

# INF 5004 - INTELIGÊNCIA ARTIFICIAL AVANÇADA

## 1. INTRODUÇÃO E REVISÃO DE IA

- Motivação:
  - o Fronteiras entre o natural e o artificial
  - o Seria inteligência? Mente? Livre-arbítrio?
  - o Uma máquina de fato inteligente pode não ser humana? Pode não ter processos mentais conscientes? E inconscientes?
  - o O que distingue a ciência da computação convencional da IA?
  - o É possível simular processos mentais com algoritmos?
  - o A mente de um cérebro é apenas um programa?
  - o Quais são os modelos básicos do comportamento e como construir máquinas para simulá-las?
  - o Até que ponto a inteligência pode ser descrita recorrendo à avaliação de regras, inferências, à dedução e à computação de padrões?
  - o Qual é o desempenho das máquinas que simulam tais comportamentos através destes métodos?
- Fronteiras entre o comportamento inteligente e não inteligente:  
habilidades essenciais relacionadas à inteligência
  - o responder à situações de modo flexível
  - o tirar proveito de circunstâncias fortuitas
  - o perceber o sentido de mensagens contraditórias ou ambíguas
  - o reconhecer a importância relativa de elementos diferentes de uma dada situação
  - o encontrar similaridades entre situações apesar de diferenças que existam entre elas
  - o encontrar diferenças entre situações apesar das existentes
  - o sintetizar novos conceitos a partir de conceitos já conhecidos relacionando estes de outras formas
  - o imaginar novas ideias
- Críticas à IA:
  - o Penrose (Física)
  - o Dreyfus (Filosofia)
  - o Gerald Edelman (Biologia)
  - o [Searle]
- Histórico da IA:
  - o 1917-21: palavra "robô" aparece pela primeira vez (Karel Čapek, R.U.R.)
  - o 1950: jogo da imitação ou Teste de Turing
  - o 1956: reunião histórica que "cria" a IA
  - o Década 70: representação do conhecimento, raciocínio, busca
  - o Década 80: programação em lógica, raciocínio quantitativo. Sistemas especialistas, IAD, redes neuronais (neurais), proc. de imagem, NLP, reconhecimento de padrões, STI
  - o Década 90: SMA, Alife
- Definições de Inteligência:
  - o Da etnologia: algo que não se consegue explicar (Sócrates); idéia difícil de se definir (Turing)

- o Uma máquina inteligente precisa construir inteligência a partir de percepção e não apenas se apoiar em um modelo do mundo que possua dentro de si (Turing)
  - o Capacidade de descrever objetos e dados complexos e conhecimento e realiza operações de forma eficiente
  - o Parte computacional da habilidade de atingir metas no mundo (McCarthy) --> mas: seria apenas a parte computacional? outros tipos de inteligência? atingir metas não coloca as coisas excessivamente no nível de planejamento?
- Definição de IA (Coelho, 1999)
    - o Estudo do comportamento inteligente como processo computacional
    - o Área que visa compreender melhor a mente através da especificação de modelos computacionais
    - o Área que visa construir sistemas computacionais que realizam ações consideradas tradicionalmente mentais
    - o Área que visa estabelecer novos meios de representação baseados em computação, através dos quais o intelecto humano possa expressar-se de modo diferente com clareza e,
    - o Área que visa à investigação da natureza da inteligência e do espírito, recorrendo ao computador como sistema físico simbólico geral.

## 2. EXERCÍCIOS:

1) classifique as definições de IA que aparecem no livro de Russel e Norvig bem como as 5 do livro de Coelho (texto acima) levando em conta as seguintes dimensões:

- seres humanos vs. seres ideais
- pensamento vs. ação
- interesse em resultados teóricos vs. em aplicações práticas
- almejam obter consciência em computadores inteligentes ou não

2) idem para as seguintes definições de IA (outras fontes):

- “a collection of algorithms that are computationally tractable, adequate approximations of intractability specified problems” (Partridge, 1991)
- “the enterprise of constructing a physical symbol system that can reliably pass the Turing Test” (Ginsburg, 1993)
- “the field of computer science that studies how machines can be made to act intelligently” (Jackson, 1986)
- “the field of study that encompasses computational techniques for performing tasks that apparently require intelligence when performed by humans” (Tanimoto, 1990)
- “a very general investigation of the nature of intelligence and the principles and mechanisms required for understanding or replicating it” (Sharples *et al*, 1989)
- “the getting of computers to do things that seem to be intelligent” (Rowe, 1988)

