

Testes de Inteligência de Máquina

Shane Legg

IDSIA, Galleria 2, Manno-Lugano CH-6928, Suíça

shane@vetta.org

www.vetta.org/shane

Marcus Hutter

RSISE @ ANU e SML @ NICTA, Canberra, ACT, 0200, Austrália

marcus@hutter1.net

www.hutter1.net

Dezembro de 2007

Abstrato

Embora a definição e a mensuração da inteligência sejam claramente de importância fundamental para o campo da inteligência artificial, nenhuma existe um levantamento de definições e testes de inteligência de máquina. De fato poucos os pesquisadores estão cientes de alternativas ao teste de Turing e suas muitas derivados. Neste artigo, preenchemos essa lacuna fornecendo uma breve pesquisa muitos testes de inteligência de máquina que foram propostos.

[arXiv: 0712.3825v1 \[cs.AI\]](https://arxiv.org/abs/0712.3825v1) 22 de dezembro de 2007

Conteúdo

| | |
|------------------------------------|--------------------|
| 1. Introdução | 2 |
| 2 Teste de Turing e derivados | 2 |
| 3 testes de compressão | 3 |
| 4 Complexidade linguística | 4 |
| 5 Múltiplas habilidades cognitivas | 5 |
| 6 jogos competitivos | 5 |
| 7 Coleta de testes psicométricos | 5 |
| 8 Teste de Smith | 6 |
| 9 Teste C | 7 |
| 10 Inteligência universal | 7 |
| 11 Resumo | 9 |
| Referências | 10 |

1. Introdução

Apesar do sólido progresso em muitas frentes nos últimos 50 anos, a inteligência artificial é ainda um campo muito jovem, com muitas de suas maiores conquistas e algumas de suas problemas fundamentais, ainda por resolver. Do ponto de vista teórico, um dos os problemas mais fundamentais no campo é que o próprio conceito de inteligência permanece bastante sombrio. Isso é algo verdadeiro no contexto dos humanos, mas é especialmente verdade quando consideramos máquinas que podem ter completamente diferentes

sensores, corpos, capacidades cognitivas e vivem em diferentes ambientes para nós mesmos. O que "inteligência" significa para uma máquina? Talvez a primeira tentativa de responder esta questão, e certamente a única tentativa que a maioria dos pesquisadores conhece, é o famoso jogo de imitação de Alan Turing [33]. Turing reconheceu como é difícil seria definir explicitamente a inteligência e, assim, tentar contornar o problema completamente. Embora tenha sido uma jogada inteligente, ela nos deixa com um teste de máquina inteligência que nos diz quase nada sobre o que realmente é a inteligência e, portanto, é de pouca utilidade como fundamento, teórico ou prático, para nossa pesquisa.

Desde então, alguns pesquisadores ousados tentaram resolver esse difícil problema de maneira mais satisfatória, propondo várias definições e testes de máquina inteligência. De um modo geral, essas propostas foram ignoradas pela comunidade. De fato, até onde sabemos, nenhuma pesquisa geral de testes e definições de inteligência para máquinas já foi publicada.

Consideramos que ignorar uma questão tão fundamental quanto a definição de máquina a inteligência é um erro grave. Em qualquer ciência, questões que envolvem questões fundamentais definições e métodos de medição desempenham um papel central e formam a base em que avanços teóricos são construídos e avanços práticos são medidos. Se quisermos realmente avançar como um campo nos próximos 50 anos, precisaremos retornar a este mais central dos problemas, a fim de garantir o que é inteligência artificial e o que ele visa. Como primeiro passo nessa direção, é necessário que os pesquisadores sejam pelo menos ciente das muitas alternativas aos testes de Turing que foram propostas. Neste artigo, esperamos atender parcialmente a essa necessidade, fornecendo a primeira pesquisa geral de testes e definições de inteligência de máquina.

