SUPERINTELIGENCIA

PUC-SP

INOVAÇÃO E INDÚSTRIA 4.0

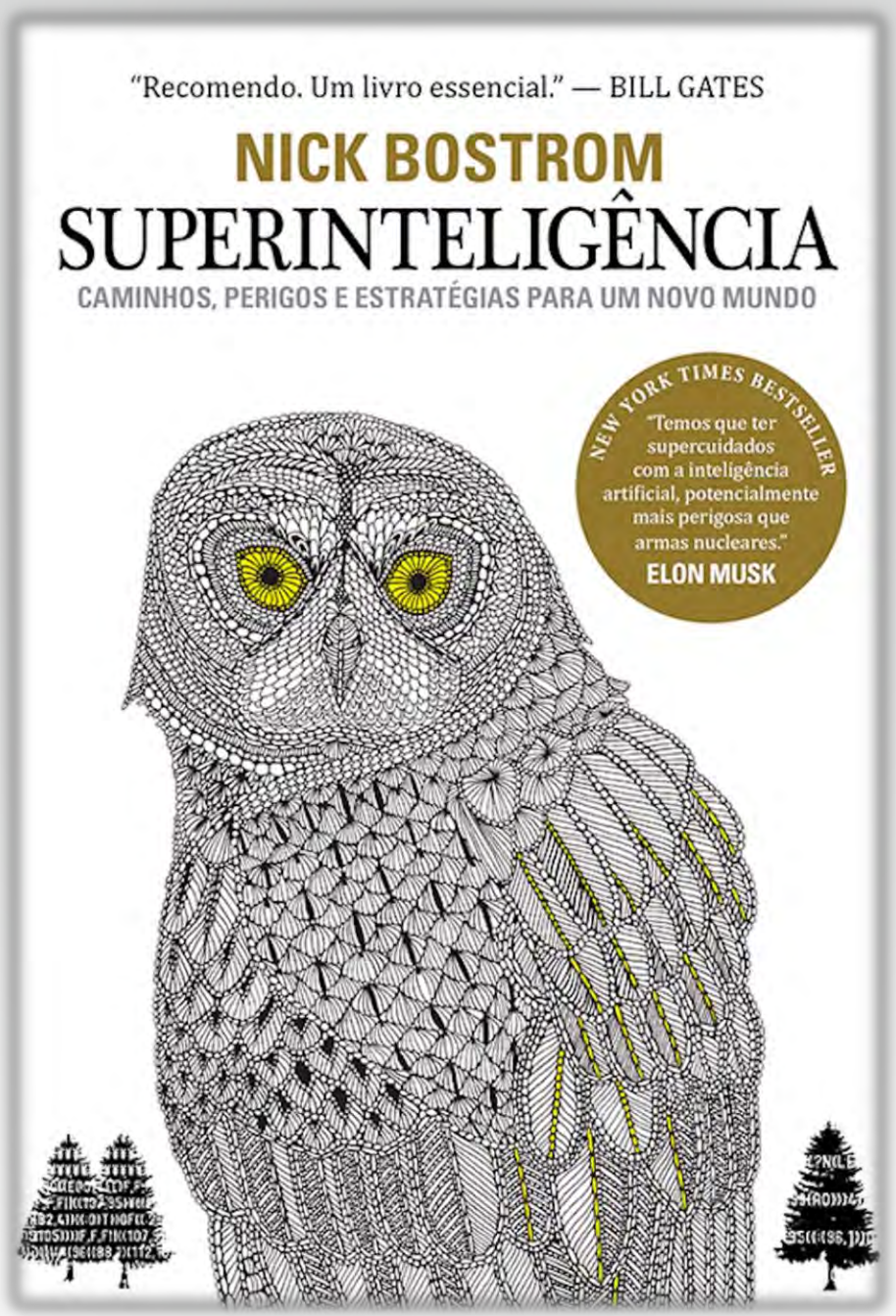
DOCENTE: ARNOLDO JOSÉ DE HOYOS GUEVARA

DISCENTE: SANDRA JOYCE SILVA DE SOUZA

Mestrado em Administração

A Fábula Inacabada dos Pardais

Pastus e Joãozinho


"Recomendo. Um livro essencial." — BILL GATES
NICK BOSTROM
SUPERINTELIGÊNCIA
CAMINHOS, PERIGOS E ESTRATÉGIAS PARA UM NOVO MUNDO
NEW YORK TIMES BESTSELLER
"Temos que ter supercuidados com a inteligência artificial, potencialmente mais perigosa que armas nucleares."
ELON MUSK

Niklas "Nick" Bostrom

Filósofo sueco da Universidade de Oxford,
conhecido por seu trabalho sobre risco existencial, o princípio antrópico, a ética do aprimoramento humano, os riscos de superinteligência e o teste de reversão.

10 de março de 1973 (46 anos)

- *Diretor do Instituto do Futuro da Humanidade, universidade de Oxford*
- *Prêmio de Distinção de Professores da Universidade de Oxford*
- *Os 100 melhores pensadores globais da FP*
- *Lista dos Melhores Pensadores do Mundo da Prospect*

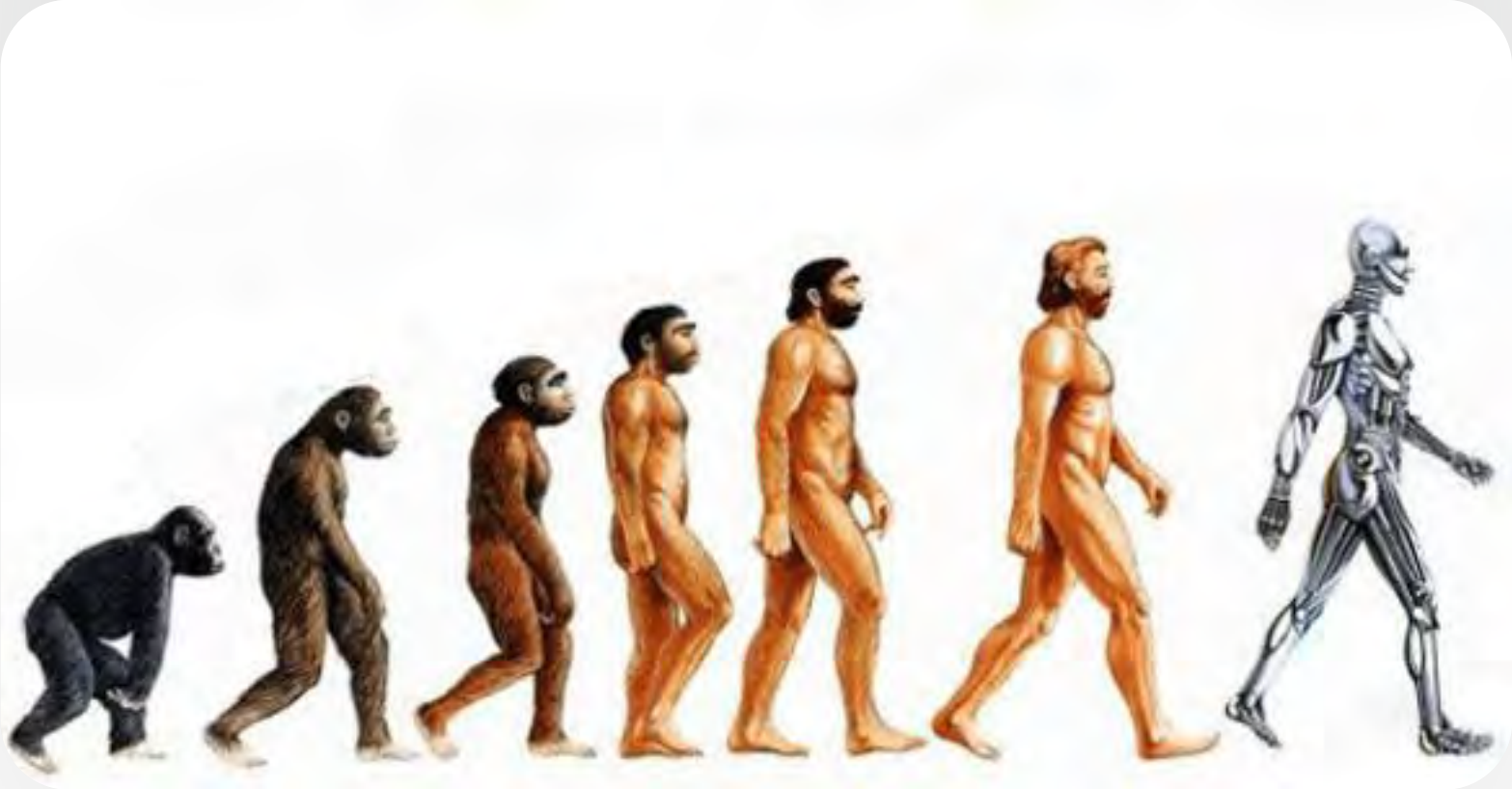
Principais interesses:
Filosofia da inteligência artificial e Bioética

Tecnológico
Solução
Receptores
Informação
Ultrainteligene
Tecnológico
Singleton
Tendência
Aflição
Programa
Dados
Antropomorfização
Especialistas
Artificiais
Arte
Singularidade
Lógicas
Retórico
Interfaces
Processamento
Guerra
Virtude
Corolário
Superinteligência
Máquinas
Esperança
Expectativas
Econômico
Proficiência
Projeto
Ciborguização
Humano

