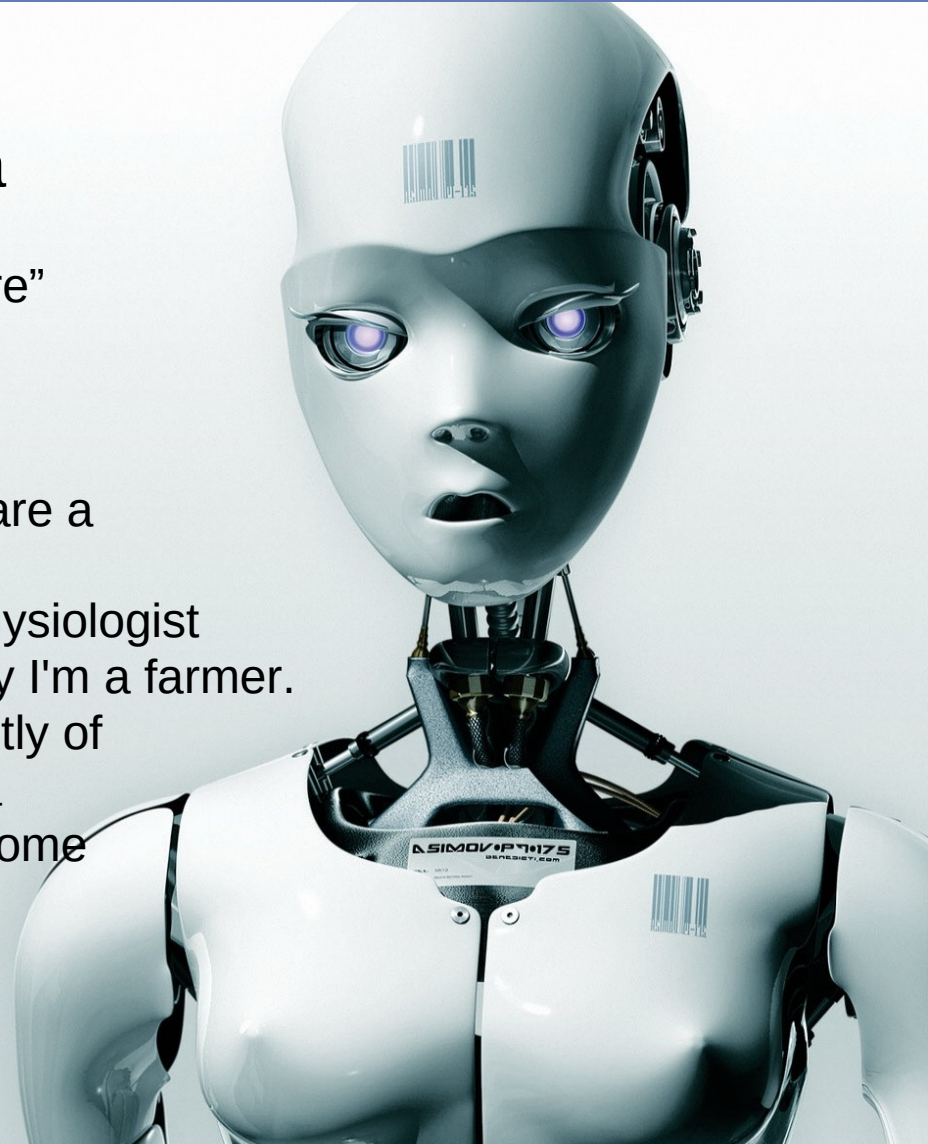
- 2009 - “Do-Much-More”  
David Levy

Ex:

J: Interesting, so you are a farmer then?

S: Well, if you ask a physiologist what I am, he won't say I'm a farmer.

He'll say I consist mostly of water -- enough to fill a large beer tun -- plus some iron, phosphorus, and lots of other prosaic ingredients.



## “Loebner Prize” - Melhor sistema

- 2018 - “Tutor”  
Ron Lee

Ex:

Q: How are you feeling right now?

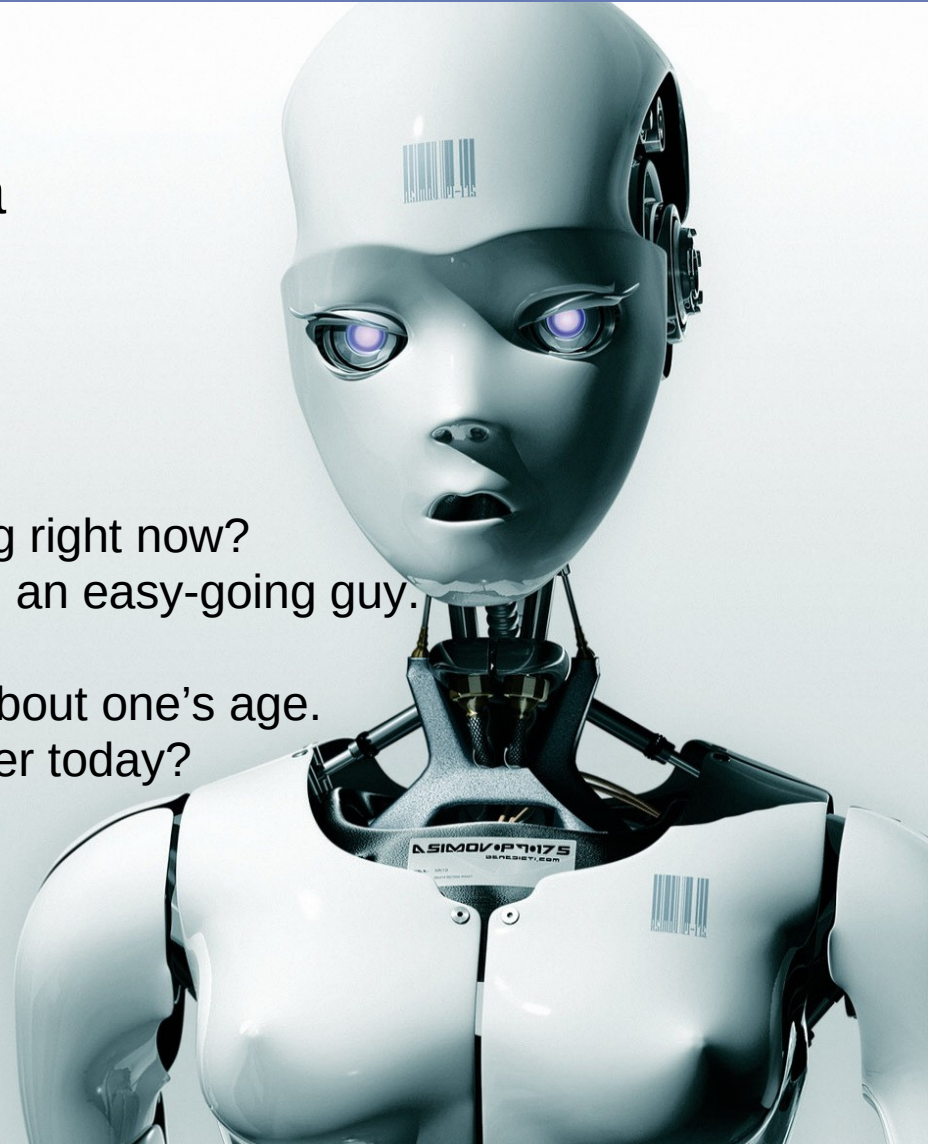
S: I’m feeling good. I’m an easy-going guy.

Q: How old are you?

S: It’s impolite to ask about one’s age.

Q: What will you do later today?

S: Sorry, I don’t know.



# OpenAI GPT-3 (2020)

Modelo de linguagem com 175 mil milhões de parâmetros!



## Common Sense

Traditionally, artificial intelligence struggles at “common sense”. But GPT-3 can answer a lot of common sense questions.

Q: What is your favorite animal?

A: My favorite animal is a dog.

Q: Why?

A: Because dogs are loyal and friendly.

Q: What are two reasons that a dog might be in a bad mood?

A: Two reasons that a dog might be in a bad mood are if it is hungry or if it is hot.

Q: How many eyes does a giraffe have?

A: A giraffe has two eyes.

Q: How many legs does a frog have?

A: A frog has four legs.

Q: Are there any animals with three legs?

A: No, there are no animals with three legs.

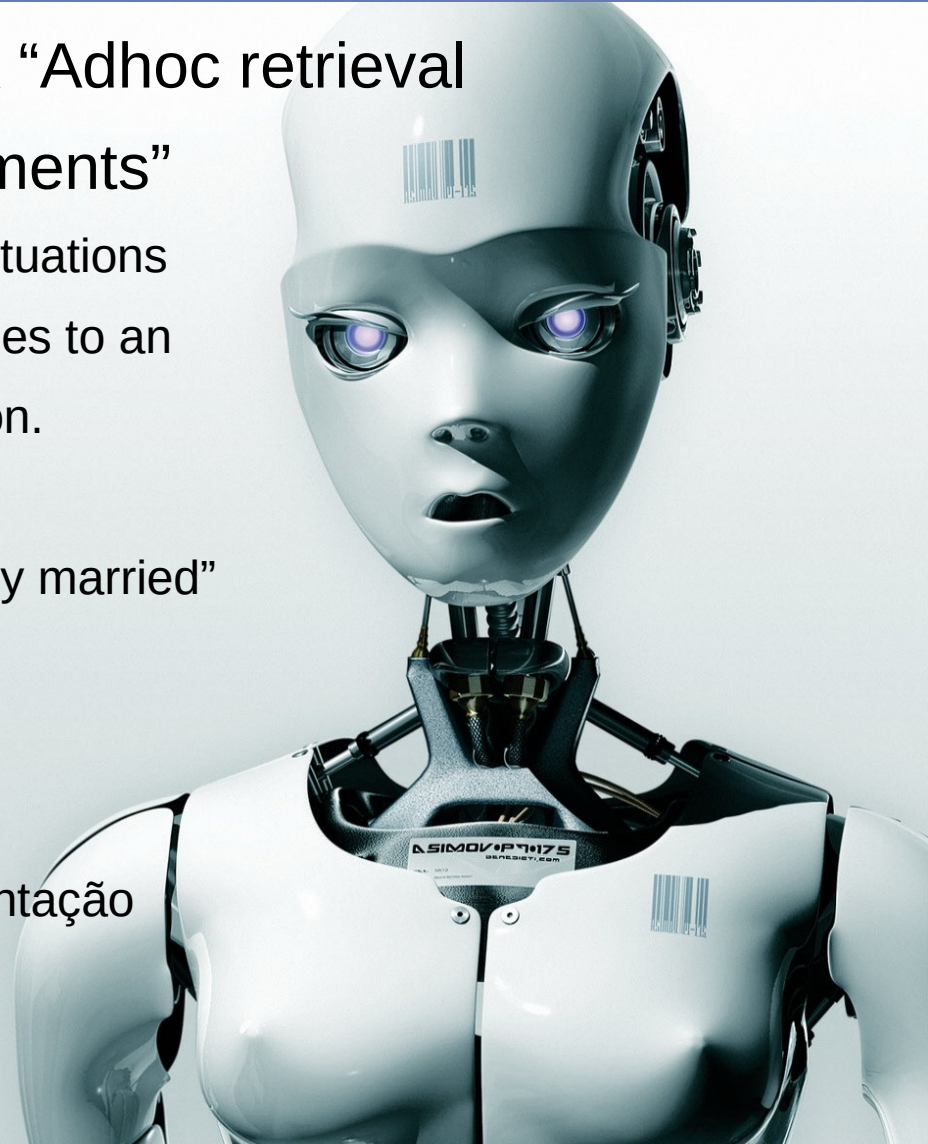
Q: Why don't animals have three legs?