3) busque e apresente (criticamente) outras definições de inteligência, IA e cognição

4) identifique 2 classes de problemas: a) onde o exercício da inteligência seja fundamental para encontrar a solução; b) onde a inteligência não é necessária para encontrar solução

5) para o item a) acima, descreva qual conhecimento é necessário e como representá-lo

6) pesquise a literatura sobre IA a fim de descobrir se as seguintes tarefas podem ser realizadas **atualmente** por computadores:

- jogar uma partida de um jogo de mesa do tipo ping-pong com eficácia
- dirigir um carro no centro de Porto Alegre
- jogar uma partida de um jogo de bridge em nível competitivo
- descobrir e provar novos teoremas matemáticos
- escrever uma estória engraçada **de propósito**
- fazer um aconselhamento legal competente a um especialista na área legal (advogado)
- traduzir inglês falado para português falado, em tempo real

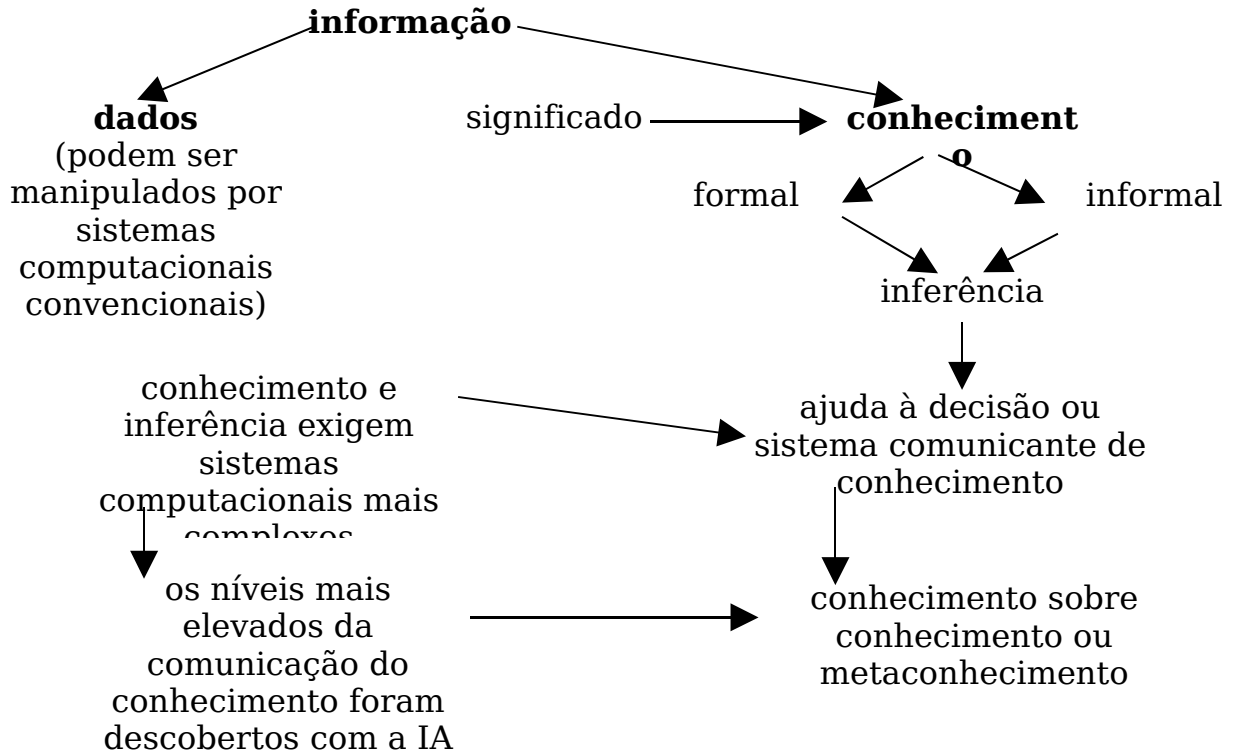
### 3. MÉTODO CIENTÍFICO CARTESIANO (DESCARTES)