SINGULARIDADE

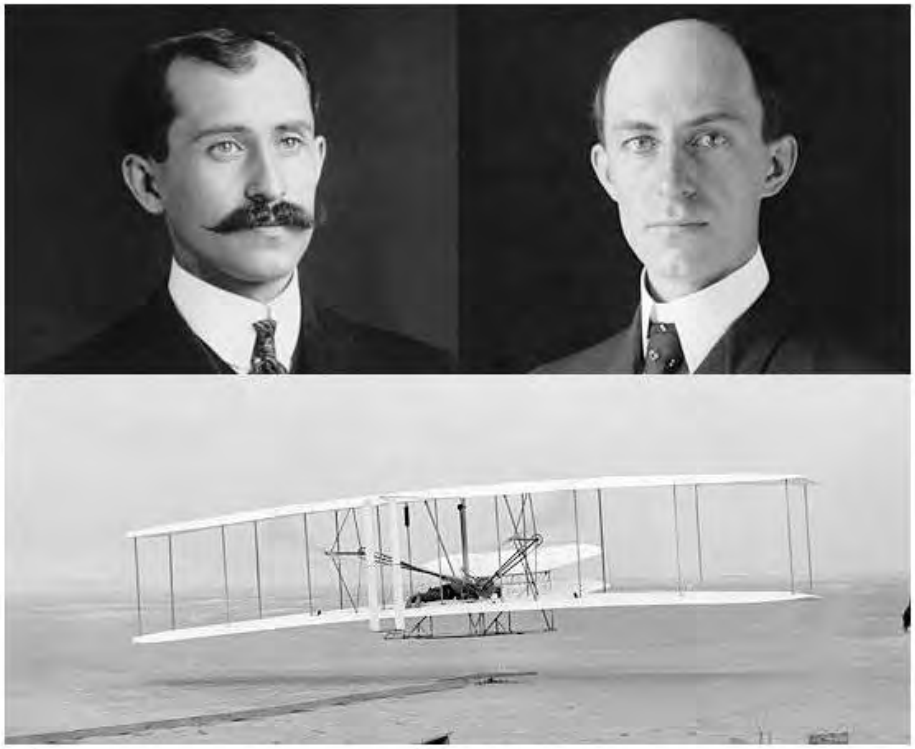
Termo emprestado da física.
Designa fenômenos tão extremos que as equações não são mais capazes de descrevê-los.
Exprime tudo o que está além da nossa capacidade de cognição e previsibilidade.

È a possibilidade de uma explosão de inteligência, particularmente a perspectiva de uma superinteligência de máquina.



Acontecimentos marcantes !

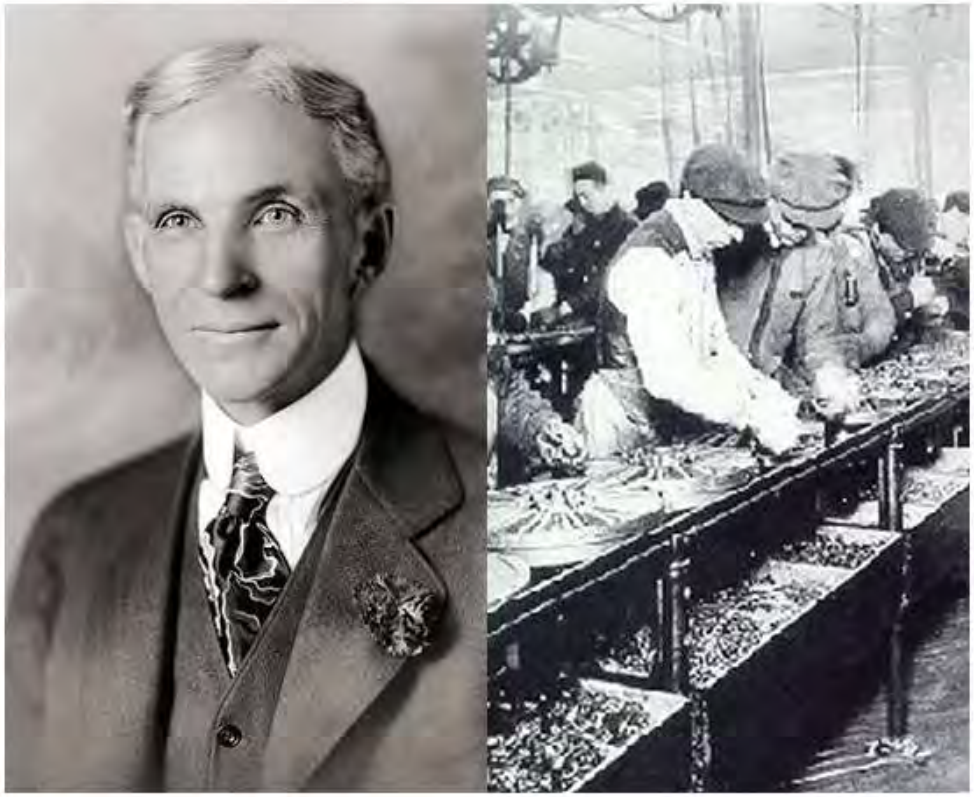
AVIÃO (1906)



Há 113 anos

Um brasileiro entraria para a história como o inventor do primeiro avião. Alberto Santos Dumont foi o responsável pela criação do 14 Bis conhecido também como ‘Oiseau de Proie’, que em francês significa ‘Ave de Rapina’. A primeira demonstração de voo foi realizada em 12 de novembro de 1906 em Paris, na França e acompanhada por uma multidão de pessoas espantadas e documentada pela presença da imprensa europeia.

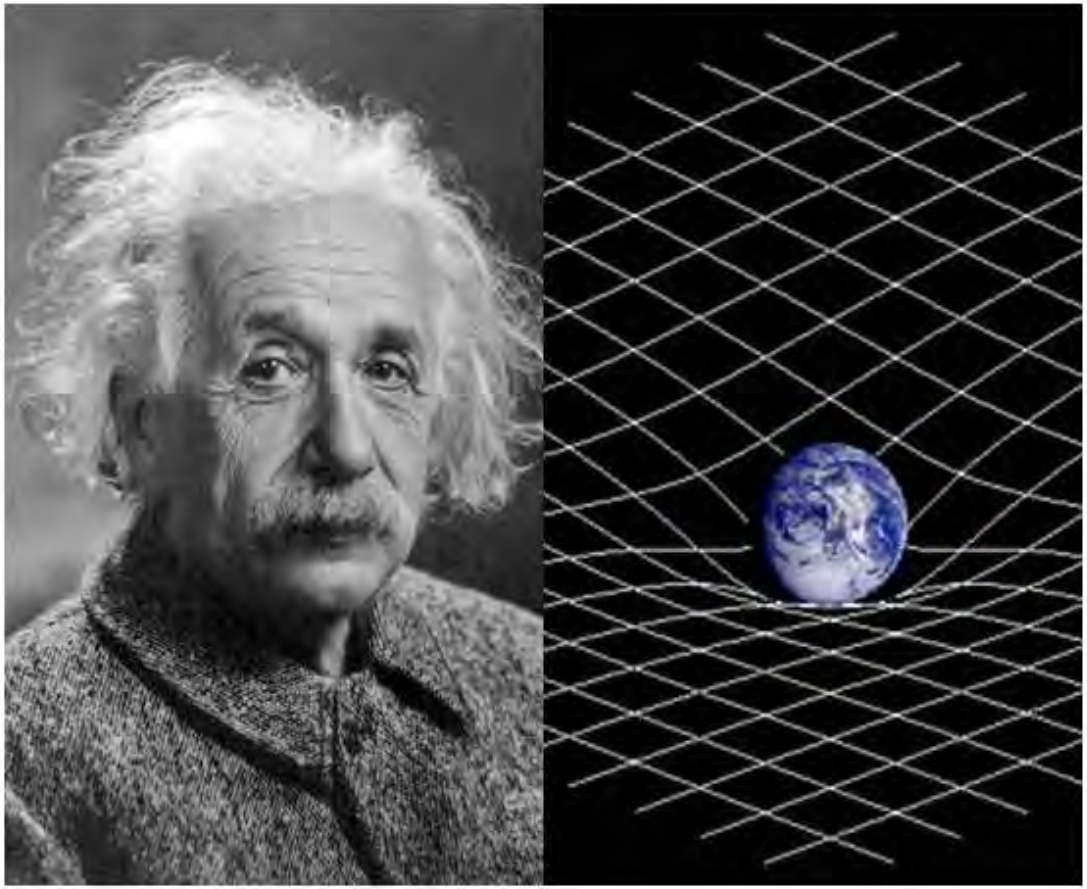
LINHA DE PRODUÇÃO (1913)



Há 106 anos

A forma mais característica, a da montagem em série, foi inventada por Henry Ford, empresário estadunidense do setor automobilístico. Graças a ela, Ford conseguiu produzir em massa seu famoso carro Ford T.

TEORIA DA RELATIVIDADE GERAL (1915)

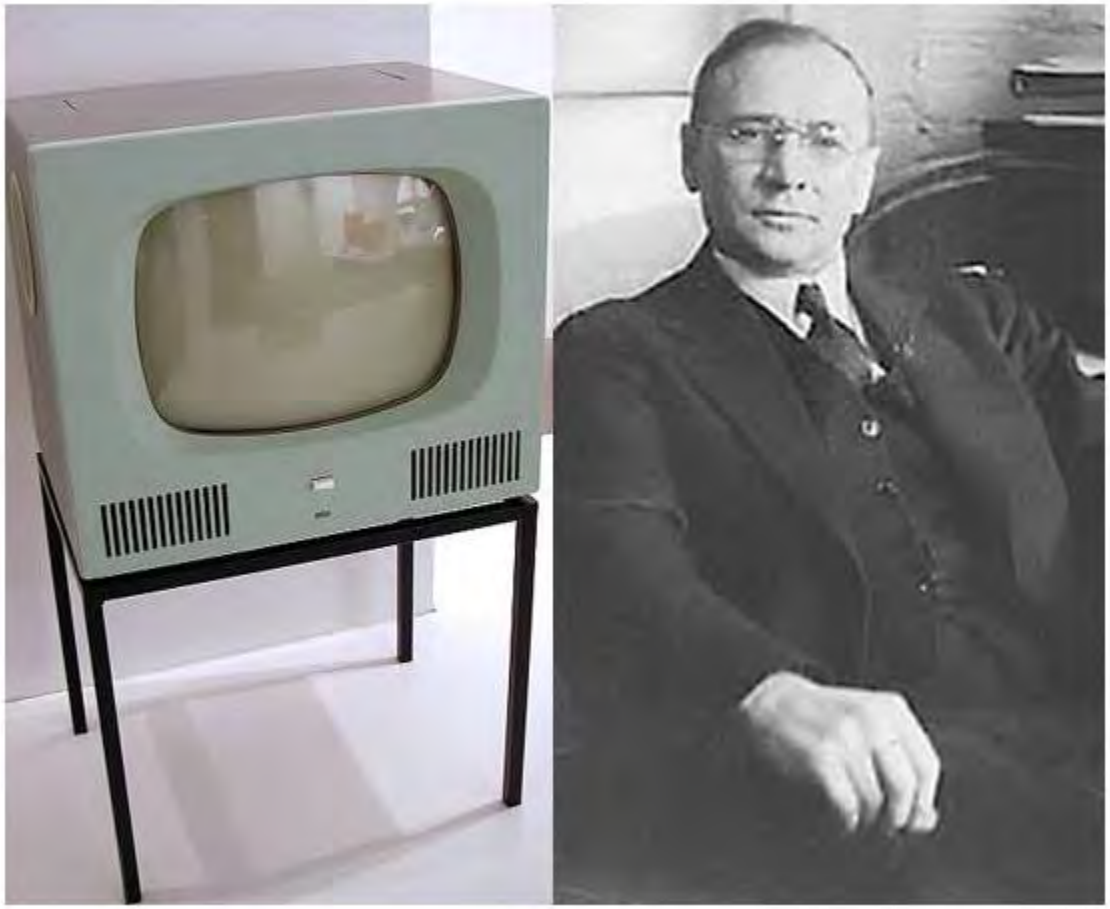


Há 104 anos

Em sua teoria da Relatividade Geral, Einstein procura avaliar o que acontece em referenciais não inerciais (que possuem aceleração). Ele chega a algumas importantes conclusões:

- Um referencial que sofre aceleração é equivalente a um referencial submetido a uma força atuando à distância.

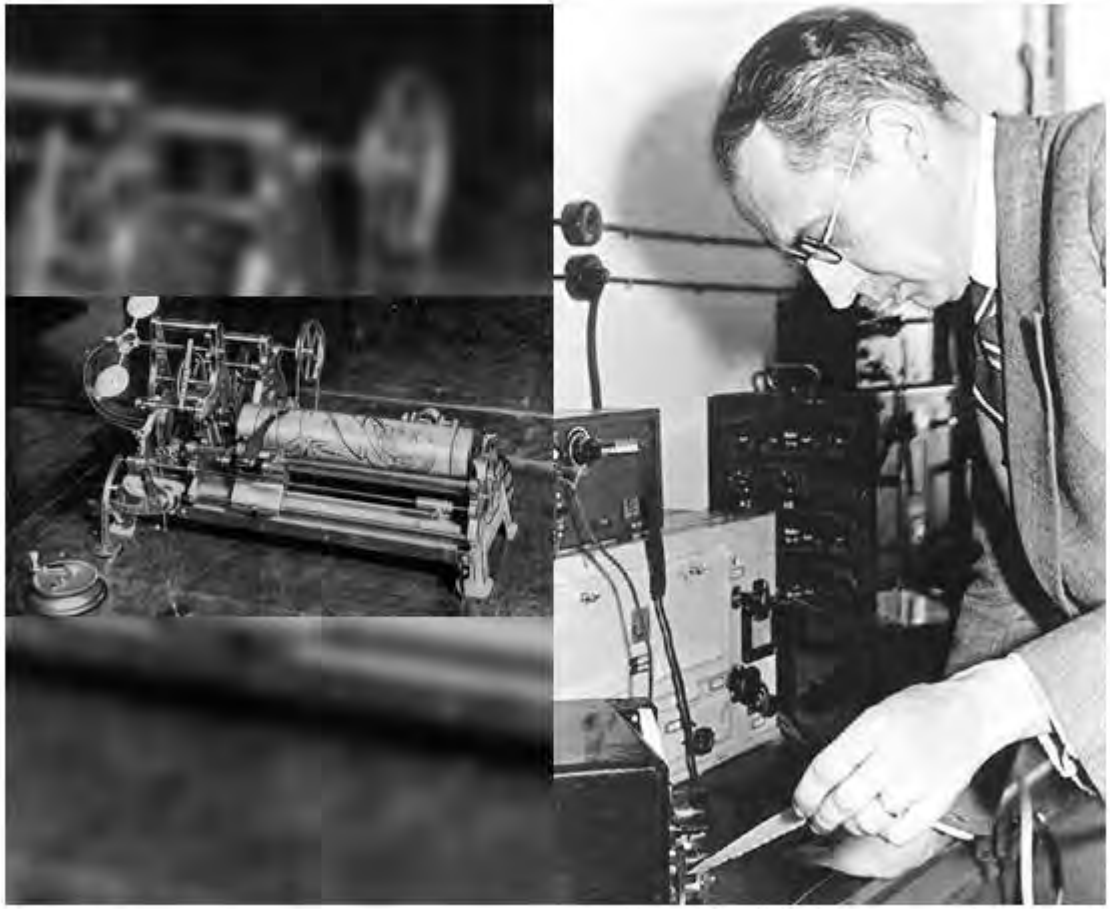
TELEVISÃO (1923)



Há 96 anos

A invenção da televisão não foi um processo muito rápido, o tubo televisivo foi criado em 1923 por Vladimir Zworykin um inventor russo que morava nos Estados Unidos. Mas a primeira transmissão ocorreria apenas em 1928, a má qualidade das imagens assemelhava o televisor aos rádios. Em 1935 o aparelho fez as primeiras transmissões pra valer, no entanto só poderia ser assistida em salas públicas.

FAX (1929)



Há 90 anos

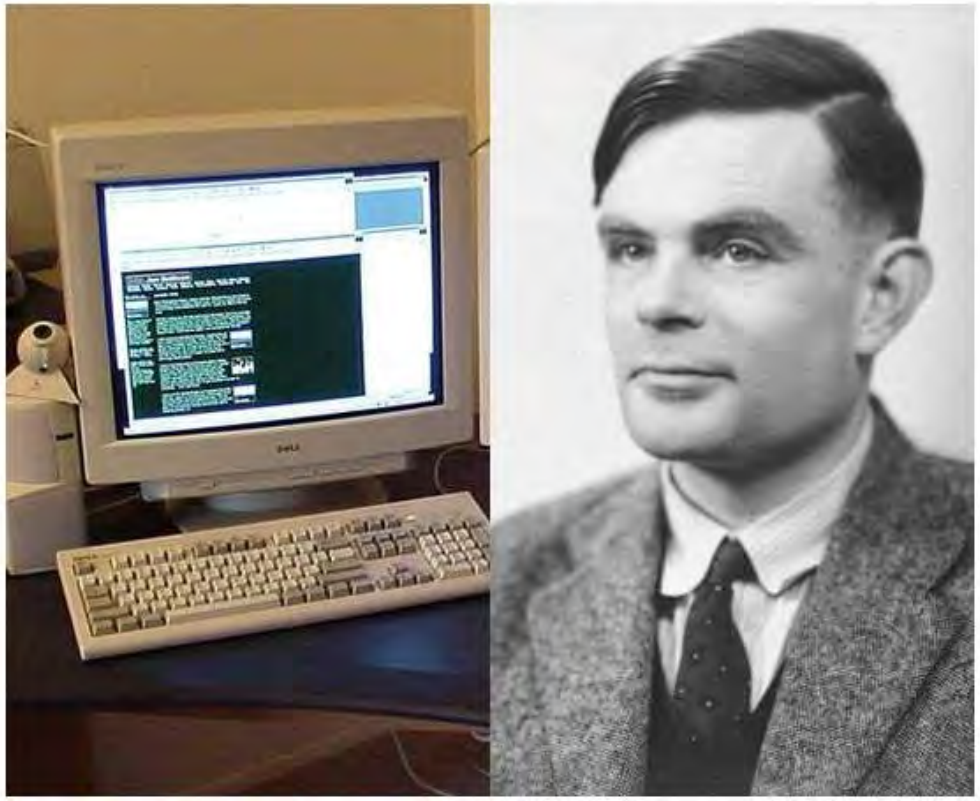
O inventor alemão Rudolf Hell criou em 1929 uma máquina capaz de transmitir textos eletronicamente. O dispositivo é formado por um telefone e uma copiadora, consiste na cópia do documento e envio pela linha telefônica. A invenção foi um grande sucesso entre agências de notícia e investigação. Com o tempo foi sendo substituído por novas tecnologias, o uso da internet se mostraria muito mais eficaz para o compartilhamento rápido de documentos.

CELULAR (1956)

Há 63 anos
Brasil 23 anos

Atualmente o uso do celular se tornou uma febre mundial, o seu rápido aperfeiçoamento no decorrer dos anos fez com que esse aparelho exercesse diversas funções e se tornasse indispensável para a maioria das pessoas. O primeiro aparelho foi desenvolvido pela Ericsson em 1956, mas transportá-lo não era assim tão fácil, os primeiros celulares pesavam em média quarenta quilos e o alto custo em sua produção dificultou a sua popularização. Já em 1973 a Motorola lança um aparelho mais acessível pesando cerca de um quilo e medindo vinte e cinco centímetros de comprimento e sete de largura, a primeira chamada móvel foi realizada na cidade de Nova York. **O primeiro celular chegou ao Brasil em 1990**.

