

AI Watch - A Evolução Histórica da Inteligência Artificial

Resumo e tradução feita por: Guilherme Peliceri

Contato: peliceriguilherme@gmail.com

Quem é AI Watch e quais os objetivos deste relatório?

AI Watch é uma iniciativa da Comissão Europeia que **monitora** o desenvolvimento, a aplicação e o impacto da inteligência artificial (IA) na União Europeia. O objetivo principal é fornecer análises baseadas em evidências para apoiar a **formulação de políticas** de IA na UE.

O **objetivo** do relatório é analisar a **evolução** das Inteligências Artificiais, e pontuar as principais **diferenças** e **similaridades** entre elas ao longo das gerações.

Sumário

- Fundamentos da IA (definições, métodos e aplicações).
- Os grandes períodos de desenvolvimento da IA (dividido em: Invernos como péssimos momentos e Primaveras como ótimos momentos).
- Análise sobre os modelos atuais.
- Uma comparação com o atual e o passado para ter insights de um possível futuro.

Observação

O relatório foi publicado em 2020, não contém atualizações até os dias atuais (julho de 2025).

Texto inteiramente revisado por IA, para evitar desinformação e minimizar lacunas de informação.

Toda informação foi retirada do material original.

1. Definições e Classificações Principais:

IA é definida como Inteligência demonstrada por uma máquina, mas também pode significar máquinas que imitam as nossas funções cognitivas, como pensar, aprender, entender.... Ela tem muitos significados, o que é diferente entre eles são as nuances.

Para a AI Watch, significa: *"Sistemas de inteligência artificial (IA) são sistemas de software (e possivelmente também de hardware) projetados por humanos que, dado um*

objetivo complexo, atuam na dimensão física ou digital ao perceber seu ambiente através da aquisição de dados, interpretar os dados estruturados ou não estruturados coletados, raciocinar sobre o conhecimento ou processar a informação derivada desses dados, e decidir a(s) melhor(es) ação(ões) a ser(em) tomada(s) para alcançar o objetivo determinado. Sistemas de IA podem tanto usar regras simbólicas quanto aprender um modelo numérico, e também podem adaptar seu comportamento ao analisar como o ambiente é afetado por suas ações anteriores."

A IA possui várias abordagens e técnicas para ser construída como o **Machine Learning** (exemplos como deep learning e reinforcement learning, pilar da IA moderna), **Machine Reasoning** (capacidade das máquinas de usar o raciocínio lógico e o raciocínio inferencial para resolver problemas e tomar decisões, pilar da IA Simbólica) e **Robotics** (uso de sensores para interação máquina e mundo real)

Artificial Narrow Intelligence (ANI): ou "Inteligência Artificial Limitada" são todas as IAs existentes hoje, que são projetadas e treinadas para executar uma tarefa específica ou um conjunto limitado de tarefas. São capazes de processar dados na velocidade da luz e fazer tarefas que humanos acham repetitivas de maneira muito eficiente.

Porém, as ANIs são **incapazes de generalizar dados**. Por exemplo, uma IA de reconhecimento facial não consegue transferir conhecimento para uma IA de geração de fala. O problema de generalização ainda não foi resolvido.

Artificial General Intelligence (AGI): ou "Inteligência Artificial Generalista" são as IAs capazes de fazer **qualquer** tarefa cognitiva de um ser humano, pois ao contrário das ANIs, elas conseguem generalizar os dados. Em filmes é mostrada como uma máquina que pensa, tem sentimentos e consciência de si mesma. Porém, não há nenhuma IA assim criada até hoje.