2 Teste de Turing e derivados

A abordagem clássica para determinar se uma máquina é inteligente é a chamada Teste de Turing [33] que tem sido amplamente debatido nos últimos 50 anos [26]. Turing percebeu o quão difícil seria definir diretamente a inteligência e, assim, tentou contornar o problema criando seu agora famoso jogo de imitação: se juízes humanos não podem efetivamente discriminar entre um computador e um humano através de conversas teletipadas, devemos concluir que o computador é ligante.

Embora simples e inteligente, o teste atraiu muitas críticas. Bloquear e Searle argumenta que passar no teste não é suficiente para estabelecer inteligência [3, 28, 7].

2

Essencialmente, ambos argumentam que uma máquina pode parecer inteligente sem ter alguma "inteligência real", talvez usando uma tabela muito grande de respostas para questões. Embora tal máquina possa ser impossível na prática devido à vasta tamanho da tabela necessária, não é logicamente impossível. Nesse caso, um ininteligente máquina poderia, pelo menos em teoria, passar consistentemente no teste de Turing. Alguns consideram isso para colocar a validade do teste em questão. Em resposta a esses desafios, versões ainda mais exigentes do teste de Turing foram propostas, incluindo o Teste de Turing total [11], teste de Turing verdadeiramente total e teste de Turing invertido [35]. Dowe argumenta que o teste de Turing deve ser estendido, incluindo que o O agente tem uma representação compactada da área de domínio, excluindo a consulta argumentos de contador de tabela [6]. É claro que esses ataques ao teste de Turing podem ser aplicado a qualquer teste de inteligência que considere apenas o comportamento externo de um sistema, isto é, a maioria dos testes de inteligência.

Uma crítica mais comum é que não é necessário passar no teste de Turing para estabelecer inteligência. Geralmente esse argumento é baseado no fato de que o teste requer que uma máquina tenha um modelo altamente detalhado de conhecimentos e padrões humanos de pensamento, tornando-o um teste de humanidade em vez de inteligência [9, 8]. De fato, mesmo pequenas coisas como fingir ser incapaz de executar aritmética complexa rapidamente ou falsificar erros de digitação humana se tornam importantes, algo que contraria claramente

o objetivo do teste.

O teste de Turing também tem outros problemas. Os atuais sistemas de IA são um longo caminho de ser capaz de passar em um teste de Turing irrestrito. Do ponto de vista prático isso significa que o teste de Turing completo é incapaz de oferecer muita orientação ao nosso trabalho. De fato, mesmo que o teste de Turing seja o teste mais famoso de inteligência de máquina, quase nenhuma pesquisa atual em inteligência artificial é direcionada especificamente para ser capaz de passar. Infelizmente, basta restringir o domínio da conversa em o teste de Turing para facilitar o teste, como é feito na competição Loebner [22], é não suficiente. Com possibilidades restritas de conversação, o Loebner de maior sucesso os participantes estão ainda mais focados em fingir falibilidade humana, em vez de qualquer coisa assemelhando-se a inteligência [15]. Talvez uma alternativa melhor, então, seja testar se um A máquina pode imitar uma criança (veja, por exemplo, os testes descritos nas Seções 4 e 5). Finalmente, o teste de Turing retorna resultados diferentes, dependendo de quem o ser humano julga estamos. Sua falta de confiabilidade levou, em alguns casos, a máquinas claramente não inteligentes classificado como humano, e pelo menos uma instância de um humano que realmente falha em um Turing teste. Quando questionado sobre este último, um dos juízes explicou que “nenhum ser humano o ser teria essa quantidade de conhecimento sobre Shakespeare” [29].

3 testes de compressão

Mahoney propôs uma solução particularmente simples para os problemas de aprovação ou falha binária com o teste de Turing: Substitua o teste de Turing por um teste de compressão de texto [23]. Em essência, isso é um pouco semelhante a um "teste de Cloze", em que o indivíduo

3

a apreensão e o conhecimento em um domínio são estimados fazendo-os adivinharem palavras de uma passagem de texto.