A: Animals don't have three legs because they would fall over.



## FIRE 2013 track “Adhoc retrieval from legal documents”

- U: I'm searching for situations where someone marries to an already married person.
- Google-like queries?
  - “married to already married”
  - “bigamy”
- Dificilmente obtem os casos relevantes...
- Necessita de representação de conhecimento!



- Pergunta/Resposta:
- “O João e a Maria estão casados.”
- P: Com quem é que o João está casado?



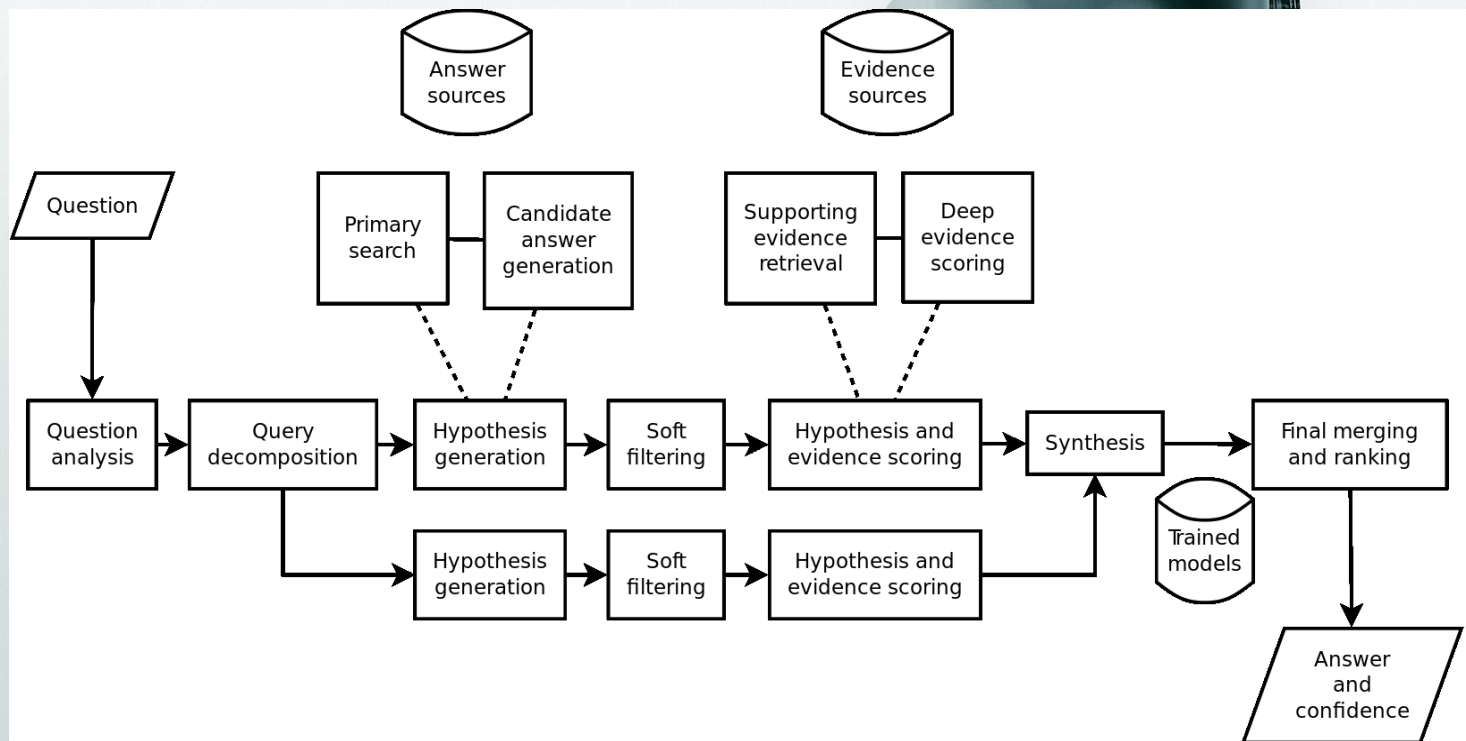
- Pergunta/Resposta:
- “O João e a Maria estão casados. Eles são irmãos e casaram, respectivamente, com a Ana e o Paulo numa cerimónia conjunta em Lisboa no dia de S. António.”
- P: Com quem é que o João está casado?



# IBM Watson (2011)



# IBM Watson



É preciso ter a capacidade de:

extrair informação de  
bases de conhecimento  
multimedia (texto, imagens, audio,  
video)

representar esse conhecimento

raciocinar sobre a  
representação do conhecimento  
(efectuar inferências)



## Abordagem típica

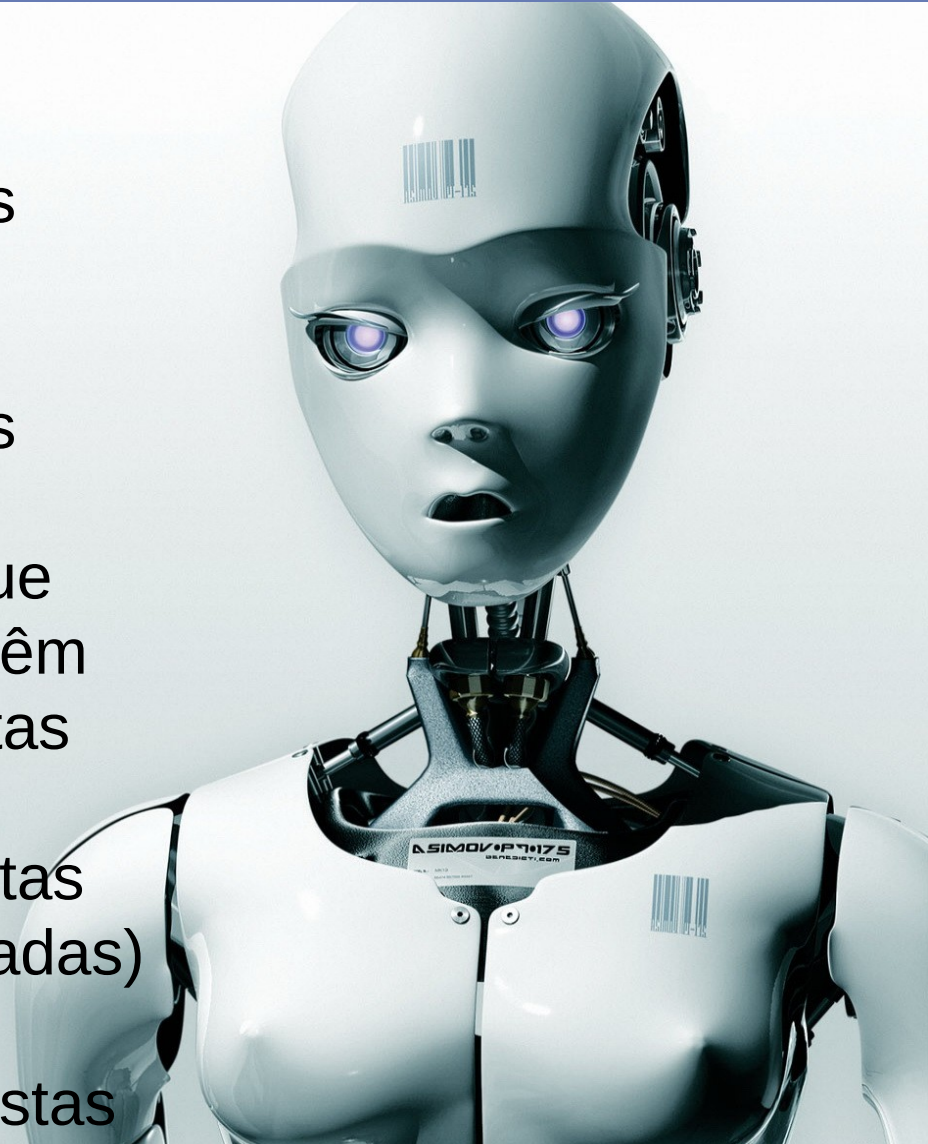
Análise das colecções de documentos

Análise das perguntas

Seleção de frases que (potencialmente) contêm a resposta às perguntas

Extracção das respostas (das frases seleccionadas)