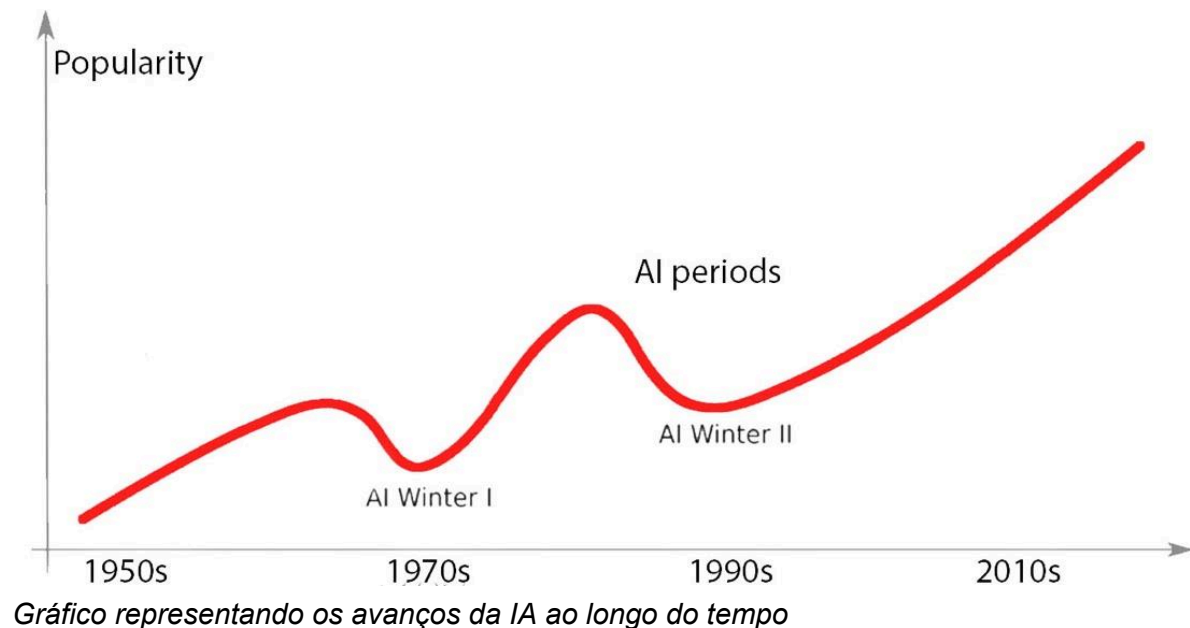
Artificial Superintelligence (ASI): ou Superinteligência Artificial é qualquer intelecto que **exceda** em muito o desempenho cognitivo dos humanos em praticamente todos os domínios de interesse. Ultrapassa a inteligência humana em todos os níveis, até mesmo dos mais brilhantes cientistas. O futuro da IA **termina** aqui.

Machine Learning: uma área da Inteligência Artificial, que estuda os algoritmos que sistemas de computador aprendem por meio da experiência. Os algoritmos de ML criam amostras de treino para aprender comportamentos padrões, para que assim, possam fazer previsões ou decisões apenas por ela mesma. Existem 4 categorias:

- **Supervised Learning Algorithms (Algoritmos de Aprendizado Supervisionado):** se baseia em mapear dados de entrada com dados de saída rotulados. Exemplo: precisamos saber se há um gato na imagem, e então mostramos várias imagens ao algoritmo. Aqueles que contém gatos possuem o rótulo "possui gato" e os que não tem, possuem o rótulo "não tem gato". Assim, o sistema aprende a reconhecer gatos em fotos desconhecidas. Precisa de muita quantidade de dados.
- **Unsupervised Learning Algorithms (Algoritmos de Aprendizado Não Supervisionado):** Não requer dados rotulados, na verdade os dados são dados ao sistema e ele tenta achar padrões e estruturas sozinho. Faz isso separando dados similares para formar "clusters".

- **Semi-Supervised Learning Algorithms (Algoritmos de Aprendizado Semi-Supervisionado):** é uma categoria entre supervisionada e não supervisionada, onde contém tanto dados rotulados como não rotulados. Os dados rotulados servem para treino e aprendizado, e os não rotulados são usados para teste e aprimoração
- **Reinforcement Learning (Aprendizado por Reforço):** o algoritmo aprende por tentativa e erro, caso faça ações certas, ganha recompensa, caso não, ganha punição. O sistema tenta sempre maximizar a quantidade de recompensa que consegue

A seguir, vamos tratar sobre onde surgiu tudo isso.



2. A Criação (1950~1970):

"Computing machinery and intelligence" (Turing 1950) - Artigo de Alan Turing, onde questiona "Máquinas podem pensar?". Então propôs o famoso teste de Turing, usado para medir a inteligência da máquina.