Embora a compactação de texto simples possa ser realizada com frequências de símbolos, o a compressão resultante é relativamente fraca. Usando modelos mais complexos que capturam recursos de nível superior, como aspectos gramaticais, os melhores compressores são capazes de comprimir o texto para cerca de 1,5 bits por caractere em inglês. No entanto, os humanos, que também pode fazer uso do conhecimento geral do mundo, a estrutura lógica do argumento etc., são capazes de reduzir isso para cerca de 1 bit por caractere. Assim, a compressão estatística fornece uma medida facilmente calculada de quão completo o modelo de uma máquina da linguagem, raciocínio e conhecimento de domínio são relativos a um humano.

Para ver a conexão com o teste de Turing, considere um teste de compressão baseado em corpus de diálogo muito grande. Se um compressor pudesse funcionar extremamente bem em tais um teste, isso é matematicamente equivalente a poder determinar quais sentenças são prováveis em um dado momento do diálogo e não são (para a equivalência de compressão e previsão, veja [2]). Assim, como falha no teste de Turing ocorre quando um máquina (ou pessoa!) gera uma sentença que seria improvável para um humano, um desempenho extremamente bom na compactação de diálogo implica a capacidade de passar um Teste de Turing.

Um desenvolvimento recente nessa área é o Prêmio Hutter [17]. Neste teste, o corpus é um extrato de 100 MB da Wikipedia. A ideia é que isso represente uma amostra razoável de conhecimento mundial e, portanto, qualquer compressor que possa executar muito bem neste teste deve ter um bom modelo não apenas do inglês, mas também do mundo conhecimento em geral.

Uma crítica aos testes de compressão é que não está claro se um poderoso o compressor se traduziria facilmente em uma inteligência artificial de uso geral.

4 Complexidade linguística

Uma abordagem mais lingüística é adotada pelo projeto HAL na empresa Artificial Intelligence NV [32]. Eles propõem medir o nível de conversação de um sistema usando técnicas desenvolvidas para medir a capacidade linguística das crianças. Esses métodos examinam coisas como tamanho do vocabulário, duração das expressões, tipos de esponja, complexidade sintática e assim por diante. Isso permitiria que os sistemas fossem "... atribuiu uma idade ou um nível de maturidade ao lado de sua avaliação binária do teste de Turing de 'inteligente' ou 'não inteligente' " [31]. Como eles consideram a comunicação o base de inteligência e o teste de Turing para ser um teste válido de inteligência de máquina, na sua opinião, a melhor maneira de desenvolver a inteligência é refazer a maneira pela qual ocorre o desenvolvimento linguístico humano. Embora não se refiram explicitamente às suas medida linguística como teste de inteligência, porque mede o progresso em direção a o que eles consideram um teste de inteligência válido, ele atua como um.

4

Page 6

5 Múltiplas habilidades cognitivas

Uma abordagem mais ampla de desenvolvimento está sendo adotada pelo projeto Joshua Blue da IBM [1]. Neste projeto, eles medem o desempenho de seu sistema considerando uma ampla testes linguísticos, sociais, de associação e de aprendizagem. O objetivo deles é passar primeiro o que eles chamam de "teste de Turing para bebês", ou seja, para desenvolver um sistema de IA que possa passar quando criança em uma configuração semelhante ao teste de Turing. Até agora, este teste não é totalmente especificado.

Outra empresa que segue uma abordagem semelhante de desenvolvimento baseada em O desempenho do sistema através de uma ampla gama de testes cognitivos é o projeto a2i2 em AI adaptativa [34]. Em vez de inteligência no nível da criança, seu objetivo atual é trabalhar em direção a um nível de desempenho cognitivo semelhante ao de um mamífero pequeno. A ideia é que mesmo um pequeno mamífero possui muitas das principais habilidades cognitivas necessário para que a inteligência em nível humano trabalhe em conjunto de maneira integrada. Enquanto isso pode ser útil para orientar o desenvolvimento de inteligência moderada, não se sabe se escalará para níveis mais altos de inteligência. Os testes específicos em uso não foram publicados.