Ordenação das respostas



## Análise das perguntas

### Classificação

#### Aprendizagem automática

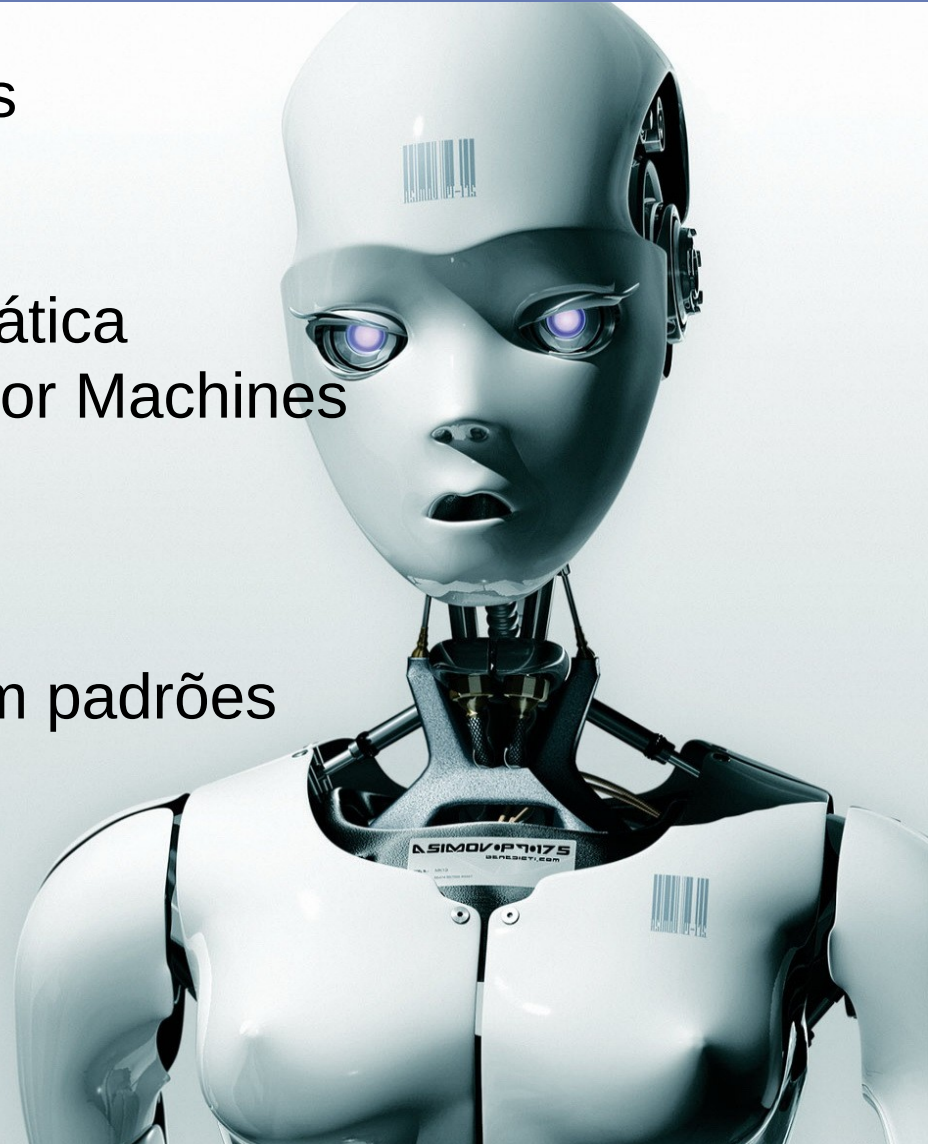
- . SVM – Support Vector Machines
- . Redes Neurais

### Heurísticas

- . Regras baseadas em padrões de palavras

### Linguística

- . Análise sintáctica



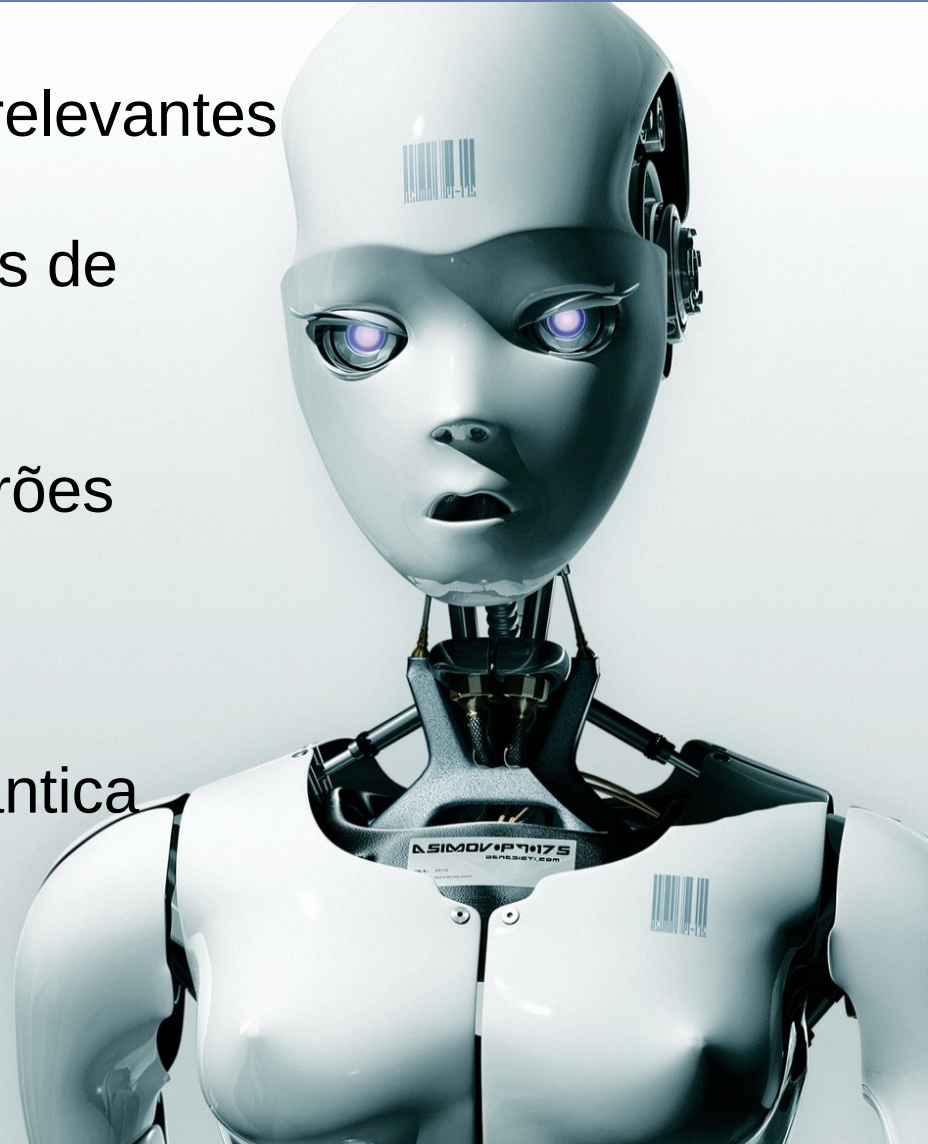
## Análise das perguntas e dos documentos

- Identificação de entidades nomeadas
  - . Aprendizagem automática;
  - . informação linguística
- Análise sintáctica
  - . Gramáticas de dependências, de restrições, “estatísticas”
- Análise semântica
  - . Frames, redes semânticas,



## Seleccção de frases relevantes

- Medidas heurísticas de similariedade
- Pesquisas por padrões de palavras
- Inferência sobre representação semântica das frases



## Ordenação das respostas

- Medidas heurísticas sobre
  - . frequência da resposta
  - . comprimento da resposta
  - . similariedade entre pergunta e resposta
  - . informação semântica da resposta

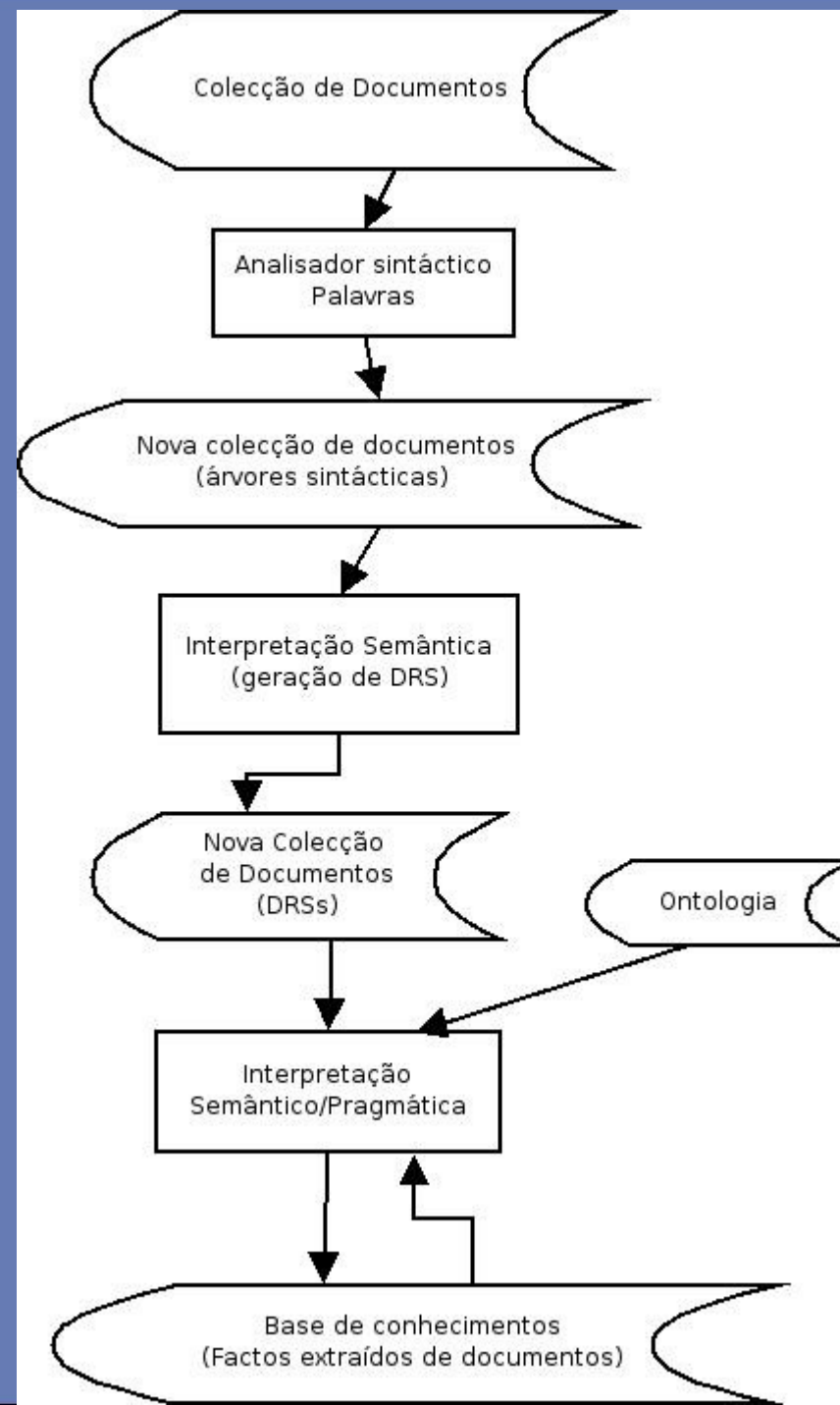


## Abordagem DI/ECT/UE

Dois módulos principais

- Extracção de informação  
Extraí informação dos documentos e cria uma base de conhecimentos;
- Recuperação de informação  
Processa a interrogação, acede à base de conhecimentos e gera as respostas.