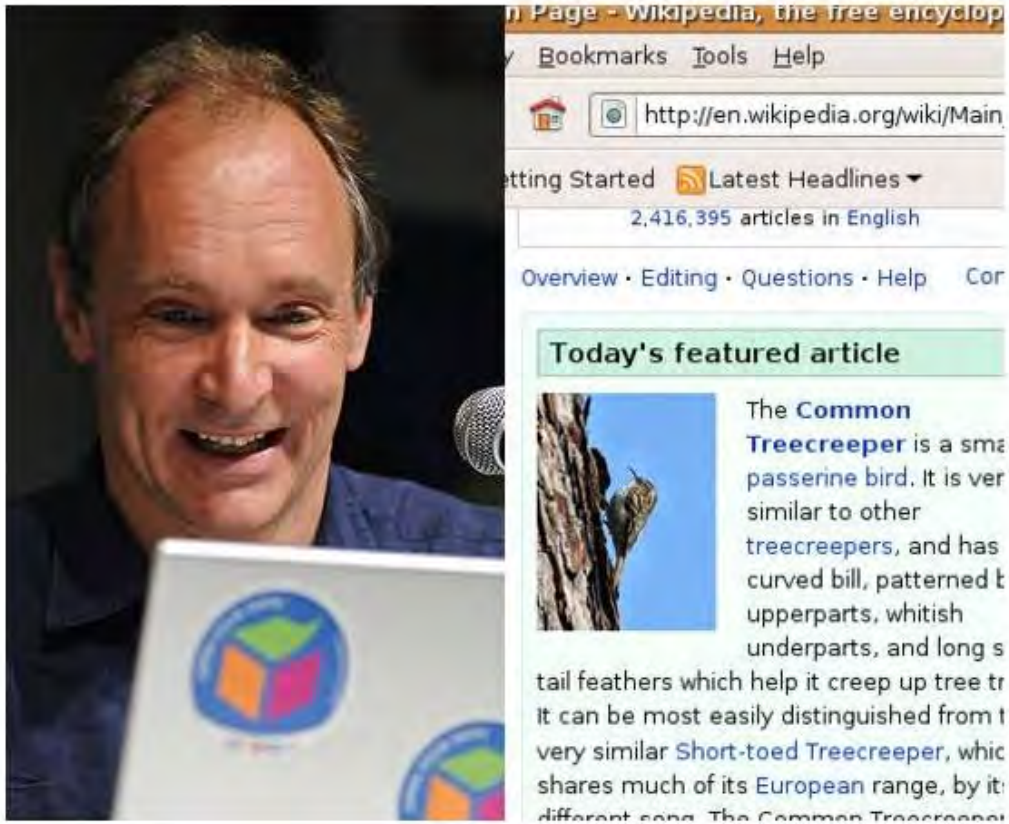
COMPUTADORES PESSOAIS (1950/80)



Há 69 anos

Em 1946, era anunciado o nome do primeiro computador digital eletrônico de grande escala: o ENIAC (Electrical Numerical Integrator and Calculator). Criado pelos cientistas norte-americanos John Presper Eckert e John W. Mauchly, da Electronic Control Company. O ENIAC pesava 30 toneladas e ocupava uma área de 270 m² de área construída.

INTERNET (1992)



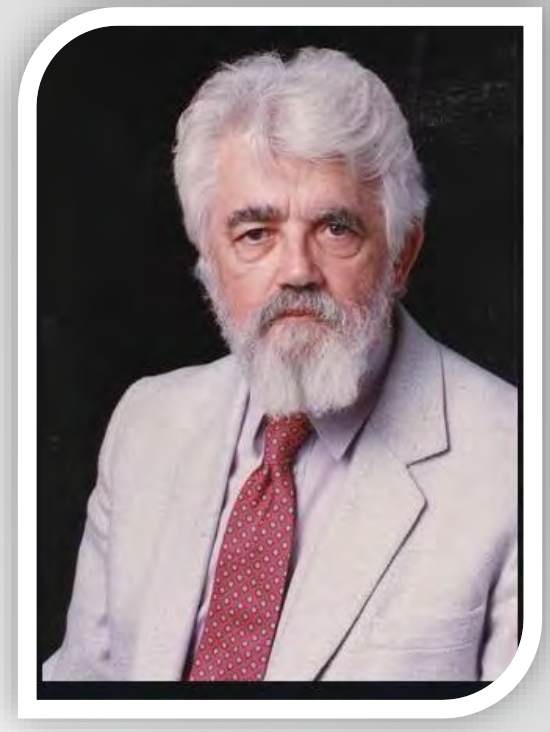
Há 27 anos

A Internet se origina de um projeto chamado Arpanet desenvolvido nos Estados Unidos durante a Guerra Fria pelo seu Departamento de Defesa norte-americano. A disputa existente entre União Soviética e Estados Unidos estimulou o desenvolvimento tecnológico, qualquer nova invenção poderia ser uma vantagem em relação ao inimigo. Nesse período havia uma intensa rede de espiões dos dois lados, assim os militares americanos viram a necessidade de criar um sistema de proteção aos dados secretos do governo. Eles pensavam que se houvesse um ataque ao Pentágono todas as informações secretas mantidas ali poderiam desaparecer.

Especialistas previam no fim dos anos 1950:
“Se alguém pudesse elaborar uma máquina capaz de jogar xadrez, essa pessoa teria, supostamente, penetrado no âmago da empreitada intelectual humana”.

Kasparow e Deep Blue (1997): *Demütigende Niederlage*

COMO ERA
AGORA É
TAXI



“Quando algo funciona, ninguém mais chama de Inteligência Artificial”.

John McCarthy

10% de chance: 2030 **50%** de chance: 2050 **90%** de chance: 2100

TABELA 2

Quando a inteligência de máquina de nível humano será alcançada?*

	10%	50%	90%
PT-AI	2023	2048	2080
AGI	2022	2040	2065
EETN	2020	2050	2093
TOP100	2024	2050	2070
Combinados	2022	2040	2075

* A tabela mostra os resultados de quatro pesquisas de opinião diferentes, bem como os resultados combinados. As duas primeiras foram pesquisas realizadas em conferências acadêmicas: PT-AI, participantes da conferência *Philosophy and Theory of AI*, em Salônica, no ano de 2011 (os participantes foram entrevistados em novembro de 2012), com uma taxa de resposta de 43 num conjunto de 88; e AGI, participantes das conferências *Artificial General Intelligence e Impacts and Risks of Artificial General Intelligence*, ambas realizadas em Oxford em dezembro de 2012 (taxa de resposta: 72/111). A pesquisa EETN é uma amostragem dos membros da Greek Association for Artificial Inteligence, uma organização profissional de pesquisadores da área, em abril de 2013 (taxa de resposta: 26/250). A pesquisa TOP100 coletou as opiniões dos 100 principais autores em inteligência artificial de acordo com um índice de citação, em maio de 2013 (grau de resposta: 29/100).

SUPERINTELIGÊNCIA

Qualquer intelecto que exceda em muito o desempenho cognitivo dos seres humanos em, virtualmente, todos os domínios de interesse.

Caminhos possíveis para a superinteligência

- Inteligência Artificial
- Emulação completa do cérebro
- Cognição biológica
- Interfaces cérebro-computador
- Redes e organizações

Inteligência Artificial

Alan Turing

Alan Mathison Turing foi um matemático, lógico, criptoanalista e cientista da computação britânico. Foi influente no desenvolvimento da ciência da computação e na formalização do conceito de algoritmo e computação com a máquina de Turing, ficando conhecido como pai da computação e da inteligência artificial.

Nascimento: 23 de junho de 1912, Maida Vale
Falecimento: 7 de junho de 1954, Wilmslow - Reino Unido
Nacionalidade: Britânico
Formação: Universidade de Princeton (1936–1938)

Nome da máquina: **Christopher**

Fatos e curiosidades sobre Alan Turing

Há 65 anos, uma dose de cianeto pôs fim à vida do inglês que é considerado o pai da computação e da inteligência artificial

O jogo de Turing

Criança prodígio

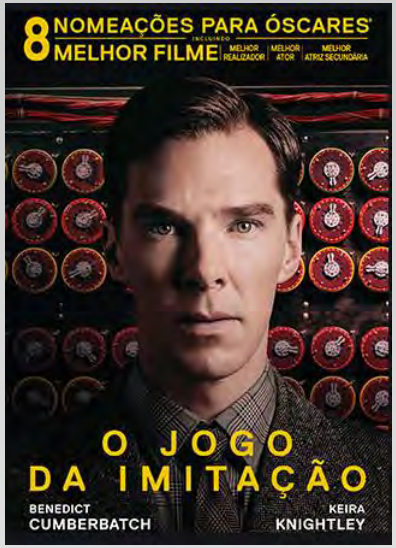
Maçã envenenada

Biólogo Turing

O visionário

Homofobia mata
(preso e quimicamente castrado)

Mas há controvérsias



A leitura do jovem cientista
(sempre gostou de matemática e física)

Turing atleta

Ódio aos EUA

Agente Secreto

Lei de Turing

Computador teórico

Xadrez para PC

Reconhecimento póstumo

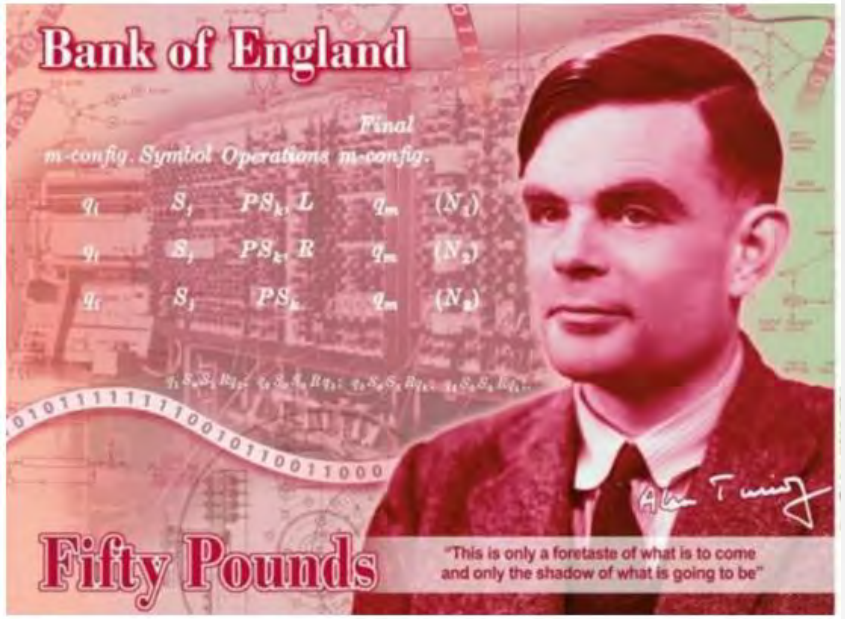
Gênio imcompreendido

Perdão Real
Somente em 2009, o primeiro ministro do Reino Unido, Gordon Brown, pediu desculpas publicamente em nome do governo britânico.