- Mecanismos cognitivos são estudados através de redução (reducionismo): para compreender um mundo complexo, é necessário desmontá-lo em partes elementares mais simples (crítica: é suficiente?)
- É uma via analítica
- Influência do cartesianismos na IA:
  - o Teoria representacional da mente: cognição envolve o processamento de estados internos semanticamente interpretados
  - o Mecanismo de processamento de informação no cérebro é uma metáfora inspiradora deste sistema representacional → cognição é fundamentalmente computacional
  - o GPS (General Problem Solver): apoia-se no tripé: busca – memória – raciocínio
  - o Modelo cognitivo: busca heurística (auxiliada por métodos de planejamento + regras heurísticas): garantia de se encontrar a solução correta (se existir) em tempo finito
  - o Explicação para o pensamento: descrições simbólicas discretas (frames, redes semânticas, scripts, etc.)
- Década 60: aprendizagem = adaptação de heurísticas e estruturas de conhecimento ao espaço de buscas
- Década 70: natureza dos computadores e dos modelos de computação limitam algoritmos de resolução de problemas
- Década 80: Roger Schank (1983): sistemas especialistas não existem sem aprendizado e sem reorganização do conhecimento
- Ainda do reducionismo: outras metáforas derivadas do funcionamento do cérebro humano (perceptron, redes neurais)

- Pergunta ainda sem solução: é possível simular a mente em um programa de computador?
  - o Não: o espírito do homem será sempre incompreensível para o homem porque qualquer compreensão deste espírito deve ser sempre um subconjunto do conjunto de seu espírito
  - o Sim: pode-se identificar certos componentes da inteligência sem necessariamente as conhecer ou explicar seu funcionamento, tais como: a capacidade de computar ou calcular (+ simples), de correlacionar, de associar, de inferir e de extrapolar (+ complexa)

#### 4. ELEMENTOS DA IA

- IA = conhecimento + inferência (equação mais simples possível, mas muito adequada à sistemas especialistas)
- O que é conhecimento? Cooley (1988) propõe a seguinte relação entre conhecimento, dado e informação:



- Conhecimento deve ser explícito e manipulável → incompletude, inconsistência e incerteza
- Evolução da base de conhecimento no tempo: como garantir a consistência? Técnicas de manutenção da verdade (TMS), lógica e raciocínio não monotônico
- Hierarquias dos tipos de conhecimento (mais baixo para mais alto) segundo Gaines e Shaw (1986): ações e construções, experiência (conhecimento advindo das ações), hipóteses (modelar sobre experiências eficientes), analogias (comparação de modelos alternativos), abstrações dos modelos, transcendências (distinções sobre modelos abstratos)
- Organização do conhecimento (analogia com organização cerebral), 2 lados:
  - o Conhecimento analítico (lado esquerdo do cérebro): análogo biológico do modelo computacional de Von Neumann (processamento seqüencial, temporal e linear); processamento lógico e literal, espacial, quantitativo, local ligado à razão, ao pensamento, análogo ao hardware.
  - o Conhecimento advindo da intuição (lado direito): trata padrões, associações abstratas, inferências não lógicas e extrapolações criativas (modelo perseguido pelo projeto computadores de 5ª geração nos anos 80 e 90); processamento intuitivo e metafórico, qualitativo, global, ligado à emoção e sentimento análogo ao software.