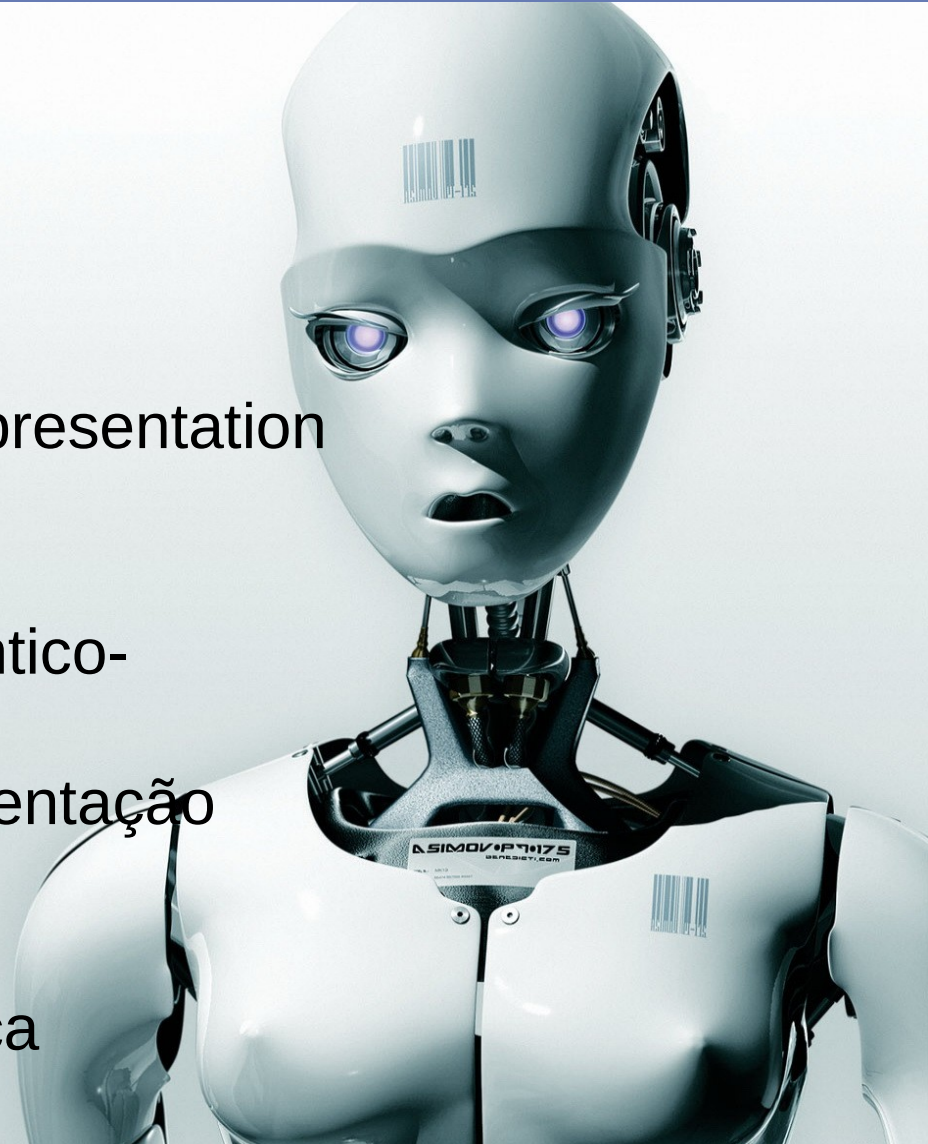
Análise lexical

Análise sintáctica

Análise semântica –  
DRT (Discourse Representation  
Theory)

Interpretação semântico-  
pragmática  
– ontologia + representação  
semântica

Representação lógica

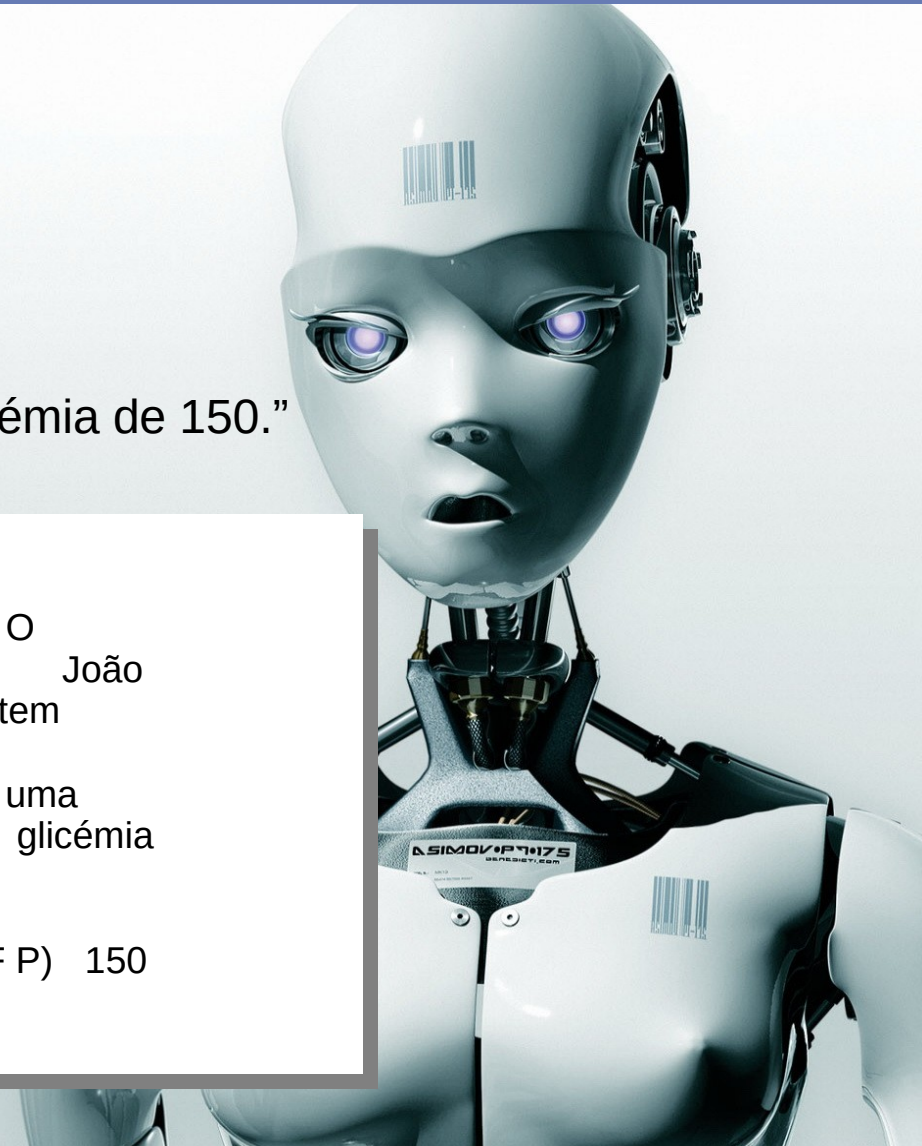


## Análise sintáctica

- “O João tem uma glicémia de 150.”

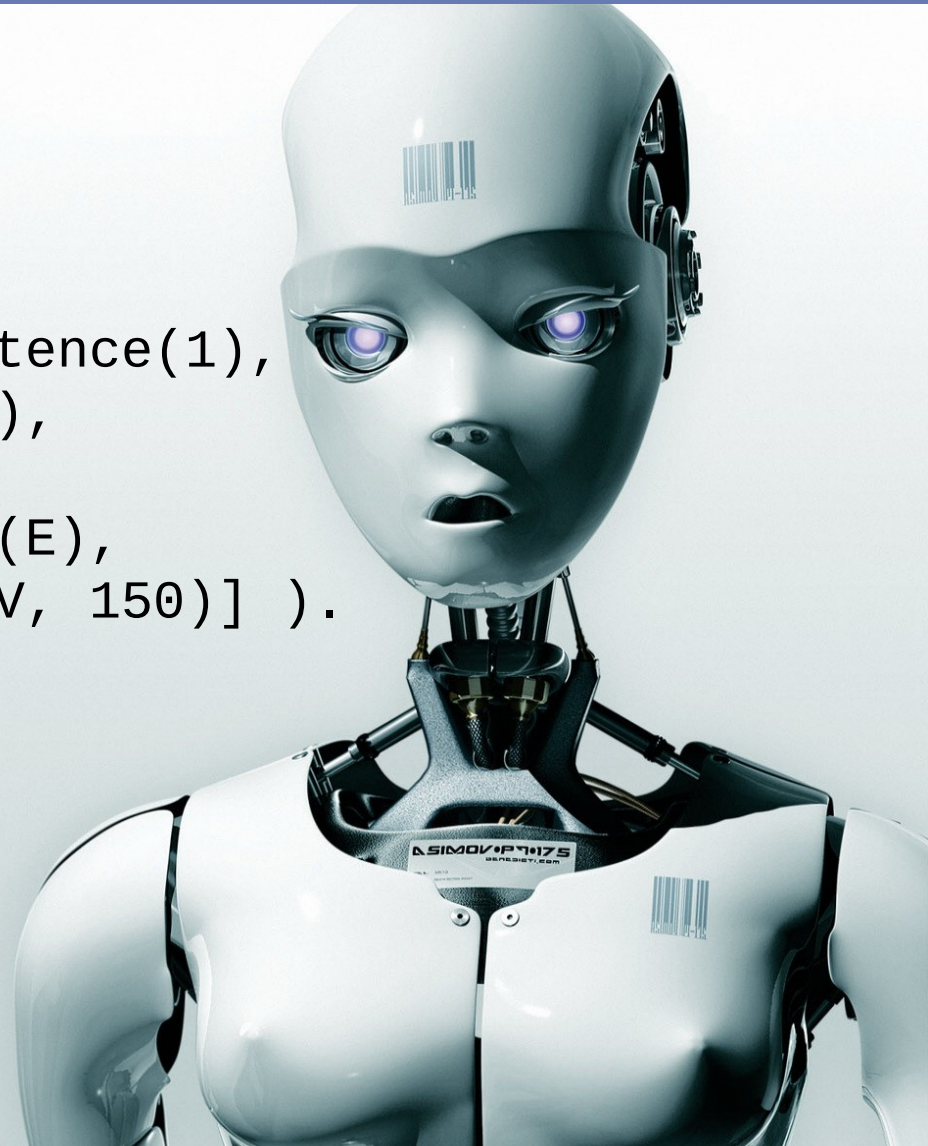
```

STA:fcl
=SUBJ:np
==>N:art('o' <artd> M S)    O
==H:prop('João' <hum> M S)   João
=P:v-fin('ter' PR 3S IND)    tem
=ACC:np
==>N:art('um' <arti> F S)    uma
==H:n('glicémia' <f-q> F S)  glicémia
==N<:pp
===H:prp('de') de
===P<:num('150' <card> M/F P) 150
=.
    
```