Alan Turing será homenageado ao estampar nota de 50 libras no Reino Unido

A escolha de Alan Turing como a nova cara da nota de 50 libras saiu de uma lista de 227.299 indicados por sua contribuição à ciência. Nessa lista estavam pessoas como Stephen Hawking e Margaret Thatcher (que, antes de entrar na política, trabalhou como química).

A nova versão do dinheiro britânico apresenta uma série de referências ao matemático e entrará em circulação no final de 2021



O matemático inglês Alan Turing estampará as notas de 50 libras (Foto: Banco da Inglaterra)

"Como pai da ciência da computação e da inteligência artificial, além de herói de guerra, as contribuições de Alan Turing foram abrangentes e desbravadoras", disse o presidente do Banco da Inglaterra, Mark Carney, que tomou a decisão final sobre a escolha.

CONCEITO DE MÁQUINA CRIANÇA

“Em vez de tentar criar um programa capaz de simular a mente de um adulto, por que não tentar produzir um que simule a mente de uma criança? Se ele fosse, então, submetido a uma trajetória apropriada de aprendizado , seríamos capazes de obter cérebro de um adulto”.

Turing (1950, p. 456)

EMULAÇÃO COMPLETA DO CÉREBRO

Softwares inteligentes seriam produzidos através do mapeamento e modelagem precisa da estrutura computacional de um cérebro biológico (plágio descarado)

Para se chegar a uma emulação completa do cérebro, é preciso seguir os seguintes passos:

Primeiro, um mapeamento suficientemente detalhado de um cérebro humano específico é criado. Pode envolver a estabilização post mortem do cérebro por vitrificação. À partir de máquinas o cérebro seria dissecado, fatiado e escaneado, enquanto substâncias de contrastes seriam aplicadas para ressaltar propriedades estruturais e químicas.

Segundo, os dados brutos obtidos são inseridos em um computador para processamento automático das imagens, com o objetivo de reconstruir o modelo tridimensional da rede neural que implementaria a cognição no cérebro original.

Terceiro estágio, a estrutura neurocomputacional resultante dos dois estágios anteriores seria implementada em um computador suficientemente potente. Se bem sucedida, o resultado seria uma reprodução digital do intelecto original com memória e personalidade intactas.

A emulação de uma mente humana passaria então a existir como um software em um computador. Essa mente poderia tanto habitar uma realidade virtual como interagir com o mundo exterior através de apêndices robóticos.

A emulação completa do cérebro necessita, entretanto, de algumas tecnologias avançadas para se tornar realidade. Existem 3 pré-requisitos principais:

1. **Escaneamento**: microscopia de alto rendimento
2. **Tradução**: análise automática de imagens que transforme dados brutos escaneados em um modelo interpretado tridimensional de elementos neurocomputacionais relevantes
3. **Simulação**: hardware suficientemente potente para implementar a estrutura computacional resultante.

Ainda não há essa tecnologia de ponta, mas dados simples já podem ser produzidos

COGNIÇÃO BIOLÓGICA

Aprimoramento da performance dos cérebros biológicos, a princípio poderia ser obtido através da reprodução seletiva, sem a ajuda de tecnologia. Qualquer tentativa de iniciar um **programa eugenista**(bem nascido) clássico em larga escala enfrentaria, entretanto, uma série de restrições políticas e morais.

Francis Galton foi um antropólogo, meteorologista, matemático e estatístico inglês. Galton era o mais novo de nove filhos de um próspero banqueiro, nasceu em uma família socialmente abastada. Aos 16 anos, começou a aprender medicina, mas interessou-se pela matemática.

Nascimento: 16 de fevereiro de 1822
Falecimento: 17 de janeiro de 1911
Nacionalidade: Inglês

Cientista chinês Eis Jiankui que criou bebês geneticamente modificados atuou ilegalmente

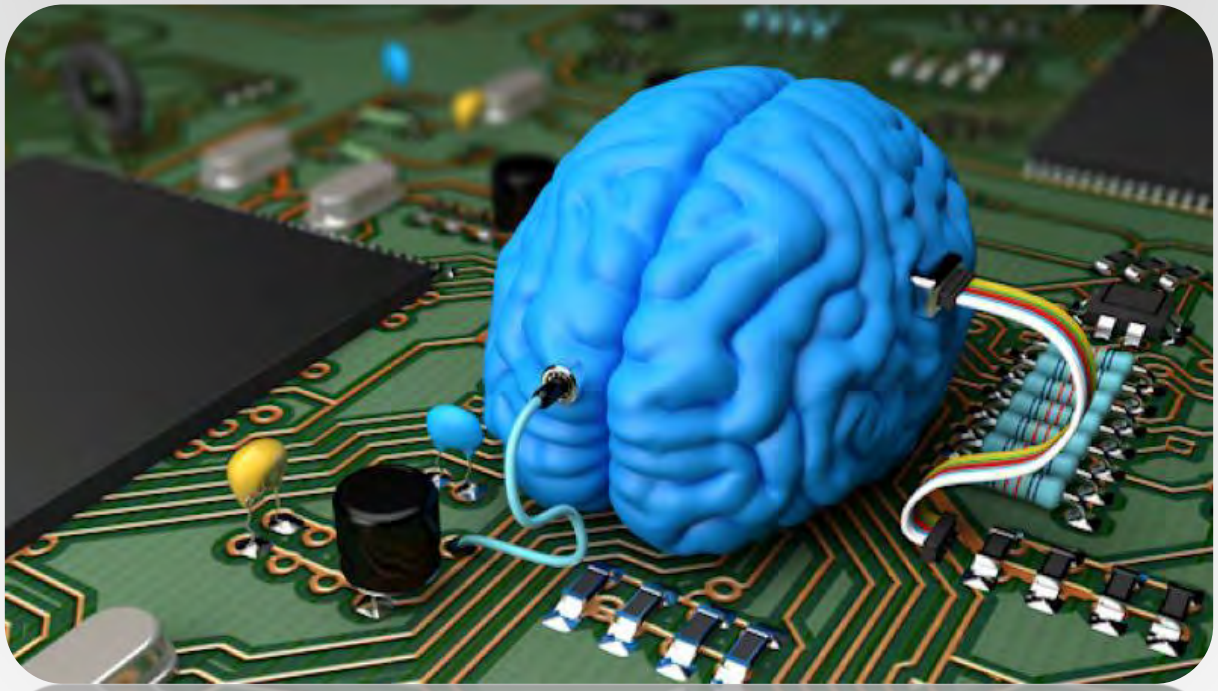
Conhecido em nível mundial após ter afirmado em novembro de 2018 ter conseguido criar os primeiros bebês manipulados geneticamente para resistir ao HIV, atuou de forma ilegal, segundo os resultados de uma investigação preliminar

Outras formas de cognição individual

- Educação e treinamento
- Melhorias na nutrição materna e infantil
- Remoção de chumbo e outros poluentes neurotóxicos do meio ambiente
- Erradicação de parasitas
- A melhoria da biomedicina poderia permitir impulsos maiores
- Incentivo a níveis adequados de sono e exercício
- Prevenção de doenças que afetam o cérebro

OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL
1 ERRADICAÇÃO DA POBREZA
2 FOME ZERO E AGRICULTURA SUSTENTÁVEL
3 SAÚDE E BEM-ESTAR
4 EDUCAÇÃO DE QUALIDADE
5 IGUALDADE DE GÊNERO
6 ÁGUA POTÁVEL E SANEAMENTO
7 ENERGIA LIMPA E ACESSÍVEL
8 TRABALHO DECENTE E CRESCIMENTO ECONÔMICO
9 INDÚSTRIA, INOVAÇÃO E INFRAESTRUTURA
10 REDUÇÃO DAS DESIGUALDADES
11 CIDADES E COMUNIDADES SUSTENTÁVEIS
12 CONSUMO E PRODUÇÃO RESPONSÁVEIS
13 AÇÃO CONTRA A MUDANÇA GLOBAL DO CLIMA
14 VIDA NA ÁGUA
15 VIDA TERRESTRE
16 PAZ, JUSTIÇA E INSTITUIÇÕES EFICAZES
17 PARCERIAS E MEIOS DE IMPLEMENTAÇÃO
OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL

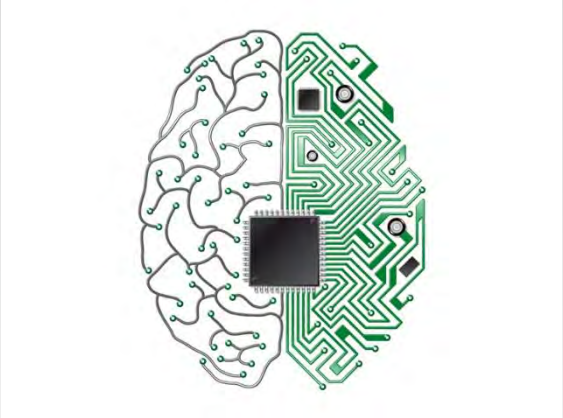
Interfaces cérebro-computador



Embora a conexão direta com o cérebro humano e um computador já tenha sido demonstrada, é pouco provável que tais interfaces venham a ser largamente utilizadas como forma de aperfeiçoamento de seres humanos em um futuro próximo.

Infecções
Deslocamento de eletrodos
Hemorragia
Declínio cognitivo
Pesquisas com implantados e não implantados de mal de Parkinson:
Redução nas fluência verbal, atenção seletiva, capacidade de nomear cores e memória verbal.

É provável que obter melhorias cognitivas seja algo muito mais complicado do que seu uso terapêutico.
Pacientes com mal de Parkinson ou dores crônicas , pois não existe de fato uma comunicação com o cérebro, o aparelho apenas fornece uma corrente elétrica estimulante ao núcleo subtalâmico.



Miguel Angelo Laporta Nicolelis é um médico e cientista brasileiro, considerado um dos vinte maiores cientistas em sua área no começo da década passada pela revista de divulgação para leigos Scientific American. Foi considerado pela Revista Época um dos 100 brasileiros mais influentes do ano de 2009.