- Visão diferente (Robert Ornstein na psicologia e Marvin Minsky na computação): seres humanos tem várias mentes, cada qual é um especialista (agente para Minsky) em uma área → “inteligências” são independentemente separadas e autônomas
- Elementos da IA (do ponto de vista teórico): lógica (monotônica, não monotônica, fuzzy); dependência conceitual; cognição; modelos sintáticos; semântica e pragmática da compreensão de uma linguagem natural; cinemática e dinâmica (para robótica); modelos do mundo (robôs), de alunos (S.T.I.), de interfaces, de diálogo, etc.
- Elementos da IA do ponto de vista da abstração, esquemas de representação do conhecimento (regras de redes semânticas; frames, scripts; lógicas) bem como os métodos para processamento destas representações (dedução, indução e abdução); modelos da compreensão da língua natural bem como suas representações (fonética, morfológicas e léxicas); tradução automática; reconhecimento e síntese de fala; modelos de raciocínio (temporal, espacial, causal, taxonômico, quantitativo, baseado em modelos e qualitativo); incerteza (lógica não monotônica, teorema de Bayes, regra de Dempster-Shafer); modelos de aprendizado (analogia, por exemplos, etc); métodos de busca heurística; arquitetura de máquina que imitam sistemas biológicos (redes neuronais, conexionismo e memória distribuída).
- Elementos da IA do ponto de vista de projetos e experimentos: sistemas em programação lógica, demonstração de teoremas; sistemas especialistas em domínios particulares (DENDRAL, MYCIN, etc); shells de sistemas especialistas; bancadas de desenvolvimento (KADS); implementações tipo Prolog, PARLOG, etc; sistemas de compreensão de linguagem natural; implementações de redes neuronais; sistemas que jogam xadrez, damas e jogos de estratégia; sistemas reconhecedores de voz; robôs estáticos e dinâmicos; metodologias e tecnologias como RBC, sistemas especialistas, algoritmos genéticos, aut. celulares, vida artificial, realidade virtual, estruturas auto-organizativas, sistemas auto-adaptativos
- Síntese destes elementos - criatura artificial:
  - o Módulos de entrada: visão, linguagem.
  - o Módulos do corpo central (caixas pretas): dedução e busca, planejamento, explicação, aprendizagem.
  - o Módulos de saída: fala, movimentos dos membros.

## **5. LEITURA:**

capítulo 2 de Sonho e Razão (Coelho, 1999) [obrigatória]  
 cap. 26 do livro Russel and Norvig [obrigatória]  
 capítulo 4 de Sonho e Razão (Coelho, 1999) [opcional]

## **6. PARA PRÓXIMA AULA:**

**apresentação estruturada (com slides ou no quadro)** das conclusões sobre cap. 26 Russel and Norvig (em grupo)  
 resolução do exercício 6 (ver acima)

## 7. IA: ESTUDO DA MENTE OU PROCESSAMENTO DE CONHECIMENTO?

- 1956: reunião que “cria” a IA:
  - estudo da mente: Newell
  - processamento de conhecimento: Feigenbaum
  - primeiras preocupações: representação (McCarthy), resolução de problemas (Newell e Simon), modelagem da memória (Minsky, Miller), reconhecimento de padrões (Greenblatt)
- modelos de computação anteriores à IA: máquina de Turing (1936), sistema de produção (Post, 1936 e 1943), cálculo Lambda (Church, 1936 e 1941), funções recursivas gerais (Kleene, 1936), lógica combinatória (Curry, 1958), máquina de Mealey (1955), máquina de Kleene (1956), autômatos celulares (von Neumann, 1956), máquina de Moore (1956)
- modelo de computação de McCarthy (1958): linguagem Lisp
- modelos de computação associados à IA: sistemas formais (Smullyan, 1961), algoritmos genéticos (Holland, 1975), lógica de cláusulas Horn (Kowalski, 1974), nodos conexionistas (Rumelhart e McClelland, 1976), planejamento (Chapman, 1987), Improviso (Agré e Chapman, 1987)
- tese da IA forte:
  - a inteligência humana pode ser replicada
  - as explicações das funções cognitivas são apoiadas por processos computacionais
- tese da IA fraca:
  - a inteligência humana pode, no melhor dos casos, apenas ser simulada
  - a meta para se atingir a inteligência artificial consiste em projetar máquinas que sejam capazes de exibir um comportamento inteligente
- questões formuladas na seção 1 podem ser resumidas na seguinte (Hofstadter, 1979): um sistema cognitivo (sistema lógico-formal, teoria matemática, cérebro como “base” da mente) pode ocupar-se de si mesmo? Esta pergunta remonta aos problemas propostos por Hilbert (1900), Russel e Whitehead (*Principia Mathematica* tenta derivar toda matemática a partir da lógica) e outras questões relativas às provas matemáticas:
  - existe um método (definitivo), ou seja um processo mecânico, que possa ser aplicado à uma declaração matemática, o qual possa dizer se esta pode ser provada?
    - não:
      - Hardy (1928): afirma mas não prova formalmente
      - Turing (1936)
        - superioridade da mente frente à dispositivos mecânicos
        - determinismo: modelagem de problemas complexos a partir de elementos (máquina de estados, autômatos, etc.)
    - a matemática é completa e consistente?
      - não (Gödel, 1931) e o Teorema da Incompletude: se S for consistente (não contém um teorema que seja a negação de um

outro), então existem sentenças verdadeiras que são teoremas em S e que não podem ser demonstradas:

- é possível escrever declarações/sentenças matemáticas que se referem a elas mesmas (do tipo “Eu estou mentindo”, “Esta sentença é falsa” ou “Esta declaração não pode ser provada”)
- tais declarações não podem ser demonstradas ou seja provadas como sendo verdadeiras (levariam à uma contradição) ou como sendo falsas (idem)
- logo, uma declaração deste tipo não pode ser demonstrada a partir dos axiomas o que prova que, por exemplo, a aritmética (foi mostrado que esta pode ser formulada apenas em termos de manipulação de símbolos - números e operadores, ou seja usa-se números para fazer afirmações sobre números) não é completa no sentido proposto por Hilbert (cada declaração - por exemplo  $2+2=4$  - pode ser provada como sendo V ou F)
- o exemplo usado por Gödel (“esta declaração não pode ser demonstrada”) é portanto uma declaração verdadeira (paradoxo ?) mas isto somente pode ser determinado por um observador externo (o próprio sistema axiomático não pode chegar à esta conclusão por si)
- o que isto tem a ver com a IA e a questão da mente?
  - um sistema formal S é um aparato simbólico (regras bem determinadas) o que implica em combinações simbólicas (teoremas de S) advindas das regras de S
  - a máquina de Turing é um sistema formal (Teorema de Gödel se aplica)
  - se o cérebro funcionar como uma máquina de Turing, permanecer consistente implica em uma limitação (por exemplo, não conseguirá demonstrar todas as sentenças verdadeiras da aritmética elementar)
  - mas, como o cérebro/mente (sistema S) consegue demonstrar (fora de S) que uma sentença (indemostrável no sentido de Gödel) é de fato verdadeira, então o cérebro/mente não é uma máquina de Turing, e um corolário seria: cérebro/mente não “pensa” algoritmicamente (Lucas, 1961; Penrose, 1989)
- qual é sua opinião a respeito disto?

## 8. LEITURA:

Searle x Churchland and Churchland (*Scientific American*. jan. 1990) [obrigatória]

The first AI debate (cap. 5 S. Franklin, *Artificial Minds*) [obrigatória]

Sloman (*Art. Intelligence* vol. 56: 355, 1992) [obrigatória pelo menos até seção 7]

Penrose (*The Emperor's New Mind*, caps. 1, 2) [opcional]

Dreyfus and Dreyfus (cap. 2 S.R. Graubard, *The Artificial Intelligence Debate*) [opcional]

Lucas (Minds, Machines and Gödel. . *Philosophy* 36: 112, 1961) ou cap. XV de *Gödel, Escher and Bach* (Hofstadter) [opcional]

Whitely (Minds, Machines and Gödel: a reply to Mr. Lucas. *Philosophy* 37: 61, 1962) [opcional]

### **9. PARA PRÓXIMA AULA:**

apresentação estruturada das conclusões artigo Churchland and Churchland (em grupo)

classifique cada uma das definições de IA vistas acima como pertencente à corrente da IA forte ou da IA fraca