Conferência de Dartmouth (1956): a IA ganhou esse nome e sua missão. E nessa época, os sistemas começaram a resolver problemas como:

1952 - Checkers: primeiro programa a demonstrar aprendizado e não só fazer atividades. Aprendia a cada jogo de damas que realizava.

1955 - Logic Theorist: comprovou 38 teoremas matemáticos e introduziu conceitos como heurísticas, processamento de lista e raciocínio como busca, conceitos fundamentais para uma IA.

1957 - Perceptron: é, essencialmente, um modelo matemático de um único neurônio. Ele recebe vários sinais de entrada, pondera a importância de cada um e, se a soma desses sinais ultrapassar um certo limiar, ele "dispara" um sinal de saída. Criação do **Conexionismo** (não é necessário regras de lógica complexa, apenas muitas unidades

simples que aprendem com a experiência e se ajustam) e fundação das redes neurais (o Perceptron é o bloco de construção básico, o "átomo" das redes neurais).

1961 - MENACE: um dos primeiros programas a jogar perfeitamente o jogo *Tic-Tac-Toe*.

1965 - ELIZA: sistema de linguagem natural que imitava um médico. Acreditavam falar com uma pessoa de verdade, mas depois de um tempo, perdia o raciocínio.

1969 - Shakey: primeiro robô móvel que foi capaz de pensar antes de agir. Quando problemas diversos eram apresentados para ele, primeiro analisava o ambiente e depois planejava as ações. Usou áreas da robótica, visão computacional e linguagem natural.

1969 - Publicação do livro Perceptron: levou ao primeiro grande "inverno da IA" (período ruim), pois expôs as limitações dos perceptrons, que não eram capazes de fazer operações XOR, derrubando o imenso otimismo de toda a comunidade. O connexionismo foi taxado como beco sem saída e o financiamento dessas pesquisas (massivas, vindas do governo americano) desapareceu. O campo das redes neurais entrou em um longo período de estagnação que durou mais de uma década.

Na época, pesquisadores já acreditavam que seus algoritmos já estavam pensando, criando e aprendendo. Quando viram que o crescimento não estava tão avançado quanto pensavam, levou a todos um grande período de decepção.

O Primeiro Grande Inverno das IAs (1970)

O péssimo período das IAs decorreu por motivos de **Expectativa Elevada e Falta de Tecnologia**. Por prometerem coisas grandiosas e não entregarem, houve grande quebra de confiança e cortes de verba (por meio do Relatório ALPAC e Relatório *The Lighthill*). Além disso, os computadores na época eram simplesmente fracos demais, faltando poder de processamento, memória e velocidade.

Os pesquisadores também descobriram a distinção entre problemas polinomiais e exponenciais.

- **Problemas Classe P** - Polinomial: são problemas tratáveis pelo computador, pois mesmo que fiquem grandes, o tempo para resolver cresce de forma controlada.
- **Problemas Classe NP-Completo** - Exponencial: o tempo de solução explode. Problemas podem pular de 10 segundos para minutos, horas, anos, séculos...

O que aconteceu foi que a maioria dos problemas combinatoriais e lógicos eram problemas NP-Completo.

3. IA Simbólica (1970~1990):

Nos anos 80, o paradigma de IA mudou para **IAs simbólicas e sistemas especialistas**. A ideia essencial era colocar todo o conhecimento de especialistas, em diversas áreas, dentro do computador e espalhar como um programa para PC.

Tinha dois componentes:

- a base de conhecimento vinda de especialistas humanos, com fatos, regras e relações. Geralmente no formato "**SE (IF) -> ENTÃO (THEN)**".
- o algoritmo de **inferência** que descrevia como manipular e combinar esses dados.

Muitas empresas venderam esses produtos nos anos 90, mas logo ficou claro sua **limitação fatal**. O conhecimento de um especialista de verdade é muito mais do que uma lista de regras. Envolve intuição, bom senso, experiência e a capacidade de lidar com situações novas e inesperadas. Tentar traduzir toda essa complexidade em regras rígidas de "IF-THEN-ELSE" era extremamente **difícil, caro** e, muitas vezes, **impossível**.

Marcos Importantes:

1970 - MYCIN: foi um sistema especializado no diagnóstico de doenças de sangue e prescrição de medicamentos. Teve boa precisão no cálculo de incerteza e acertava bem nos diagnósticos.