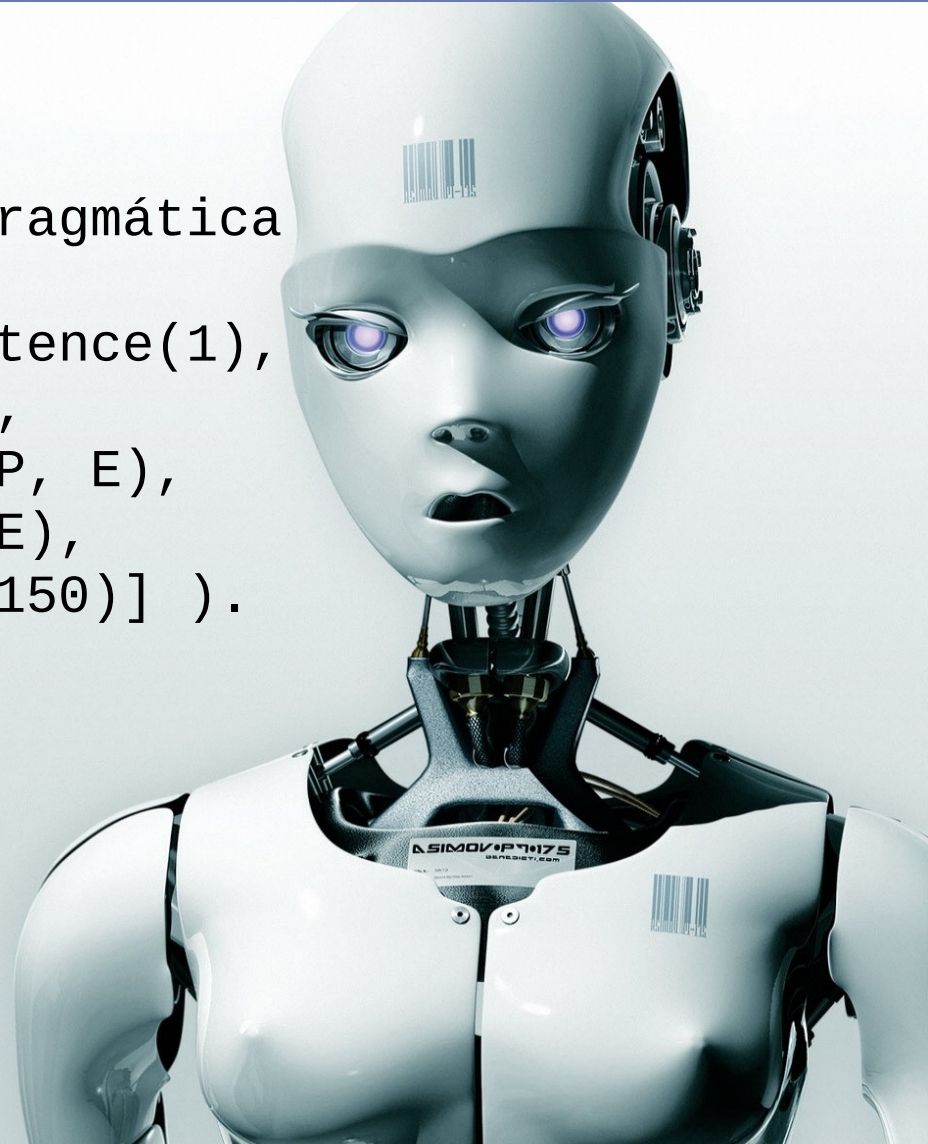
## Análise semântica

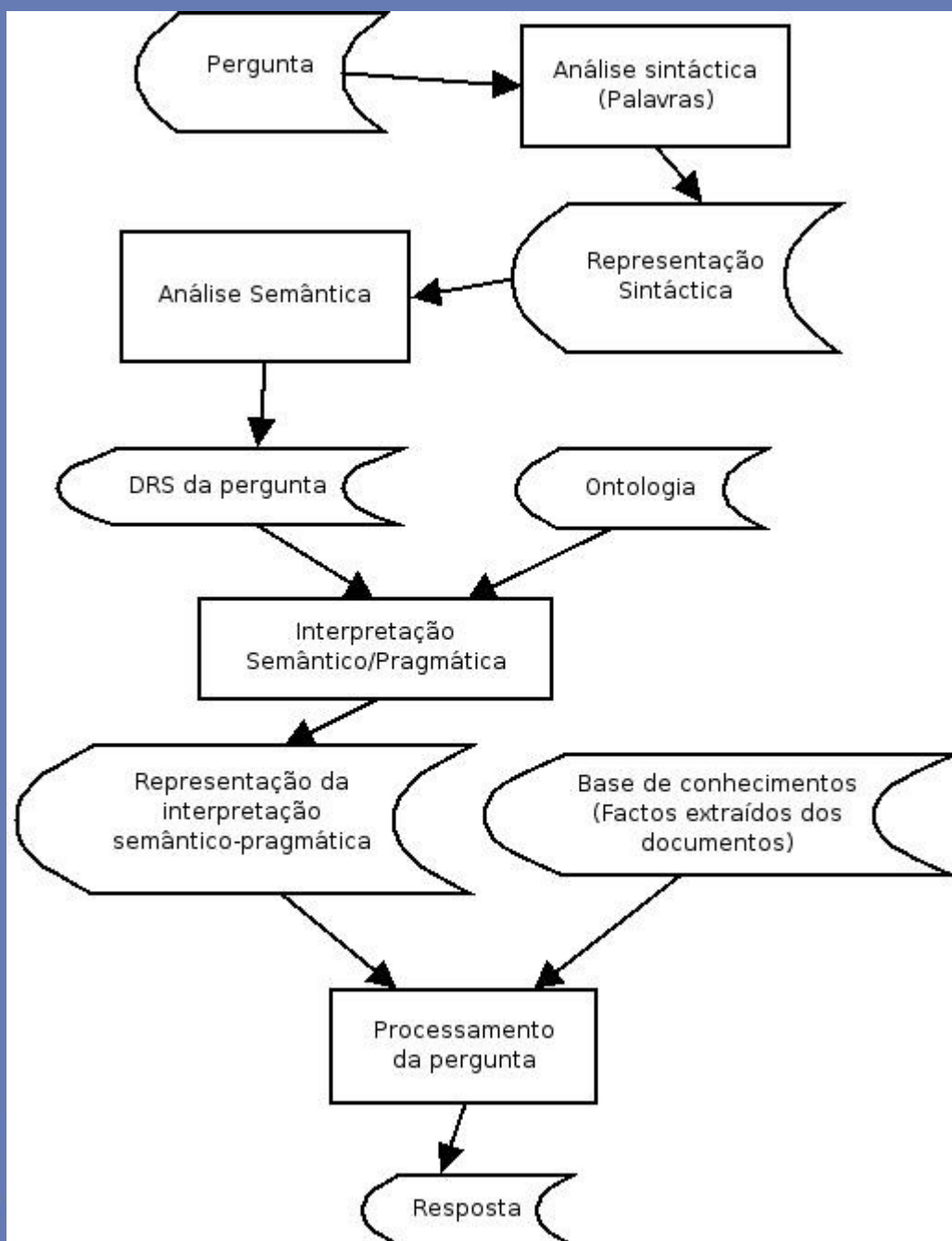
```
sentence(doc(1)-sentence(1),
[P,E,V] : [pessoa(P),
nome(P, 'João'),
tem(P, E), glicémia(E),
rel(de, E, V), num(V, 150)] ).
```



## Análise semântico-pragmática

```
sentence(doc(1)-sentence(1),
[P,E,V] :[pessoa(P),
nome(P, 'João'),tem(P, E),
exame(E), glicémia(E),
valor(E, V),num(V, 150)] ).
```





Análise sintáctica da interrogação

Análise semântica da interrogação

Interpretação semântico-pragmática da interrogação

Processamento da interrogação:

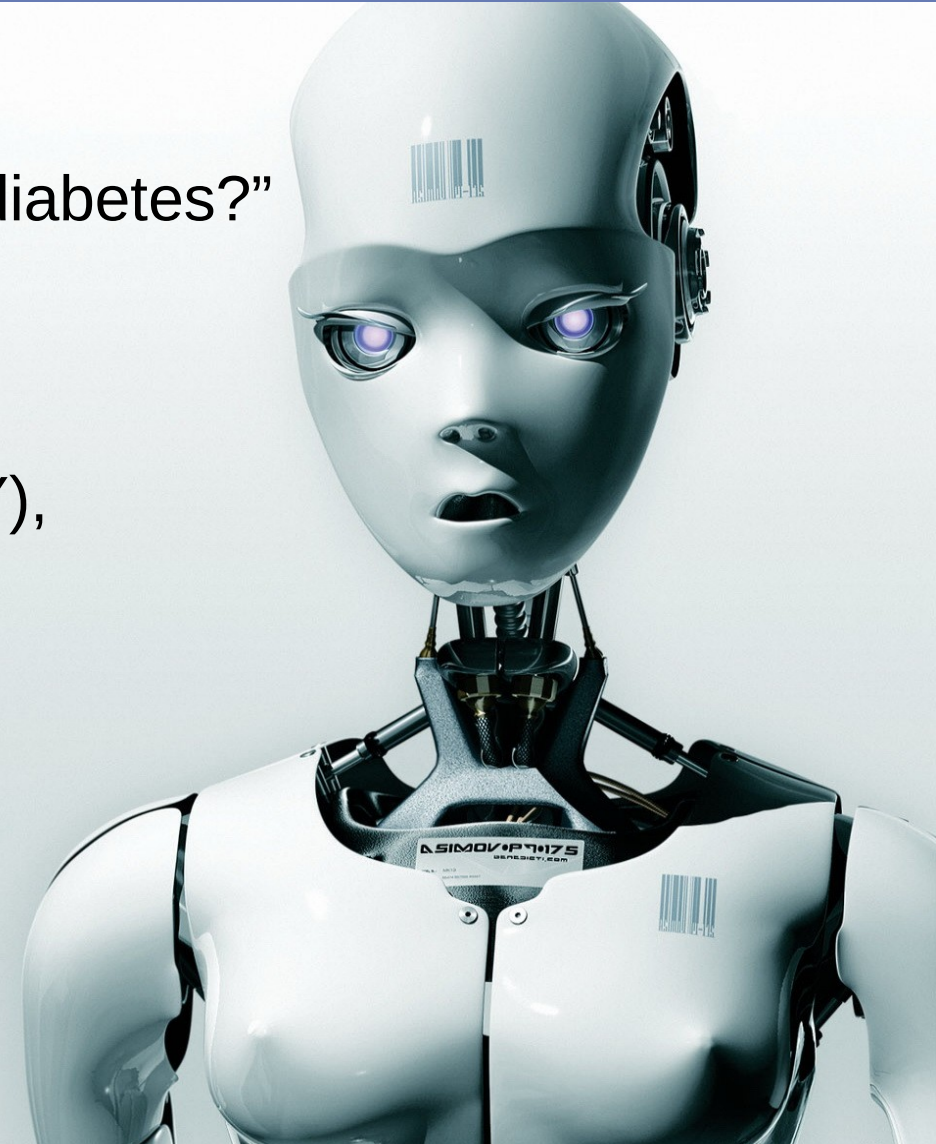
- Inferência na base de conhecimentos



“Quem tem diabetes?”