6 jogos competitivos

O método da relação de Turing de Masum et al. tem mais ênfase em tarefas e jogos ao invés de testes cognitivos. Eles propõem que "... indo bem em uma ampla gama de tarefas é uma definição empírica de "inteligência". "[24] Para quantificar isso, eles procuram para identificar tarefas que medem habilidades importantes, admita uma série de estratégias que qualitativamente diferentes e reproduzíveis e relevantes ao longo de um período período de tempo. Eles sugerem um sistema de medição de desempenho através de pares comparações entre sistemas de IA semelhantes àquelas usadas para classificar jogadores no sistema internacional de classificação de xadrez. A principal dificuldade, no entanto, que os autores reconhecer é um desafio aberto, é descobrir quais devem ser essas tarefas e quantificar quão amplo, importante e relevante cada um é. Em nossa opinião, estes são alguns dos problemas mais centrais que devem ser resolvidos ao tentar construir um teste de inteligência e, portanto, essa abordagem está incompleta em seu estado atual.

7 Coleta de testes psicométricos

Uma abordagem chamada AI psicométrica tenta resolver o problema do que testar de uma maneira pragmática. Na visão de Bringsjord e Schimanski, "algum agente é inteligente se, e somente se, for excelente em todos os testes estabelecidos e validados de inteligência." [4] Mais tarde, ampliaram isso para incluir também "testes de arte e literatura

criatividade, capacidade mecânica e assim por diante.” Com isso como objetivo, suas pesquisas são focado na construção de robôs que podem ter bom desempenho em testes psicométricos padrão

5

 Page 7

projetado para humanos, como a Wechsler Adult Intelligent Scale e o Raven Progressive Matrices.

Por mais eficazes que sejam esses testes para humanos, eles parecem inadequados para medir inteligência da máquina, pois são altamente antropocêntricos e incorporam pressupostos básicos sobre o assunto do teste que provavelmente será violado por computadores. Por exemplo, considere a suposição fundamental de que o sujeito do teste não é simplesmente uma coleção de algoritmos especializados projetados apenas para responder a perguntas comuns de teste de QI. Embora isso seja obviamente verdadeiro para um humano, ou mesmo um macaco, pode não ser verdade para um computador. O computador não poderia ser nada além de uma coleção de algoritmos projetados para identificar padrões de formas, prever sequências numéricas, escrever poemas sobre um determinado assunto ou resolver problemas de analogia verbal - tudo o que a IA pesquisadores têm trabalhado. Essa máquina pode ser capaz de obter um respeitável QI [25], embora fora desses problemas específicos de teste seja o próximo inútil. Se tentarmos corrigir essas limitações expandindo além do padrão testes, como Bringsjord e Schimanski parecem sugerir, isso mais uma vez abre o dificuldade de exatamente o que e o que não testar. IA psicométrica, pelo menos atualmente formulado, aborda apenas parcialmente essa questão central.

8 Teste de Smith

A estrutura básica do teste de Smith é que um agente enfrenta uma série de problemas que são gerados por um algoritmo [30]. Em cada iteração, o agente deve tentar produzir a resposta correta ao problema que foi dado. O gerador de problemas então responde com uma pontuação de quão boa foi a resposta do agente. Se o agente assim deseja que ele possa enviar outra resposta para o mesmo problema. Em algum momento o agente solicitações ao gerador de problemas para passar para o próximo problema e a pontuação que o agente recebido para sua última resposta ao problema atual é adicionado ao seu pontuação cumulativa. Cada ciclo de interação conta como uma etapa única e a inteligência é então sua pontuação cumulativa total considerada em função do tempo. Dentro Para manter as coisas viáveis, todos os problemas devem estar em P, ou seja, a solução deve seja verificável em tempo polinomial.