1968~1970 - SHRDLU: foi o auge da IA simbólica, permitia o usuário a ter uma conversa em inglês com o computador e o comandar dentro de um mundo de formas geométricas (mas era limitado apenas dentro desse mundo bem definido)

1972 - A linguagem Prolog foi criada.

1979 - O carrinho de Stanford conseguiu cruzar uma sala evitando obstáculos sem ajuda, conceito de carro autônomo.

1982 - Hopfield Net: era um novo tipo de rede neural criada. Era capaz de recuperar dados limpos e completos de dados incompletos. Backpropagation foi um algoritmo que resolveu um problema fundamental do livro Perceptrons. Propôs uma maneira eficiente de treinar redes neurais com múltiplas camadas e minimizar erros de previsão.

Ressuscitou o Conexionismo.

1983 - Surgimento do ID3: um algoritmo que gera árvores de decisão a partir de um conjunto de dados. Precursor do C4.5, um algoritmo de IA que usamos até hoje.

Houve muito investimento, e a confiança na IA voltou! Porém...

O maior **problema** dos sistemas especialistas era **conseguir conhecimento**. Conseguir tempo para os especialistas ensinarem as máquinas era difícil pois eles eram muito requisitados dentro da empresa. As pesquisas dos sistemas especialistas focaram em ferramentas para conseguir de forma mais eficiente o conhecimento. Mas, enquanto isso ocorria, as linguagens de programação desses sistemas ficaram para trás, não conseguia acompanhar o desempenho da linguagem C, e ficaram **ultrapassados**.

Enquanto isso, cientistas continuaram pesquisando sobre máquinas inteligentes, mas não deram o nome de IA, pois a decepção já estava atrelada a esse nome.

4. Machine Learning e Deep Learning (1990~Atualmente):

Nos anos 90~2010 a IA ganhou força novamente. Aplicações foram usadas para mineração de dados, robótica industrial, logística, inteligência de negócios, software bancário, diagnóstico médico, sistemas de recomendação e mecanismos de busca. Pesquisadores descobriram que muitos dos **problemas** da IA já tinham sido **resolvidos** nas áreas da **matemática, economia e pesquisas operacionais**.

Mas, o nome IA não era usado, pois ainda havia más lembranças no público geral, então nomes como informática, sistemas baseados em conhecimento, otimização de algoritmos... surgiram. Assim, fundos de **investimento** vieram tranquilamente.

Em 2006, a professora da Universidade de Stanford, Fei-Fei Li contribuiu para a **mudança de paradigma** da IA. O pensamento na época era *"Não conseguimos analisar uma única imagem direito, quem dirá milhares. Temos que melhorar o algoritmo!"*. O pensamento da professora era o contrário, o problema não era o algoritmo, mas sim a **falta de dados**. *"Para que uma IA aprenda a ver como um humano, ela precisa ver uma quantidade de exemplos tão vasta e variada quanto um humano vê"*. De maneira mais técnica, foi uma mudança de abordagem **"model-driven"** (focada em criar melhores algoritmos) para **"data-driven"** (focada em usar mais dados).

Criação da ImageNet: para provar sua teoria, ela liderou um projeto monumental, uma base de dados gigantesca e gratuita com **milhões** de imagens da internet, rotuladas por humanos. O projeto foi lançado em 2009.

Para estimular ainda mais a pesquisa, a equipe da ImageNet criou uma Olimpíada de Visão Computacional (a ILSVRC). Em 2011, os melhores sistemas criados tinham 25% de taxa de erro, erravam 1 entre 4 imagens.

Big Bang de 2012: uma equipe da Universidade de Toronto, liderada por Geoffrey Hinton, criou a AlexNet, que atingiu incríveis 16% de taxa de erro. Em alguns anos, a taxa de erro das IAs caiu para porcentagens mínimas. E isso fez com que o **paradigma da IA mudasse** para **Deep Learning (DL)**.