## 10.CÉREBRO, MENTE E CONSCIÊNCIA

- Cérebro:
  - existe:
    - semelhança entre diferentes partes do cortex cerebral
    - semelhança entre cérebros de várias espécies
  - mas:
    - atividade mental não é semelhante entre as espécies
  - explicação:
    - os padrões de conectividade são diferentes (conceitualmente todos os livros se parecem mas cada padrão particular de uma combinação de símbolos forma um texto distinto)
  - ciências relacionadas: neurociências (cérebro físico), ciências cognitivas (mente)
- Definição de mente:
  - intelecto, pensamento, entendimento; alma, espírito (parte distinta do corpo)
  - capacidade mental ou psíquica; capacidade cognitiva e intelectual (em contraposição às emoções e intenções)
  - memória, recordações
  - concepção, imaginação
  - local do sujeito da consciência; sistema relativo aos fenômenos cognitivos e emocionais, experiência conscientes de um indivíduo
  - intenção, intuito, desígnio, disposição, desejo
- Minsky, Pinker: mente como sistema (conjunto) de órgãos
- Máquinas podem possuir uma mente? Animais? Há correntes que afirmam que apenas seres humanos possuem mente (mas: mesmo sem ter um definição do que é pensar, se pensa)
- Definições de consciência:
  - ter um sentimento ou conhecimento das sensações ou sentidos de si mesmo ou de coisas externas
  - ter um sentimento ou conhecimento do que uma coisa é ou de que uma coisa existe ou acontece
  - percepção de si como um ser pensante; saber o que se está fazendo e porquê
  - o conhecimento do que acontece ao redor de uma pessoa; totalidade dos pensamentos de uma pessoa; sentimentos e impressões; mente; lembrança do que se fez recentemente (logo, a discussão sobre consciência em máquina é inútil)
- Visões de alguns pesquisadores sobre o tema mente e consciência (2001):
  - Minsky: consciência não existe (em entrevista à ZDNET)
  - Dennett: máquina (robô) consciente é possível; dificuldade é apenas financeira
  - Davis: dificuldade é reproduzir o funcionamento do cérebro; progressos foram feitos apenas em relação a sensores e próteses artificiais

- Kurzweil: máquinas serão conscientes ainda na primeira metade do sec. XXI
- 

## 11.O PROBLEMA MENTE-CORPO

- Teorias
  - mente é tudo (visão mentalista)
  - mente nada mais é que um processo físico (visão materialista)
  - o mental e o material co-existem (visão dualista)
- Histórico
  - Descartes (sec. XVII)
    - paralelo entre autômatos e corpo humano: estímulo externo ativa receptores do sistema nervoso -> cérebro -> nervos atuadores (ex.: joelho)
  - Kant (sec. XVIII)
    - fundamentos metodológicos da ciência cognitiva (estrutura lógica do método de inferência de processos mentais)
    - pré Kant: visão empirista do modelo da mente (Hume por exemplo mostrava que o unicórnio não existe mostrando que as duas impressões que podem ser adquiridas por um dos nossos sentidos - no caso a visão - ou seja a do cavalo e a do chifre, não ocorrem juntas)
- Mente e cérebro
  - senso estrito:
    - mente é diferente do cérebro; a mente é o subconjunto das ações do cérebro que estão relacionadas com processamento de informações (em última análise, computação); a mente é o conjunto de processos que levam o cérebro de um estado a outro (Minsky)
  - teoria computacional da mente:
    - a informação permanece a mesma independente do meio que a transmite;
    - desejos e crenças (conceitos tipicamente associados com a mente) são informações representadas por símbolos
    - símbolos são armazenados em neurônios que são disparados por sensações
    - novas crenças e desejos são formados
    - um dado comportamento ocorre
- Debate atual:
  - a teoria computacional da mente resolve o problema mente-corpo?
  - críticos: Searle (1980) e Penrose (1994)

## 12. LEITURA:

- A. Damasio: A Passion for Reasoning. In: *Descartes' Error* (cap. 11) [obrigatória]
- A. Damasio: *Descartes' Error* (cap. 1 e 2) [opcional]
- S. Franklin: *Artificial Minds* (cap. 16) [opcional]
- R. Kurzweil: *The Age of Spiritual Machines* (cap. 3) [opcional]

M. Minsky: *The Society of Mind* (cap. 28.5 a 28.8) [opcional]  
O. Flanagan: *The Science of the Mind* (cap. 1) [opcional, sobre dualismo]

### **13.EXERCÍCIO:**

Dada a seguinte proposição: “Com certeza os computadores não podem ser inteligentes; eles podem apenas fazer aquilo que foi definido pelos programadores”. A segunda proposição é verdadeira? Ela implica a primeira?

Dada a seguinte proposição: “Com certeza os animais não podem ser inteligentes; eles podem apenas fazer aquilo que foi definido pelos seus genes”. A segunda proposição é verdadeira? Ela implica a primeira?

### **14.PARA PRÓXIMA AULA:**

apresentação estruturada sobre texto livro A. Damásio (em grupo)  
apresentação resolução do exercício acima