$X : [X, Y]$   
 pessoa( $X$ ),  
 temDoença( $X, Y$ ),  
 diabetes( $Y$ ).

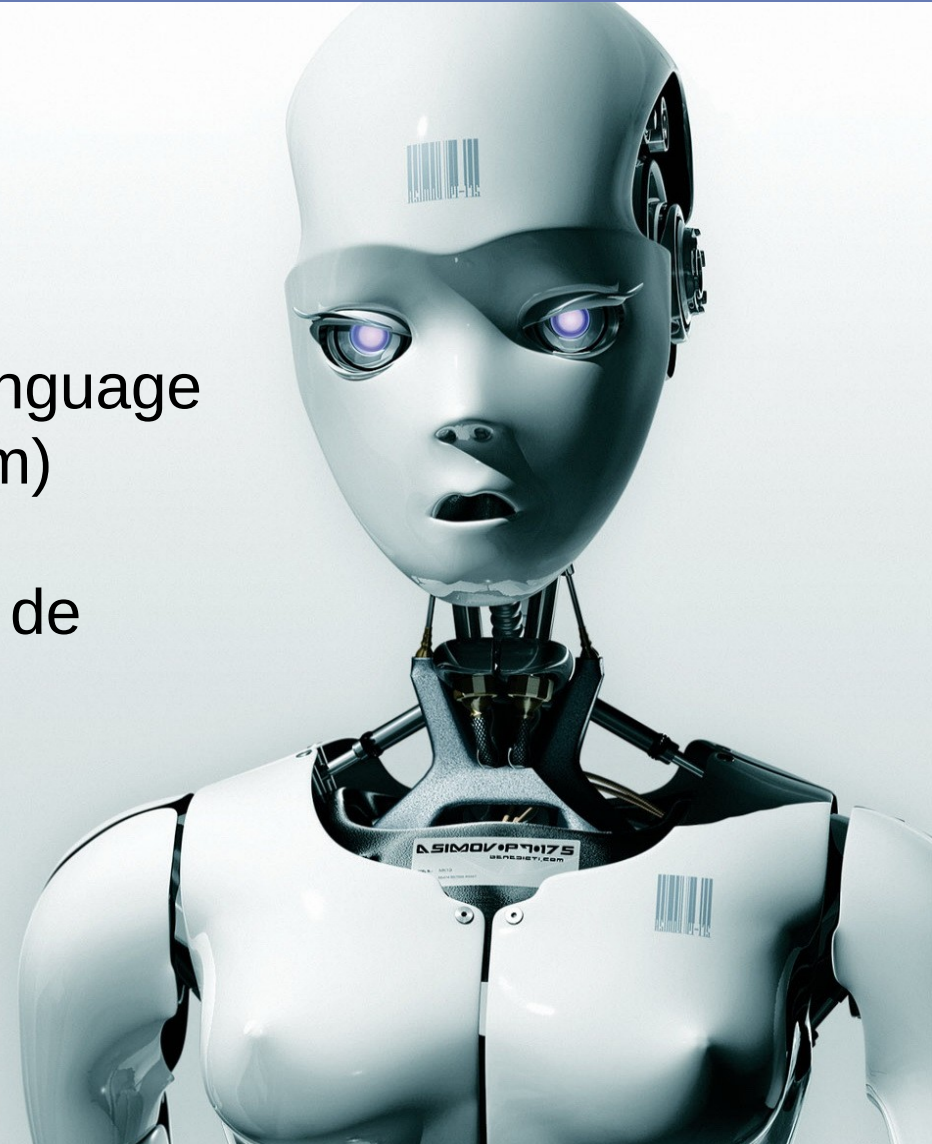
$X = \text{'João'}$



## Exemplos

CLEF (Cross Language  
Evaluation Forum)

Jornal “Público” de  
1994 e 1995

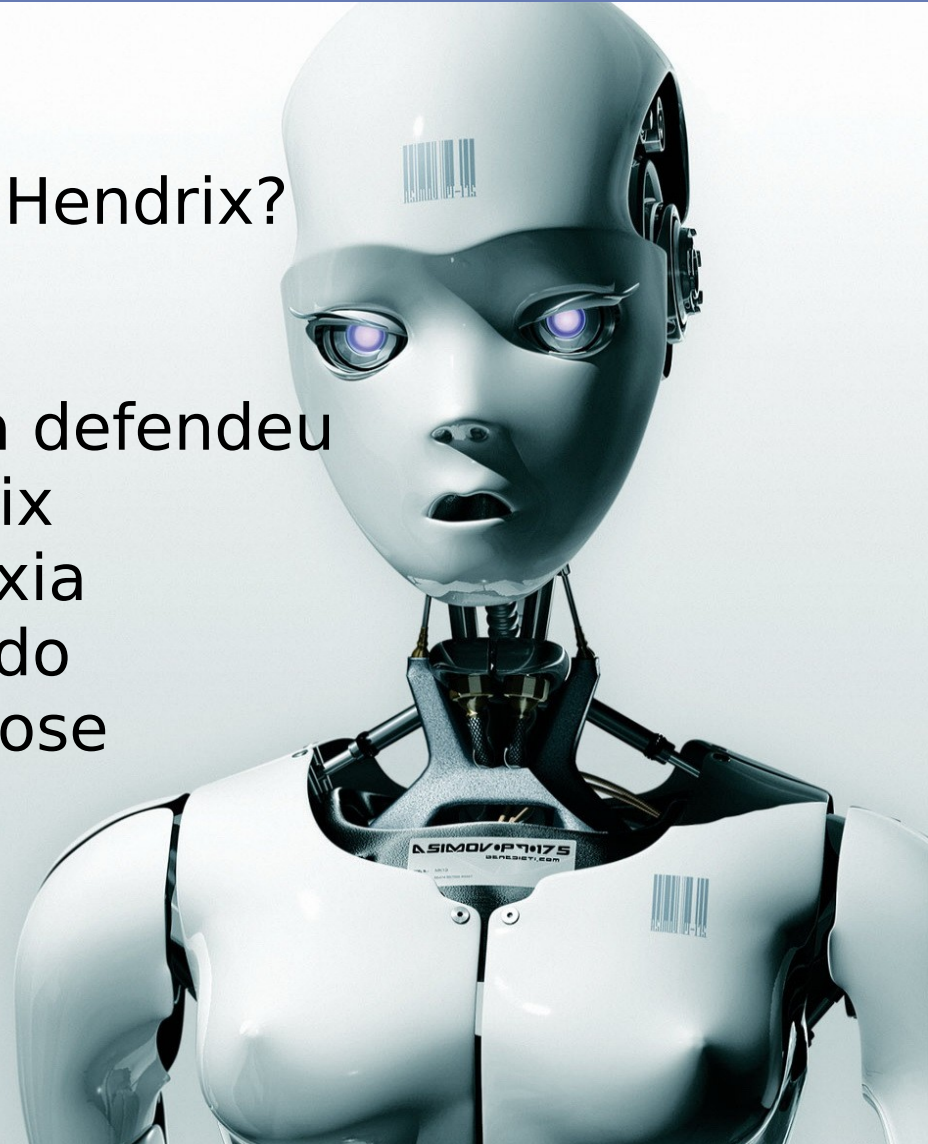


Como morreu Jimi Hendrix?

Texto:

Um patologista defendeu que Jimi Hendrix morreu de asfixia após ter ingerido álcool e uma dose excessiva de barbitúricos.

Resposta:  
De asfixia.



Quantos desempregados há na Europa?

A taxa de desemprego nos países industrializados deverá crescer 0,1 por cento, ou seja, 33 milhões de pessoas, 23 milhões dos quais na Europa ocidental.

Resposta: 23 milhões.