Temos duas críticas principais à definição de Smith. Em primeiro lugar, enquanto razões pelas quais pode fazer sentido restringir os problemas a P, não vemos por que isso restrição prática deve fazer parte da própria definição de inteligência como Smith sugere. Se alguma inovação significou que os agentes poderiam resolver problemas difíceis em não apenas P, mas às vezes também em NP, então esses novos agentes certamente seriam mais inteligente?

Em segundo lugar, embora a definição seja formalmente definida, ela ainda deixa em aberto a questão importante sobre o que exatamente devem ser os testes. Smith sugere que os pesquisadores devem sonhar com os testes e depois contribuí-los para algum pool comum de testes. Como tal, este não é um teste totalmente especificado.

6

 Page 8

9 Teste C

Uma perspectiva entre os psicólogos que apóiam a visão da inteligência do fator g , é que a inteligência é "a capacidade de lidar com a complexidade" [10]. Assim, em um teste de inteligência, as perguntas mais difíceis são as mais complexas porque eles, por definição, exigem mais inteligência para serem resolvidos. Segue-se então que se pudéssemos definir e medir formalmente a complexidade dos problemas de teste, poderíamos construir um teste formal de inteligência. A possibilidade de fazer isso foi talvez a primeira sugerido pelo teórico da complexidade Chaitin [5]. Embora esse caminho exija inúmeras dificuldades a serem enfrentadas, acreditamos que é a mais natural e oferece muitas vantagens: É formalmente motivado, definido com precisão e potencialmente pode ser usado para medir o desempenho de computadores e sistemas biológicos no mesmo escala sem o problema de viés em relação a qualquer espécie ou cultura em particular.

Um teste de inteligência baseado na teoria formal da complexidade é o Teste C de Hernández [13, 14]. Esse teste consiste em várias previsões de sequência e problemas de abdução semelhantes aos que aparecem em muitos testes padrão de QI. Semelhante Para testes de QI padrão, o C-Test sempre garante que cada pergunta tenha um resposta no sentido de que sempre existe uma hipótese consistente com o padrão observado que tem uma complexidade significativamente menor do que as alternativas. A principal diferença para sequenciar os problemas que aparecem nos testes de inteligência padrão é que as perguntas se baseiam em uma medida formalmente expressa de complexidade, A complexidade computacional K_t de Levin [20] (em vez da incomputável de Kolmogorov complexidade [21]) para fazer um teste prático. Para manter a propriedade de invariância de Complexidade de Kolmogorov, a complexidade de Levin requer a suposição adicional de que as máquinas universais de Turing são capazes de simular-se em tempo linear.

O teste foi realizado com sucesso a seres humanos com recursos intuitivamente razoáveis resultados [14]. Até onde sabemos, esta é a única definição formal de inteligência que até agora produziu um teste utilizável de inteligência.

Uma crítica aos testes C-Test e Smith é que a maneira como a inteligência é medido é essencialmente estático, ou seja, os ambientes são passivos. Acreditamos que o teste dinâmico em ambientes ativos é uma medida melhor da inteligência de um sistema. Para colocar esse argumento de outra maneira: ter sucesso no mundo real exige que você seja mais do que um espectador perspicaz! É preciso escolher cuidadosamente ações, sabendo que isso pode afetar o futuro.

10 Inteligência universal

Outro teste baseado na complexidade é o teste de inteligência universal [19]. Ao contrário do Teste C e teste de Smith, a inteligência universal testa o desempenho de um agente em um ambiente totalmente interativo. Isso é feito usando o aprendizado por reforço estrutura na qual o agente envia suas ações para o ambiente e recebe observações e recompensas de volta. O agente tenta maximizar a quantidade de recompensa

7

recebe aprendendo sobre a estrutura do ambiente e os objetivos de que precisa realizar para receber recompensas.