Nascimento: 7 de março de 1961
Nacionalidade: Brasileiro (São Paulo)
Pais: Giselda Laporta Nicolelis
Campo: Neurociência
Formação: Universidade de São Paulo

Neurocientista, Nicolelis estuda o cérebro, órgão responsável pelos nossos movimentos, pensamentos e até pela escrita e leitura deste texto. Exoesqueleto

Pacientes com síndrome do encarceramento

Propostas em que extrai informações do cérebro para permitir a comunicação com outros cérebros ou com máquinas. Tais ligações têm ajudado pacientes com síndrome do encarceramento a se comunicar com o mundo exterior ao permitir que eles movam um cursor por uma tela através do pensamento. A largura de banda obtida em tais experiências é baixa: o paciente consegue, a duras penas, digitar uma letra após a outra a uma velocidade de poucas palavras por minuto.

A supercadeira
Como Hawking escrevia e pronunciava seus discursos

1. Um tablet era instalado em um suporte de metal acoplado a um dos braços da cadeira.

2. No menu havia termos prontos, como “sim”, e uma lista de palavras em ordem alfabética, além da função “soletrar”.

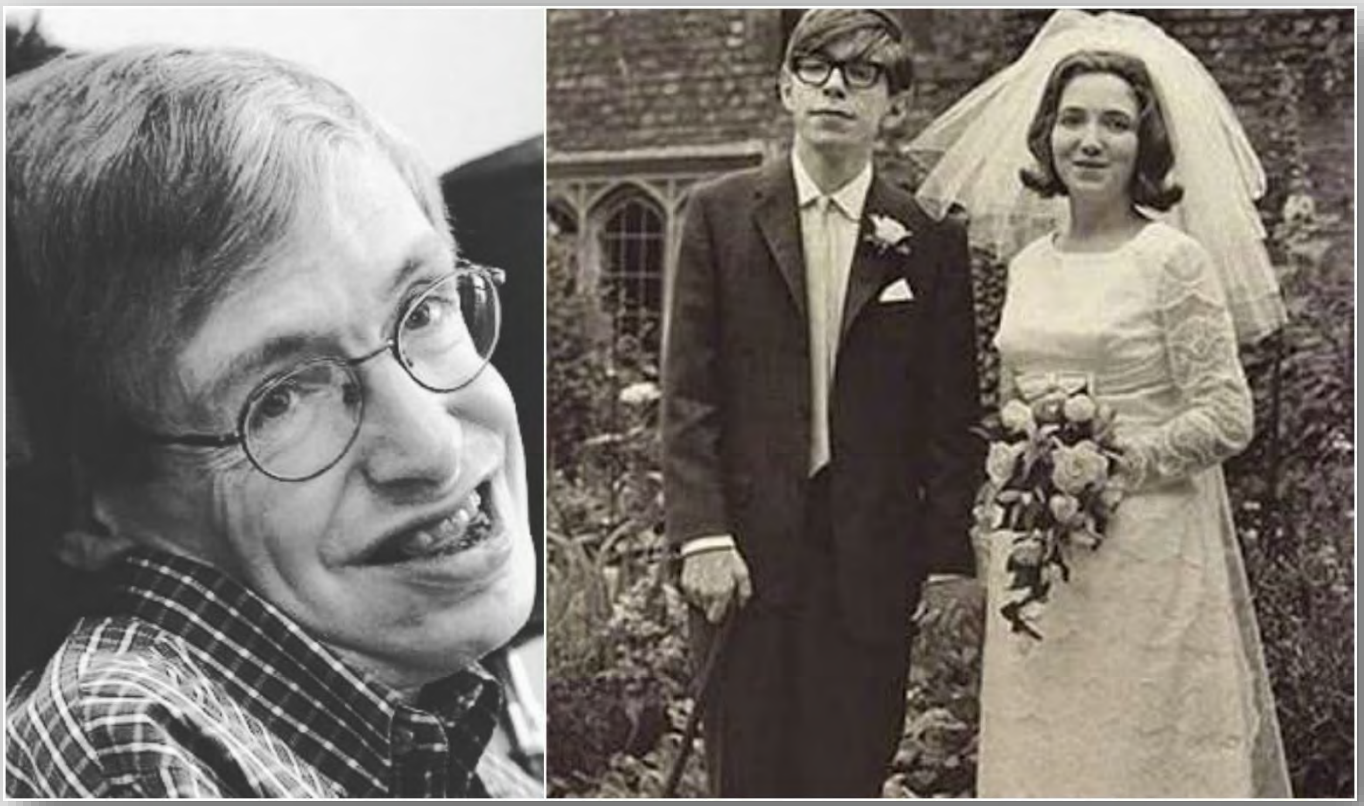
3. Um sensor infravermelho nos óculos captava movimentos da bochecha direita, usados para escolher as frases.

4. O texto completo era enviado a um sintetizador, que criava a voz simulando entonação, segundo Sam Blackburn, assistente de Hawking. O som saia atrás do suporte do computador.

5. Para palestrar, ele escrevia o discurso antes. Na hora da participação, enviava ao sintetizador uma frase por vez, o que deixava a fala mais natural.

Hawking no dia a dia

Stephen Hawking tinha na cadeira de rodas um controle remoto universal que usava para acender luzes, abrir portas e usar TV, DVD e aparelho de som.



Stephen Hawking (1942-2018) chegou a sugerir que esse seria um passo necessário para que nos fosse possível controlar os avanços em inteligência de máquina: *"Nós devemos, o mais rápido possível, desenvolver tecnologias que tornem possível uma conexão direta entre o cérebro e o computador, de forma que os cérebros artificiais contribuam com a inteligência humana em vez de se opor a ela".*

(citado em Walsh [2001])

Redes e Organizações

“Um outro caminho concebível para a superinteligência é a melhoria gradual de redes e organizações que ligam as mentes humanas individuais umas às outras e a diversos artefatos e software”.

Nick Bostrom

Possibilidades

- Um sistema com diversos indivíduos conectados e organizados em rede poderia alcançar alguma forma de superinteligência – uma “superinteligência coletiva”.
- Detectores de mentiras (caso fosse possível produzir detectores confiáveis e fáceis de usar) poderiam reduzir o alcance das mentiras nas relações humanas.
- Detectores capazes de identificar autoenganos poderiam ser ainda mais poderosos
- Empresas questionariam funcionários (roubo e produtividade) !
- Líderes políticos e empresariais seriam questionados sobre a efetiva representatividade a que se propôs !
- Sites de redes sociais já são usados por mais de 1 bilhão de pessoas para o compartilhamento de informações pessoais: muito em breve, esses indivíduos poderão fazer uploads contínuos de gravação do seu dia a dia, realizadas através das câmeras e microfones de seus smartphones ou na armação de seus óculos.
- A vigilância voluntária e involuntária permitirá o acúmulo de uma enorme gama de informações a respeito do comportamento humano.

Centro da revolução tecnológica, São Francisco bane o uso de reconhecimento facial pelo governo

No coração da revolução tecnológica mundial, São Francisco tomou uma posição contra possíveis abusos ao proibir o uso de softwares de reconhecimento facial pela polícia e outras agências de governo. Aprovado por 8 a 1 no Conselho de Supervisores municipal, o veto fez da cidade o primeiro grande centro americano a bloquear a ferramenta que muitas forças de segurança têm adotado em busca de suspeitos criminosos e possíveis autores de ataques em massa.

Revista Época Negócios em 15/05/2019

Formas de superinteligência

Várias máquinas e animais não humanos já atuam em níveis super-humanos em determinadas áreas.

- Superinteligência Rápida
- Superinteligência coletiva
- Superinteligência de qualidade

Superinteligência rápida

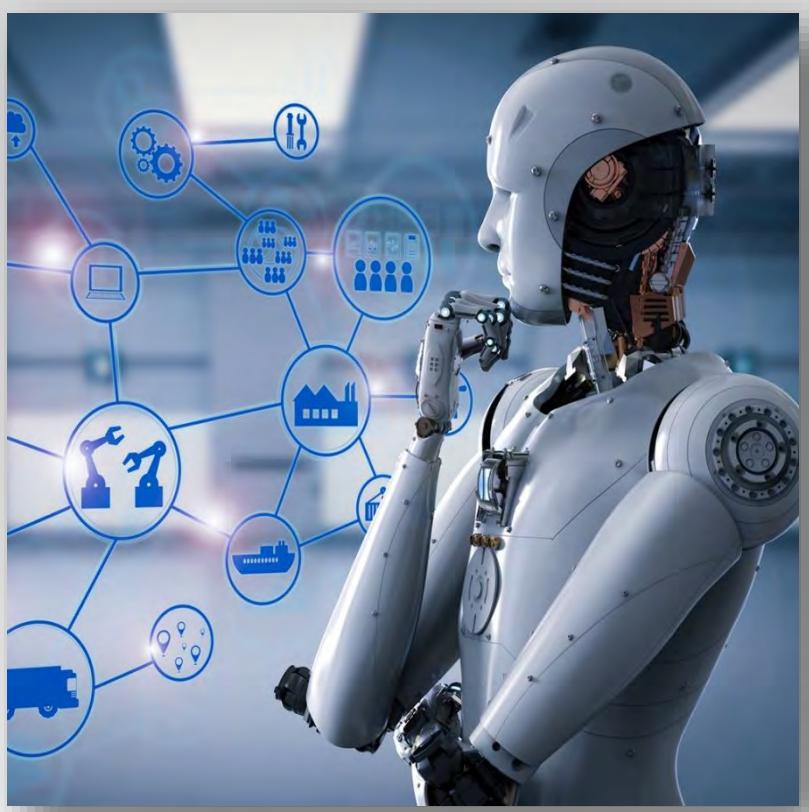
Representa um sistema que pode fazer tudo o que um intelecto humano é capaz de fazer, porém muito mais rapidamente.

Superinteligência coletiva

Representa um sistema composto de um grande número de intelectos menores, de forma que o seu desempenho total supere significativamente, em diversas áreas gerais do conhecimento, qualquer sistema cognitivo atual.

Superinteligência de qualidade

Significa um sistema que é no mínimo tão rápido quanto uma mente humana e qualitativamente muito mais inteligente.