Deep Learning é uma **sub-área** de Machine Learning e IA. Introduziu uma arquitetura de multi camadas de redes neurais que aprendem a reconhecer padrões em diferentes níveis de abstração (o aprendizado das camadas se somam, a primeira aprende o básico, a segunda pega esse básico e soma mais aprendizado...). Arquiteturas de redes neurais **DL** incluem **redes neurais profundas, redes de crenças profundas, redes neurais recorrentes e redes neurais convolucionais (CNN)**.

Mudança no paradigma de programação de IAs:

O modelo clássico da programação de IAs era dar os dados e definir as regras para manuseio dos mesmos, e assim o modelo entregava as respostas (dado + regra = resposta). A mudança foi entregar as respostas junto com os dados, para que assim o modelo raciocine as regras sozinho (**dados + resposta = regras**). Assim o sistema treina, ao invés de ser programado. Além disso, houve imensa melhoria quando as regras passaram de IFs, ELSEs e THENs para **pesos**. Um sistema de Deep Learning pode ter bilhões de parâmetros de pesos.

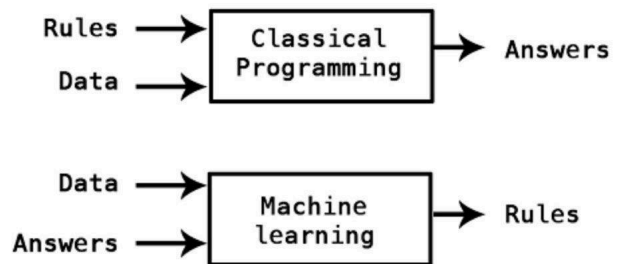
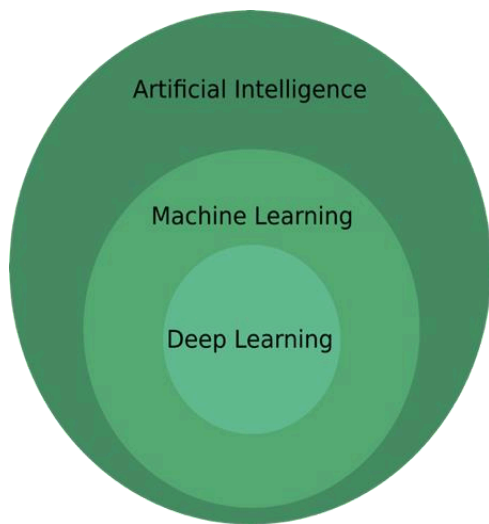
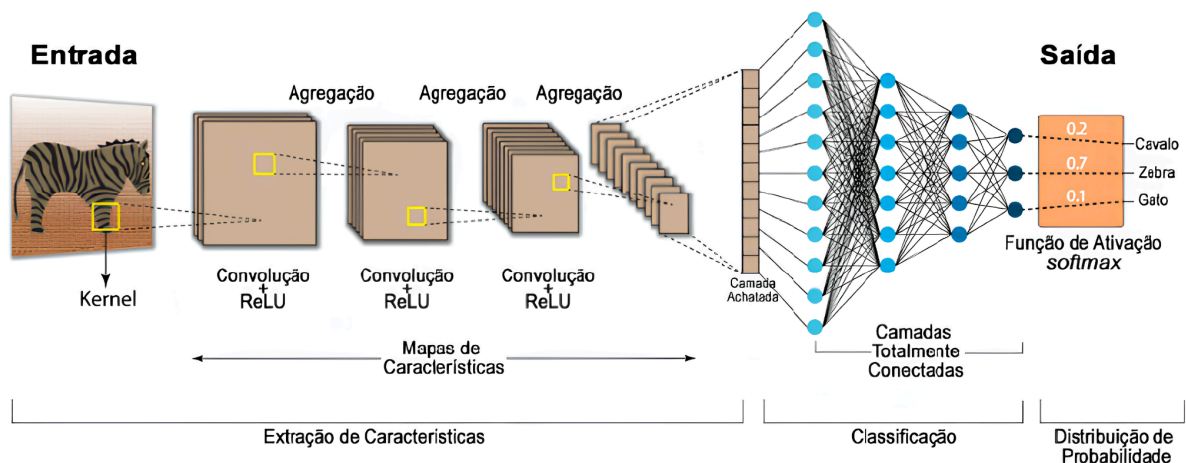