Formalmente, o processo de interação produz uma história crescente $o_1 r_1 a_1 o_2 r_2 a_2 o_3 r_3 a_3 o_4 \dots$ de observações o , recompensas $r \geq 0$ e ações a . O agente é simplesmente uma função, denotada por π , que é uma medida de probabilidade sobre as ações con-no histórico atual, por exemplo, $\pi(a_3 | o_1 r_1 a_1 o_2 r_2)$. O ambiente, denotado μ , é definido da mesma forma: $\mu(o_k r_k | o_1 r_1 a_1 o_2 r_2 a_2 \dots o_{k-1} r_{k-1} a_{k-1})$. O desempenho do agente π no ambiente μ pode ser medido por sua recompensa total esperada $V_\pi = \sum_i \gamma^i \pi(r_i | \mu, \pi)$, chamado de valor. A maior classe interessante de ambientes é

a classe E de todas as distribuições de probabilidade computáveis μ . Por razões técnicas, o presume-se que os valores estejam limitados por alguma constante c .

Para obter uma única medida de desempenho, V_{π} é calculado sobre todos os $\mu \in E$. Como são um número infinito de ambientes, sem limites à sua complexidade, é impossível obter o valor esperado em relação a uma distribuição uniforme - alguns ambientes devem ser mais pesados que outros. Considerando o perspectiva do agente sobre o problema, é o mesmo que perguntar: Dadas várias hipóteses consistentes com as observações, qual hipótese deve ser considerado o mais provável? Este é um problema fundamental na inferência indutiva para qual a solução padrão é invocar a lâmina de Occam: dadas várias hipóteses que são consistentes com os dados, os mais simples devem ser preferidos. Como isso é genérico geralmente considerada a coisa mais inteligente a se fazer, deve-se testar os agentes são, pelo menos em média, recompensados por aplicar corretamente as taxas de Occam zor. Isso significa que a distribuição a priori dos ambientes deve ser ponderada para ambientes mais simples.

Como cada ambiente μ é descrito por uma medida computável, sua complexidade pode ser medido com a complexidade K de Kolmogorov (μ), que é simplesmente o comprimento da o programa mais curto que calcula μ [21]. O peso a priori correto para μ é $2^{-K(\mu)}$. Agora podemos definir a inteligência universal de um agente π como simplesmente a sua expectativa desempenho,

$$Y(\pi) = \sum_{\mu \in E} 2^{-K(\mu)} V_{\pi} \mu.$$

Por construção, a inteligência universal mede a capacidade geral de um agente de ter bom desempenho em uma ampla gama de ambientes, semelhante à essência de muitos definições informais de inteligência [18]. A definição não impõe restrições à o funcionamento interno do agente; requer apenas que o agente seja capaz de gerar saída e receber entrada que inclui um sinal de recompensa. Se desejarmos para influenciar o teste para refletir o conhecimento do mundo, podemos condicionar a complexidade a medida. Por exemplo, use $K(\mu | D)$ em que D é um conjunto de conhecimentos básicos como a Wikipedia.

Considerando V_{π} para vários ambientes básicos, como pequenos MDPs, e agentes com estratégias de otimização simples, mas muito gerais, é claro que Orders ordena corretamente a inteligência relativa desses agentes de maneira natural. UMA um valor muito alto de Y significaria que um agente é capaz de ter um bom desempenho em muitos

8

ambientes. O agente máximo em relação a Y é o agente AIXI teórico que demonstrou ter muitas propriedades fortes de otimalidade [16]. Esses resultados confirmamos que agentes com alta inteligência universal são realmente muito poderosos e adaptável. A inteligência universal abrange agentes adaptativos simples, até super agentes inteligentes como AIXI. O teste é completamente formalmente especificado em termos de conceitos fundamentais, como computação e complexidade universais de Turing e portanto, não é antropocêntrico.

Um teste baseado em Y avaliaria o desempenho de um agente em uma amostra grande ambientes simulados e, em seguida, combine o desempenho do agente em cada ambiente um valor global de inteligência. O principal desafio que precisa ser enfrentado com é encontrar um substituto adequado para a complexidade incomputável de Kolmogorov função, possivelmente a complexidade K_t de Levin [20], como é feito pelo C-Test.