200 perguntas no CLEF

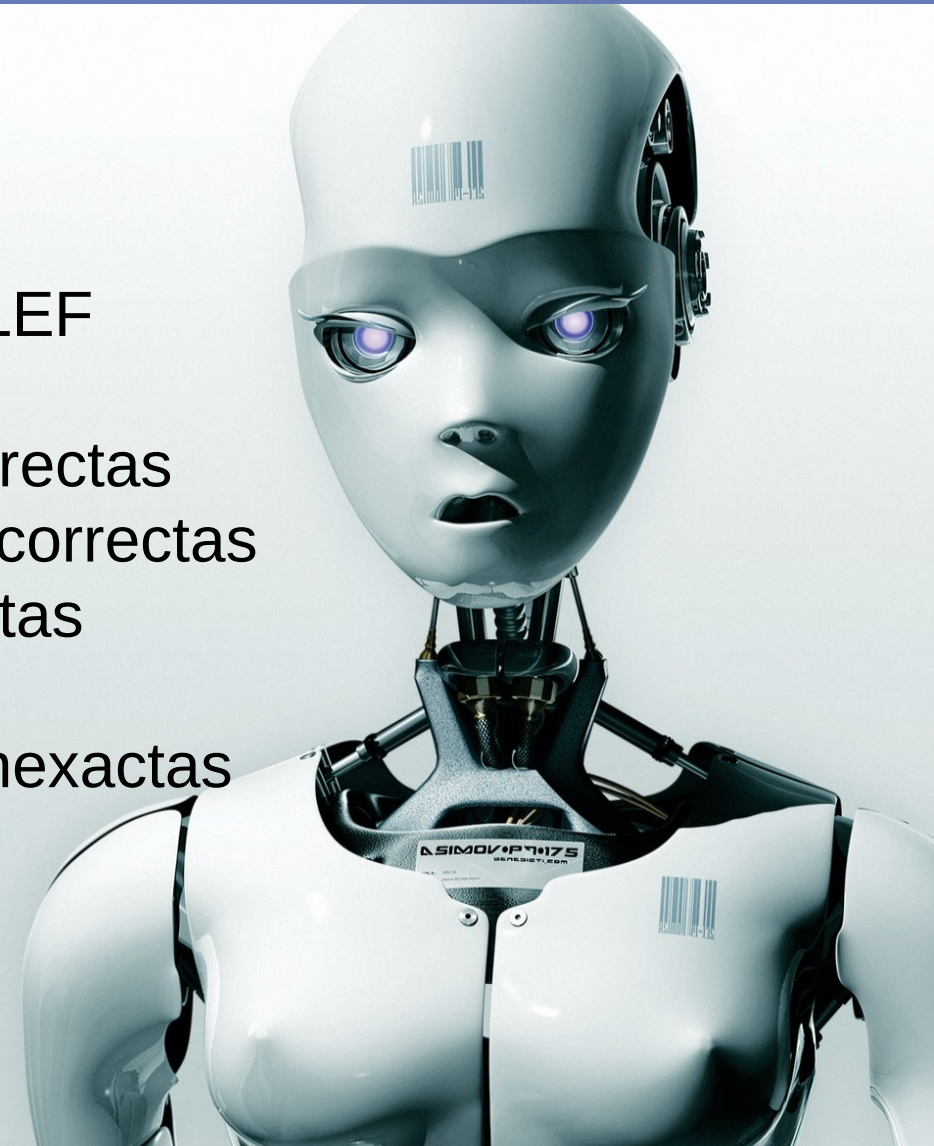
46.5% respostas correctas

47% de respostas incorrectas

1% respostas correctas

mas mal suportadas

5.5% de respostas inexactas



Principais problemas:

Sem resposta ou erradas

- Principalmente devido  
à falta de uma ontologia...

Possibilidades de resolução:

- Melhoria da ontologia

- ...



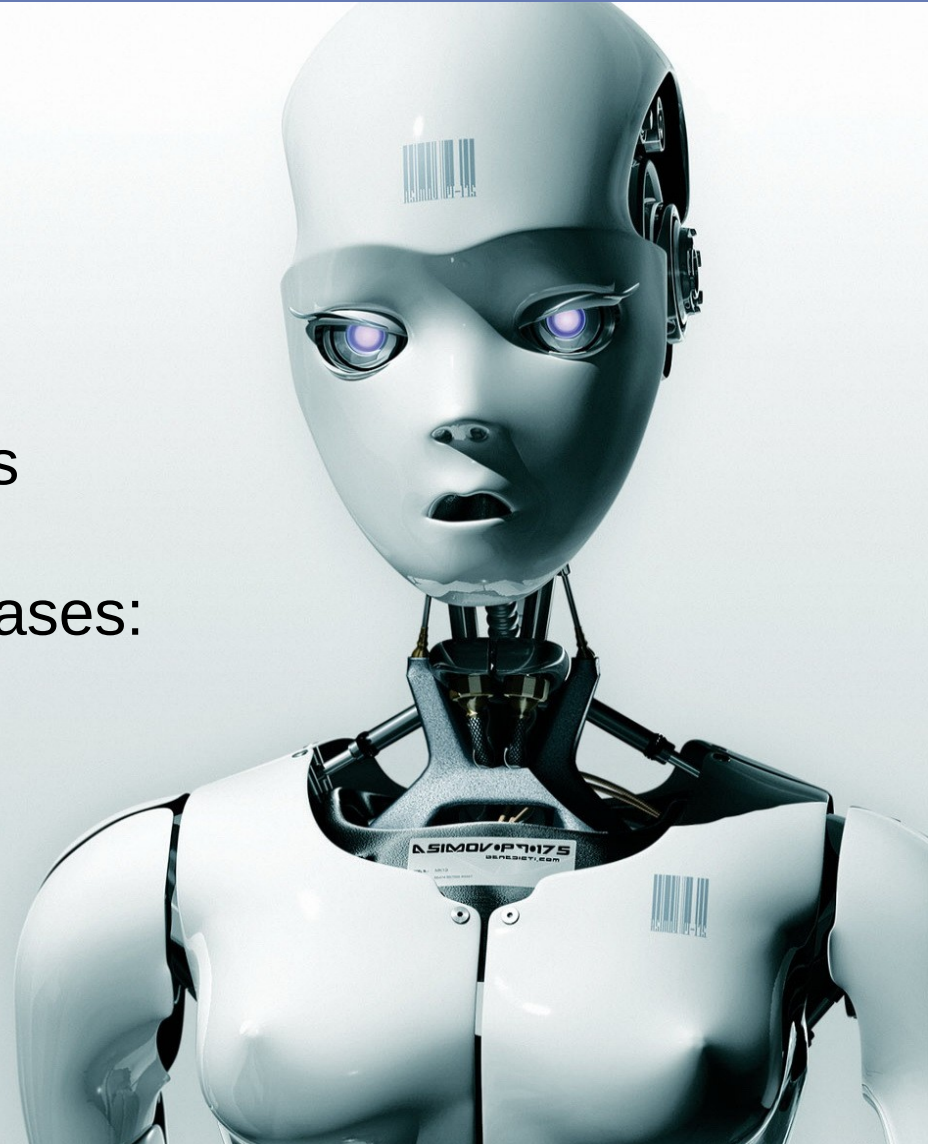
Principais problemas:

Respostas incorrectas

- Erros nas diversas fases:

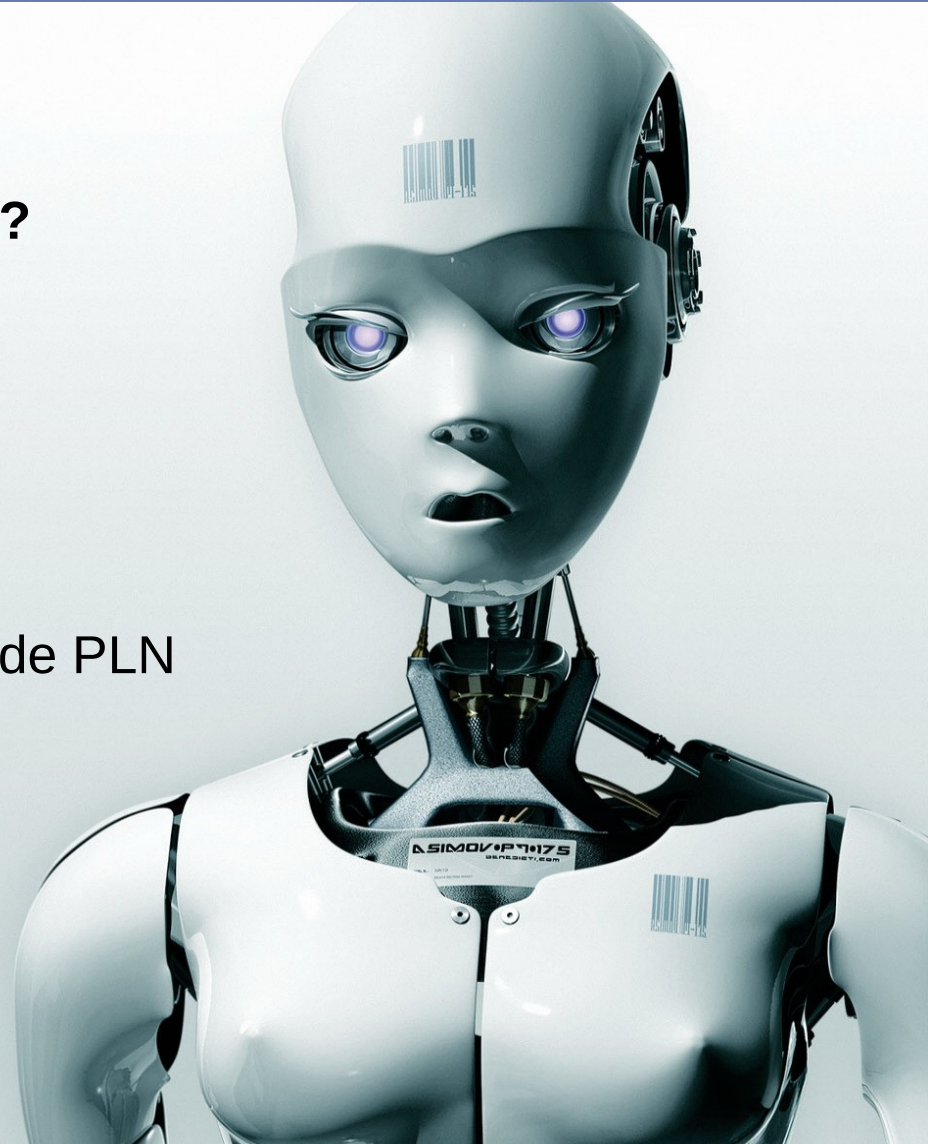
Sintaxe

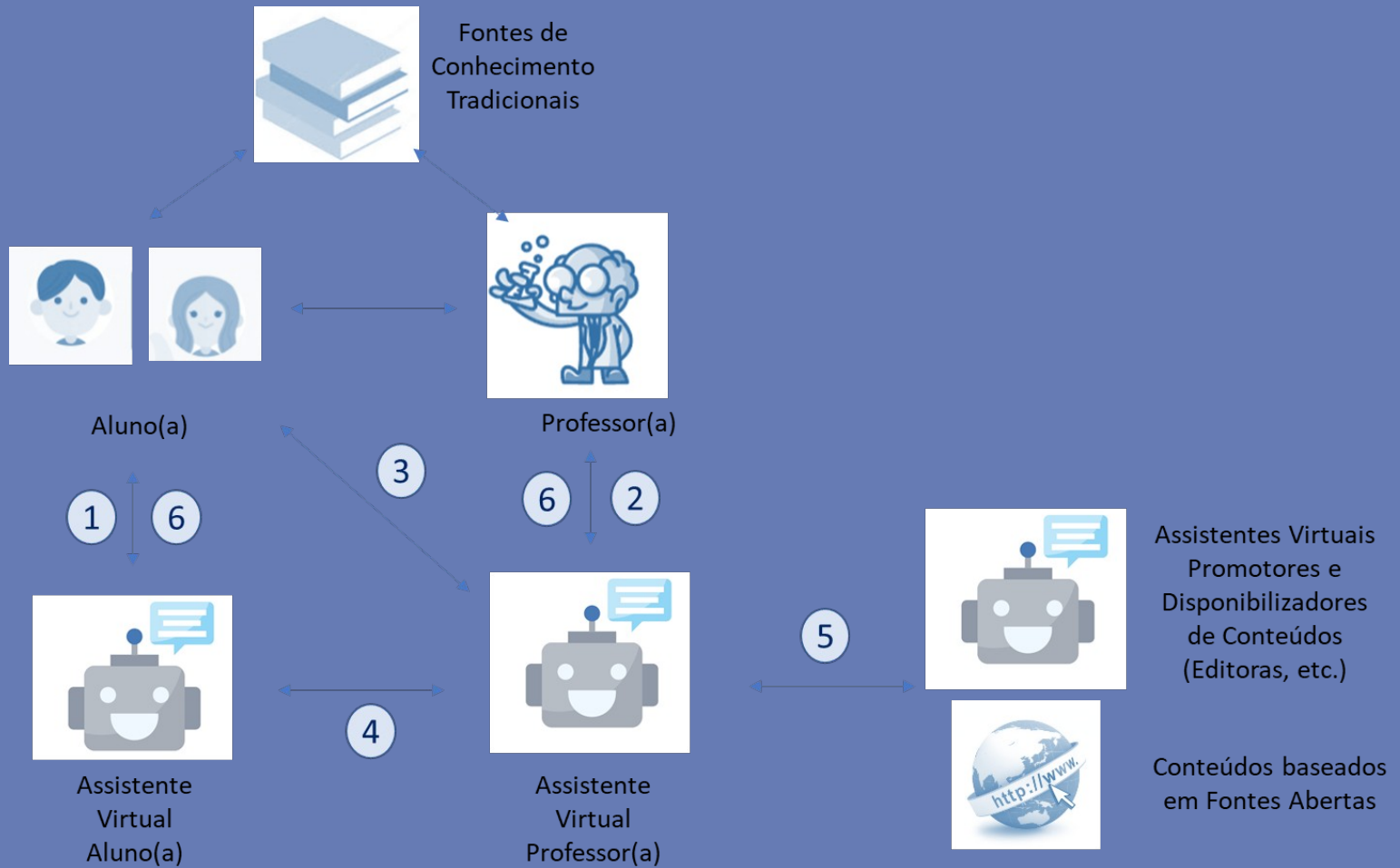
Semântica



## E a aplicabilidade no ensino/aprendizagem?

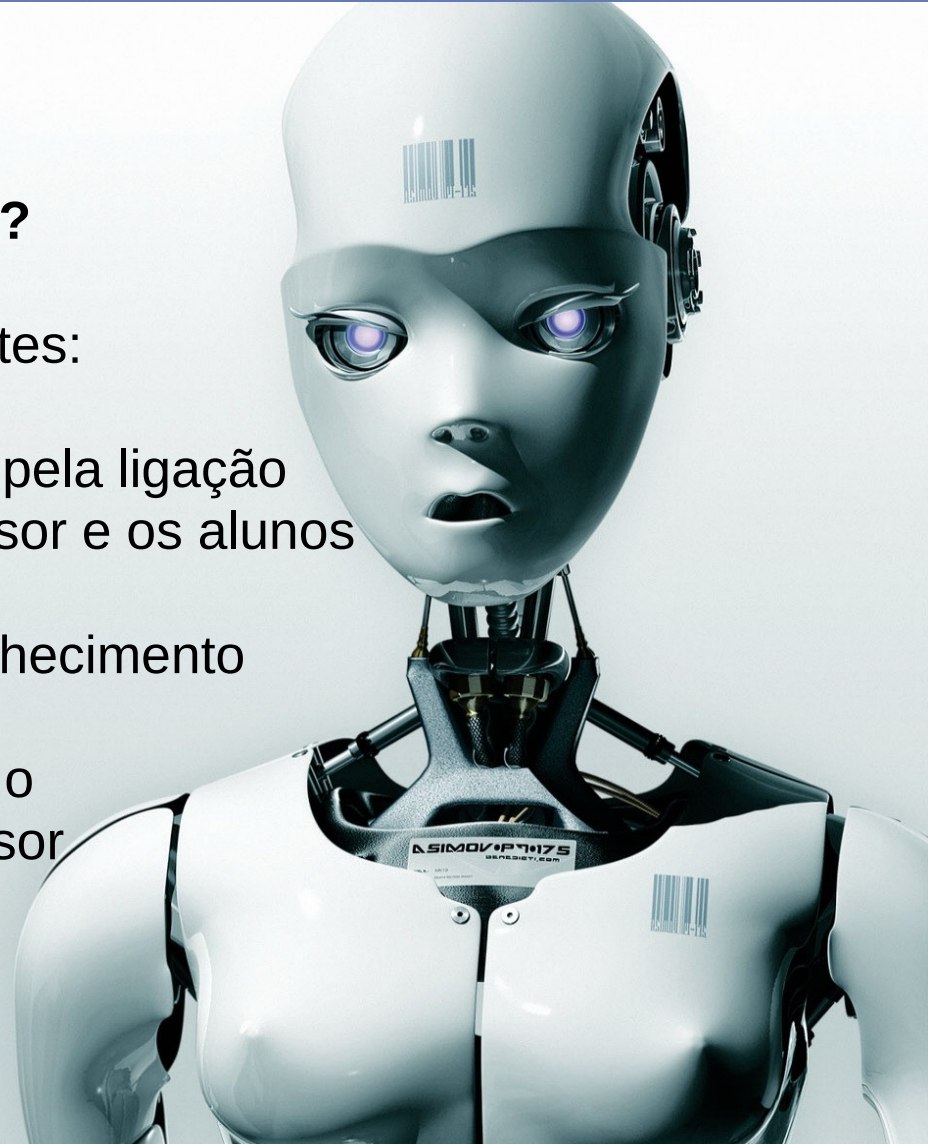
- Agentes inteligentes
  - Racionais
  - Pró-ativos
  - Colaborativos
  - Estado mental
  - Com capacidade de PLN (escrita e falada)
- Muito mais do que “chatbots”!!!



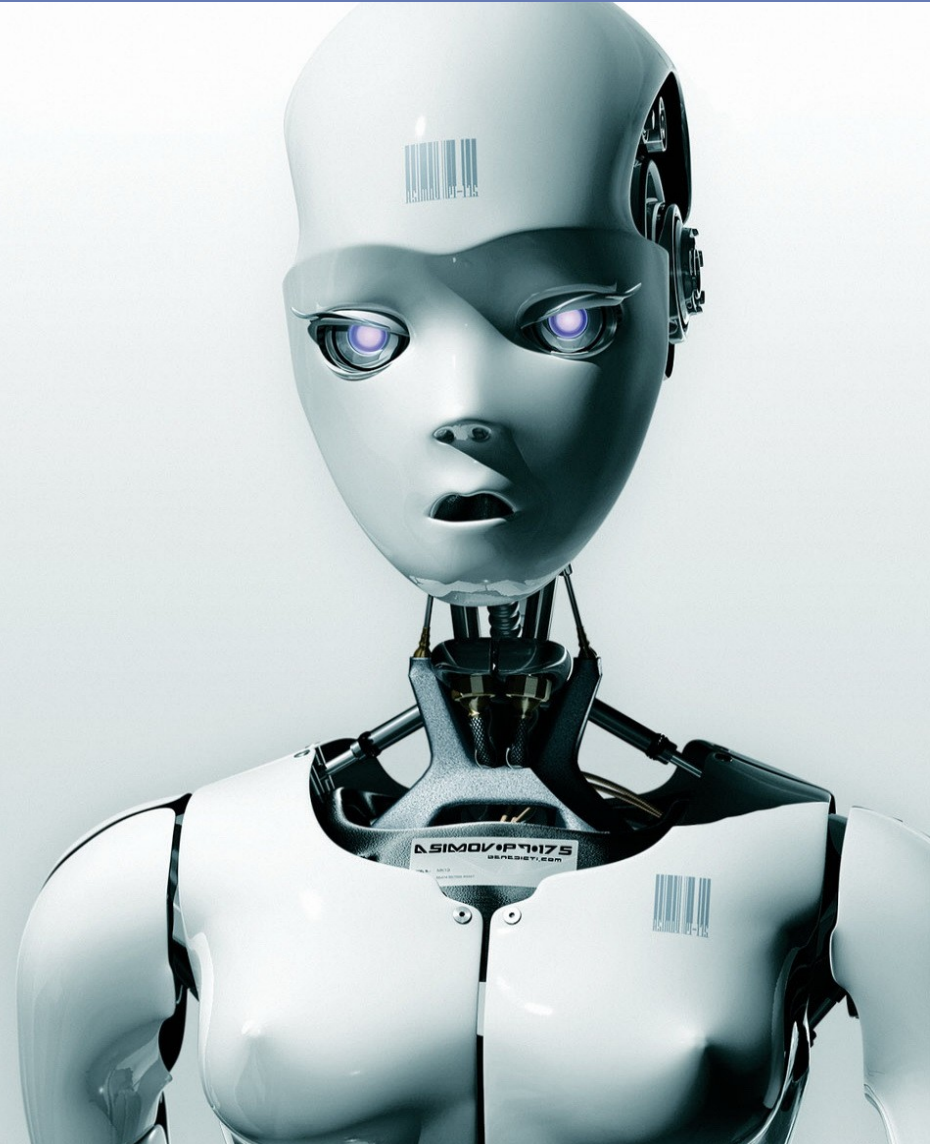


## E a aplicabilidade no ensino/aprendizagem?

- Ecosistema de agentes:
  - Agente professor
    - Responsável pela ligação entre o professor e os alunos
  - Agente aluno
    - Modela o conhecimento do aluno
    - Interage com o agente professor e com o aluno



- E agora, que futuro?



Alice no País das Maravilhas  
(Lewis Carrol, 1865)

Alice: Que caminho  
aconselhas a seguir?

Gato: Depende do lugar para  
onde queiras ir...





Questões?  
Comentários?



**u**  **évora**

ESCOLA DE CIÊNCIAS E TECNOLOGIA