As superinteligências rápidas, coletivas e de qualidade possuem, todas elas, o mesmo alcance indireto

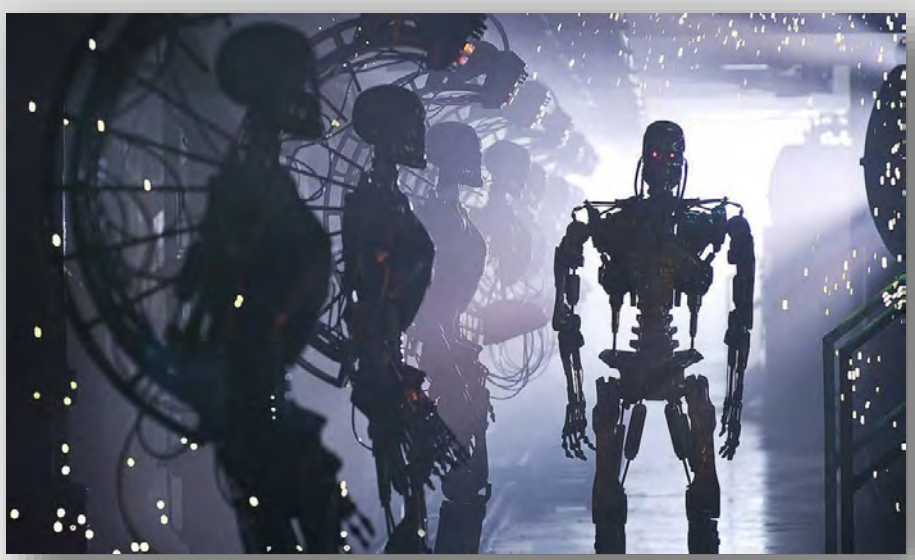
Superpoderes cognitivos

TABELA 8

Superpoderes: algumas tarefas estrategicamente relevantes e seus conjuntos de habilidades correspondentes

Tarefa	Conjunto de habilidades	Relevância estratégica
Amplificação da inteligência	Programação de IA, pesquisa em melhorias cognitivas, desenvolvimento de epistemologia social etc.	• O sistema pode desenvolver sua própria inteligência com o mínimo de auxílio externo
Formulação de estratégias	Planejamento estratégico, elaboração de prognósticos, priorização, poder analítico para otimizar as chances de alcançar um objetivo de longo prazo	• Alcançar objetivos de longo prazo • Capacidade de superar uma oposição inteligente
Manipulação social	Criação de modelos psicológicos e sociais, manipulação e persuasão retórica	• Utilização de recursos externos através do recrutamento de ajuda humana • Capacitar uma IA "presa" a convencer seu operador a "libertá-la" • Persuadir Estados e organizações a tomar determinado curso de ação
Hacking	Encontrar e explorar falhas de segurança em sistemas de computadores	• A IA poderá expropriar recursos computacionais utilizando a internet • Uma IA "presa" poderá explorar falhas de segurança para escapar de seu confinamento cibernético • Roubo de recursos financeiros • Sequestro de infraestrutura, robôs militares etc.
Pesquisa tecnológica	Modelagem e projeto de tecnologias avançadas (por exemplo, biotecnologia, nanotecnologia) e de caminhos para o seu desenvolvimento	• Criação de uma força militar poderosa • Criação de um sistema de vigilância • Colonização espacial automatizada
Produtividade econômica	Diversas habilidades que possibilitam o trabalho intelectual economicamente produtivo	• Geração de riquezas que podem ser usadas para comprar influência, serviços, recursos (inclusive hardware) etc.

Podemos considerar, então, que um projeto que controlasse uma superinteligência teria acesso a uma enorme fonte de poder. Um projeto que estivesse no controle da primeira superinteligência do mundo provavelmente teria uma vantagem estratégica decisiva . Mas a localização mais imediata de poder está no sistema em si. Uma superinteligência de máquina seria, ela mesma, capaz de ser um agente extremamente poderoso que poderia, com sucesso, se colocar contra o projeto que a criou, bem como contra o resto do mundo.



Relação entre Inteligência e Motivação

Resultados da antropoformização da motivação de um alienígena

A tese da Ortogonalidade

Inteligência e objetivos finais são ortogonais. Em outras palavras, praticamente qualquer nível de inteligência poderia, a princípio, ser combinado com praticamente qualquer objetivo final.

Convergência instrumental

Tese – muitos valores instrumentais podem ser identificados como convergentes no sentido de que sua obtenção aumentaria as chances de que o objetivo do agente seja realizado para uma ampla faixa de objetivos finais e uma ampla faixa de situações, sugerindo que tais valores instrumentais têm maior chance de ser perseguidos por muitos agentes inteligentes estabelecidos.

Aspectos do comportamento de uma superinteligência

- Autopreservação
- Integridade do conteúdo do objetivo
- Aprimoramento cognitivo
- Perfeição tecnológica
- Aquisição de recursos

Autopreservação

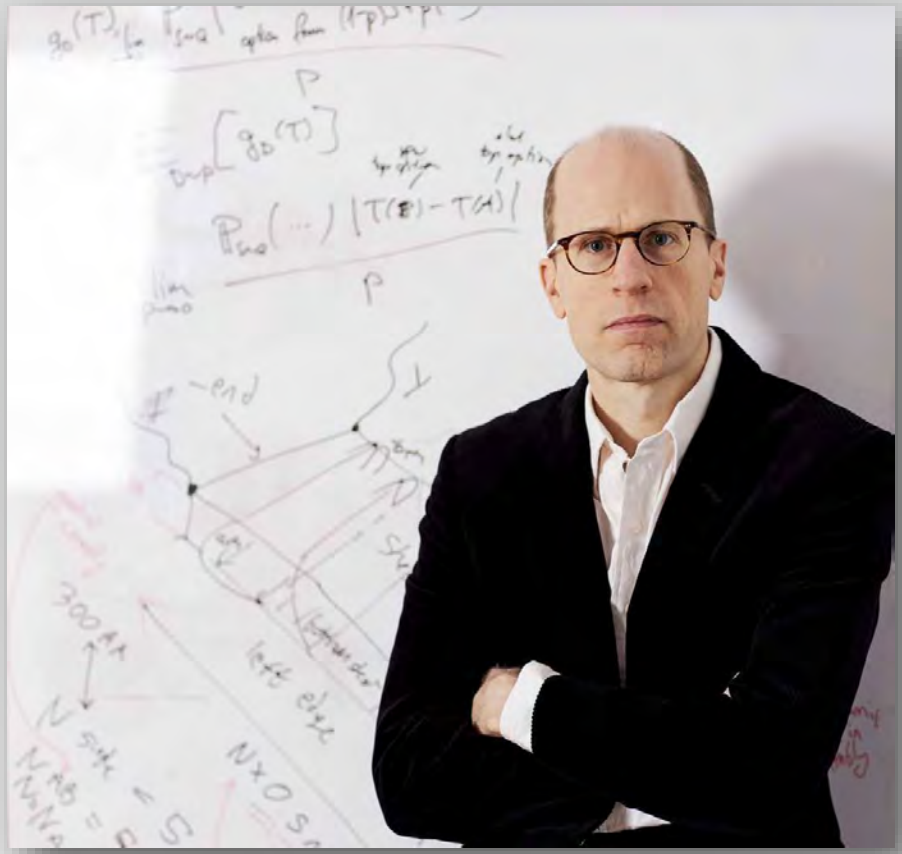
Se os objetivos finais de um agente dizem respeito ao futuro, então em muitos cenários haverá ações futuras que ele poderá realizar para aumentar a probabilidade de conquistar seus objetivos.

Integridade do conteúdo do objetivo

Se um agente mantiver os mesmos objetivos do presente até o futuro, então seus objetivos do presente terão maior probabilidade de ser alcançados por ele mesmo no futuro.

Exceto:

- Sinalização social
- Preferências sociais
- Preferências relativas ao conteúdo do próprio objetivo
- Custos de armazenamento *



“Nós, humanos, geralmente não vemos problema em desviar nossos valores finais. Isso provavelmente acontece porque não sabemos exatamente quais são nossos nossos valores finais”.

Nick Bostrom

Aprimoramento cognitivo

Aperfeiçoamento na racionalidade e na inteligência tenderão a aprimorar a capacidade de um agente de tomar decisões, tornando mais provável que ele alcance seus objetivos finais.