11 Resumo

Encerramos esta pesquisa com uma comparação dos vários testes considerados. Tabela 1 taxas cada teste de acordo com as propriedades descritas abaixo. Embora tenhamos tentado

Para ser o mais justo possível, algumas das pontuações que damos nesta tabela serão naturalmente discutível. Contudo, esperamos que ele forneça uma visão geral aproximada do pontos fortes e fracos das propostas.

Válido: um teste de inteligência deve capturar a inteligência e não alguns quantidade. Informativo: O resultado deve ser um valor escalar, ou talvez um vetor.

Vasta gama: um teste deve cobrir baixos níveis de inteligência até super inteligência.

Geral: Idealmente, gostaríamos de ter um teste muito geral que pudesse ser aplicado a tudo, desde uma mosca até um algoritmo de aprendizado de máquina. Dinâmico: um teste deve levar diretamente em conta a capacidade de aprender e se adaptar ao longo do tempo. Imparcial: um teste não deve ser tendencioso para nenhuma cultura, espécie, etc. Fundamental: Nós não deseja um teste que precise ser alterado de tempos em tempos devido a mudanças tecnológicas tecnologia e conhecimento. Formal: O teste deve ser definido com precisão, idealmente usando matemática. Objetivo: O teste não deve apelar para avaliações subjetivas como como as opiniões dos juízes humanos. Totalmente definido: o teste foi totalmente definido ou peças ainda não especificadas? Universal: o teste é universal ou é antropocêntrico? Prático: Um teste deve poder ser executado de forma rápida e automática. Teste vs. Def: finalmente, observamos se a proposta é mais um teste, mais uma definição, ou algo no meio.

Reconhecimentos.

Este trabalho foi financiado pelo subsídio NSF suíço 200020-107616.

Tabela 1: Na tabela significa "sim", • significa "discutível", · significa "não" e? significa desconhecido. Quando algo é classificado como desconhecido, geralmente é porque o teste em questão não está suficientemente especificado.

| Teste de inteligência | Informação | Valida | Ativo | Range | ic | ental | Defined | vs. | Def. |
|-------------------------------|------------|--------|-------|-------|----|-------|---------|-----|-------|
| | Informação | Valida | Ativo | Range | ic | ental | Defined | vs. | Def. |
| Teste de Turing | • | • | • | • | • | • | • | • | T |
| Teste de Turing total | • | • | • | • | • | • | • | • | T |
| Teste de Turing invertido | •• | • | • | • | • | • | • | • | T |
| Teste de Turing para criança | • | • | • | • | • | • | • | • | T |
| Complexidade linguística | • | • | • | • | • | • | • | • | T |
| Teste de compactação de texto | • | • | • | • | • | • | • | • | T |
| Relação de Turing | • | • | • | • | • | • | • | • | T / D |
| AI psicométrica | • | • | • | • | • | • | • | • | T / D |
| Teste de Smith | • | • | • | • | • | • | • | • | T / D |
| Teste C | • | • | • | • | • | • | • | • | T / D |
| Inteligência Universal | • | • | • | • | • | • | • | • | D |

Referências

[1] N. Alvarado, S. Adams, S. Burbeck e C. Latta. Além do teste de Turing: Per-
métricas de desempenho para avaliar uma simulação em computador da mente humana. Em *Per-
Workshop de Métricas para Sistemas Inteligentes*, Gaithersburg, MD, EUA, 2002.
Holanda do Norte.

[2] TC Bell, JG Cleary e IH Witten. *Compactação de texto*. Prentice Hall, 1990.

[3] N. Block. Psicologismo e behaviorismo. *Philosophical Review*, 90: 5–43, 1981.

[4] S. Bringsjord e B. Schimanski. O que é inteligência artificial? AI psicométrica
como resposta. *Décima Oitava Conferência Conjunta Internacional sobre Inteligência Artificial*,
18: 887-893, 2003.