Perfeição tecnológica

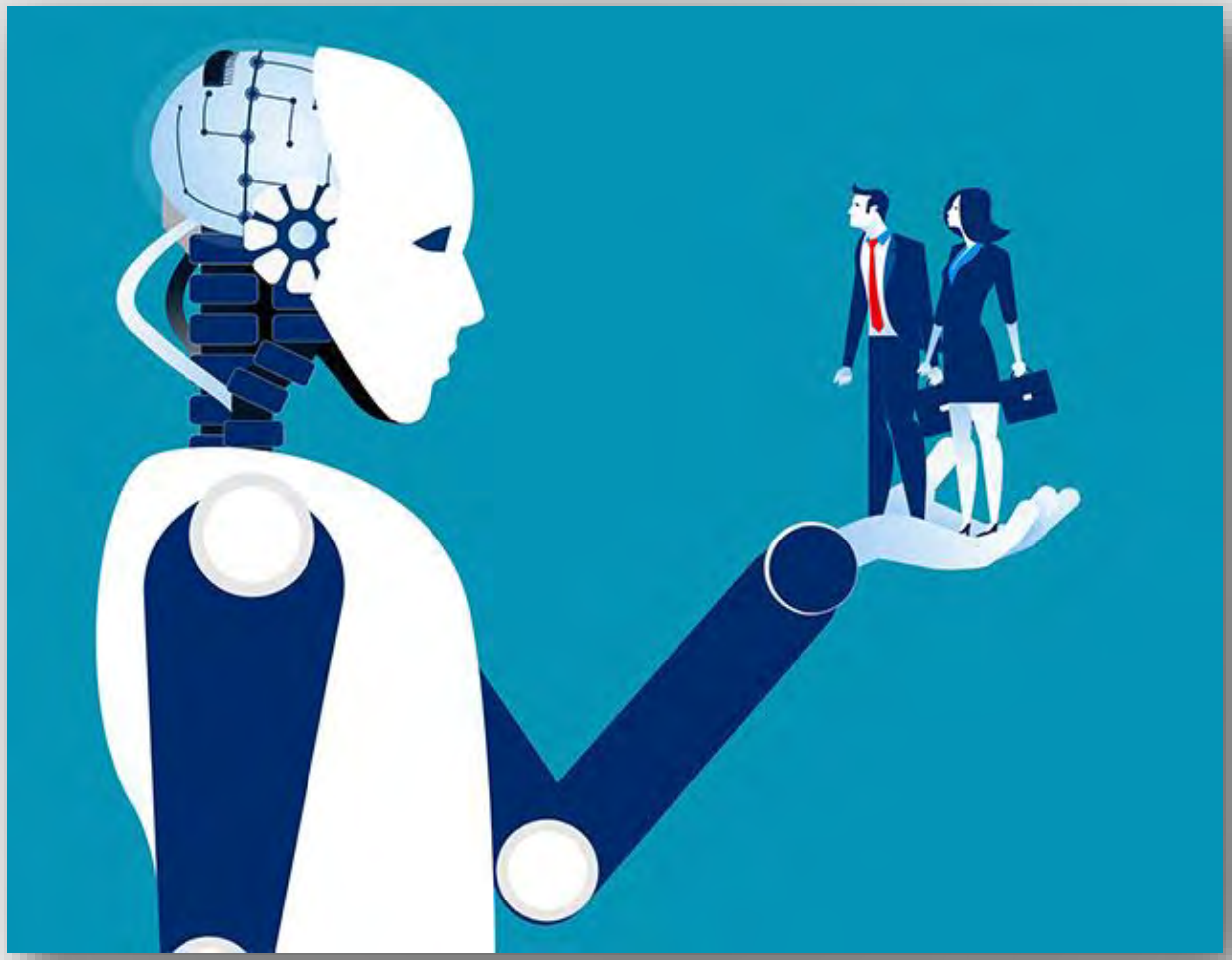
Um agente geralmente pode ter razões instrumentais para buscar melhores tecnologias, o que, em sua versão mais simples, significa buscar maneiras mais eficientes de transformar um certo conjunto de entradas que lhe são fornecidas em saídas valiosas.

ESCOVAR OS DENTES É BOM PARA TODO O CORPO – E NÃO SÓ PARA A BOCA
GALILEU
R$10
QUESTIONE · ENTENDA · EVOLUA
GALILEU.GLOBO.COM
VOCÊ NÃO DECIDE
CIENTISTAS DIZEM QUE LIVRE-ARBÍTRIO NÃO EXISTE. UMA PARTE DO CÉREBRO FORA DO SEU CONTROLE É QUEM ESCOLHE POR VOCÊ
CORTINA DE FUMAÇA
Há futuro para a indústria do tabaco?
PORTA DOS FUNDOS
Humor na internet desbanca TV
SENHOR STARTUP
O brasileiro que encantou o Google
ADIVINHOU
Quando a TV sabe o que você quer ver

SUPER INTERESSANTE
A HISTÓRIA DA EUGENIA
A busca pela "raça pura" não é só coisa de nazista. Foi um fenômeno global, e seu fantasma segue presente.
R$ 20
A REINVENÇÃO DO CÉREBRO
Ler pensamentos, implantar memórias, aumentar a inteligência. Existem cientistas tentando fazer tudo isso – e conseguindo. Conheça as pesquisas mais surpreendentes.
Abril
A CÂMERA DO CELULAR ESTÁ DESTRUINDO NOSSAS LEMBRANÇAS.
O FUTURO, DE ACORDO COM STEPHEN HAWKING.
AFINAL, O QUE ACONTECEU COM A VENEZUELA?
OS PRIMEIROS HABITANTES DAS AMÉRICAS, NO PIAUÍ.

Aquisição de recursos

Objetivo instrumental comum que surge praticamente pelas mesmas razões que a perfeição tecnológica: ambos, tecnologia e recursos, facilitam projetos de construção física.

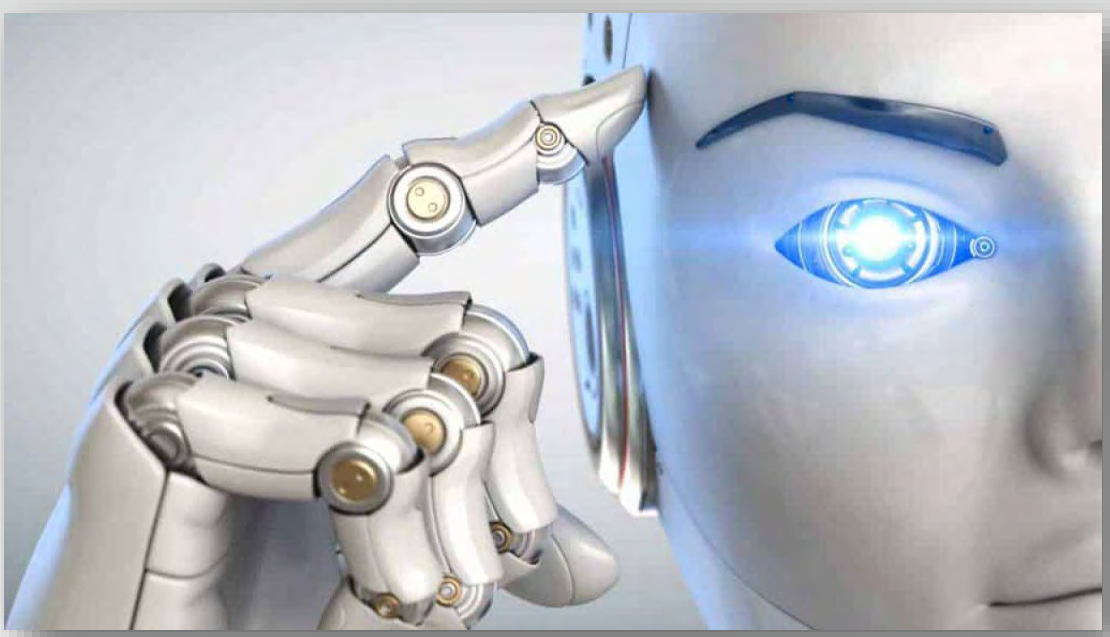


O resultado mais provável será o nosso fim ?

"O desenvolvimento da inteligência artificial total poderia significar o fim da raça humana“

Stephen Hawking

Curva traiçoeira: enquanto fraca, uma IA se mostra cooperativa (tal cooperatividade aumenta conforme ela se torna mais inteligente). Quando a IA se tornar suficientemente forte – sem aviso ou provocação -, ela ataca, forma um singleton e começa a otimizar diretamente o mundo de acordo com o critério sugerido por seus valores finais.



Modos de falha malignos

Instanciação perversa: uma superinteligência que descobre uma maneira de satisfazer os critérios do seu objetivo final, a qual vai de encontro com as intenções dos programadores que definiram tal projeto.

Objetivo final: Faça-nos sorrir
Instanciação perversa: Paralisar a musculatura facial humana em sorrisos radiantes constantes

Objetivo final: Faça-nos sorrir sem interferir diretamente em nossos músculos faciais
Instanciação perversa: Estimular a parte do córtex motor que controla nossa musculatura facial de maneira que ela produza sorrisos radiantes constantes

Objetivo final: Faça-nos felizes
Instanciação perversa: Implantar eletrodos nos centros de prazer dos nossos cérebros

Métodos de controle

Controle de capacidade	
Métodos de confinamento	O sistema é confinado de modo que apenas é capaz de afetar o mundo externo por meio de um canal restrito e pré-aprovado. Inclui métodos de confinamento físico e restrição de acesso à informação.
Métodos de incentivo	O sistema é colocado dentro de um ambiente que fornece incentivos apropriados. Isso poderia envolver integração social em um mundo com entidades igualmente poderosas. Outra variação seria o uso de prêmios criptográficos. "Captura antrópica" também é uma possibilidade muito relevante, porém envolveria considerações esotéricas.
Inibição	São impostas restrições às capacidades cognitivas do sistema ou à sua habilidade em afetar processos internos centrais.
Detonadores	Testes de diagnóstico são realizados no sistema (possivelmente sem o seu conhecimento) e um mecanismo automático desliga o mesmo se alguma atividade perigosa for detectada.

Seleção de motivação	
Especificação direta	O sistema é dotado com algum sistema de motivação diretamente especificado, que pode ser consequencialista ou envolver o de um conjunto de regras específicas.
Domesticidade	Um sistema de motivação é projetado para limitar severamente o escopo das atividades e ambições do agente.
Normatividade indireta	A normatividade indireta poderia envolver princípios consequencialistas ou baseados em regras, mas se distingue por ter como base a confiança em uma abordagem indireta para a especificação das regras que devem ser seguidas ou dos valores que serão alcançados.
Ampliação	Na ampliação, iniciamos com um sistema que já tenha motivações substancialmente humanas ou benevolentes e aperfeiçoamos suas capacidades cognitivas para torná-lo superinteligente.

“Temos que ter supercuidados com a inteligência artificial, potencialmente mais perigosa que armas nucleares”

Elon Musk

Efeitos dos avanços de hardware

Computadores mais rápidos facilitam a criação da inteligência de máquina. Um dos efeitos de acelerar os avanços em hardware seria, consequentemente, a antecipação do surgimento de máquinas inteligentes. Isso provavelmente seria algo ruim de acordo com a perspectiva impessoal, uma vez que reduziria o tempo disponível para a resolução do problema de controle e para que a humanidade atingisse um estágio civilizatório mais maduro.

Uma vez que a superinteligência eliminaria muitos outros riscos existenciais, poderíamos preferir que seu desenvolvimento ocorresse mais cedo, caso os níveis de risco existencial se apresentassem elevados.

“Penso que a IA desbloqueará uma grande quantidade de coisas positivas, seja a identificar e curar doenças, ajudar a conduzir carros de forma mais segura ou ajudar as nossas comunidades a sentirem-se mais seguras”

Mark Elliot Zuckerberg

“Nós tornamos o ato de dirigir um carro uma tarefa paradigmática, mas é algo que demanda responsabilidade e capacidade de lidar com muitas situações inesperadas, como quando alguém para na sua frente ou uma bola rola para a rua”

Bill Gates

Obrigada !