Imagem 1 mostrando como a deep learning é uma pequena área dentro da IA
Imagem 2, demonstrando qual foi a mudança de pensamento para desenvolver o sistema

Diferenciando-as:

Problema - Reconhecer gatos em imagens

Algoritmos clássicos iriam aplicar diferentes métodos para identificar cada parte do gato (olhos, boca, formato), necessitando de códigos grandes e complexos, enquanto a **Deep Learning** quase automaticamente já reconhece a imagem (nesse caso, modelos convolucionais seriam mais adequados). Ela automaticamente reconhece os gatos após analisar vários exemplos de imagens com e sem gatos.



exemplo do modelo convolucional

Técnica Transfer Learning: Consiste em pegar IAs **previamente treinadas** em problemas gerais (ex: reconhecimento de imagens diversas) e depois, **treiná-la** para a finalidade **específica** necessária (ex: treinar uma IA boa em reconhecimento de imagens no geral e ensiná-la a reconhecer gatos). Visivelmente aumenta a velocidade de treino, salva recurso computacional e produz modelos com mais qualidade.

Hoje a evolução das IAs é muito contribuída pela disponibilidade de informação na internet, como códigos, frameworks, artigos científicos, datasets.

- Github para códigos
- Frameworks de IA disponíveis para download na internet
- Artigos científicos no arXiv
- Competições e códigos no Kaggle

O conteúdo aberto ao público contribuiu muito para a **proliferação e acolhimento** das IAs.

Acontecimentos importantes:

1989 - Sistemas processaram cerca de 10-20% dos cheques descontados à mão e códigos postais nos Estados Unidos.

1989 - Invenção do Q-Learning (Learning from Delayed Rewards): basicamente permite a IA fazer ações para obter recompensas maiores no futuro, e não recompensas pequenas e imediatas. É um método "**model-free**" (livre de modelo). Isso significa que o agente não precisa de um mapa ou de um entendimento prévio das regras do ambiente para aprender, o que foi uma revolução em termos de praticidade para o Aprendizado por Reforço.

1993 - Schmidhuber: o algoritmo resolveu uma tarefa de conhecimento muito profundo.

1995 - Aplicação de Support Vector Machines (SVMs ou Máquinas de Vetores de Suporte): algoritmos extremamente metódicos de classificação que usam o truque chamado "Kernel Trick" (permite ao SVM mapear os dados para dimensões ainda mais altas, encontrando maneiras de separar dados que parecem inseparáveis em dimensões mais baixas).

1995 - um carro autônomo percorreu 4,500km apenas com uma pessoa nos freios e acelerador.

1996 - Garry Kasparov, maior jogador de xadrez da época perdeu para a máquina. O sistema de IA Simbólica apenas aprendeu todos os movimentos do jogo e os calculou.

1997 - A invenção de uma arquitetura de Rede Neural Recorrente (RNN) muito mais poderosa, chamada **LSTM**.

1998 - Melhoria no Aprendizado Baseado em Gradiente: combinou de forma eficiente o algoritmo Stochastic Gradient Descent (SGD) com o Backpropagation.

2002 - TD-Gammon: o sistema atingiu melhores resultados em gamão do que os melhores do mundo, usando a combinação de Rede Neural + Aprendizado por Reforço + Auto-jogo (self-play, a máquina jogou milhões de vezes contra ela mesma para aprender).

2005 - O robô de Stanford andou corretamente e sozinho 131 milhas numa trilha do deserto não mapeada.

2009 - ImageNet lançou um dataset com 3.2 milhões de dados gratuitamente.

2011 - O supercomputador **Watson**, da IBM, venceu os dois maiores campeões humanos do programa de TV americano Jeopardy!, usando conceitos avançadíssimos da IA Simbólica.

2012 - AlexNet ganhou a competição da ImageNet, apresentando o Deep Learning ao mundo.

2014 - GANs (Redes Adversárias Generativas):

As GANs introduziram uma arquitetura de Deep Learning com duas redes neurais competindo entre si: uma "Geradora", que cria conteúdo (como imagens falsas de rostos), e uma "Discriminadora", que tenta identificar se o conteúdo é real ou falso.

2015 - DeepRL: masterizou cerca de 2600 jogos de Atari, com apenas poucos pixels e a pontuação como input.

2016 - AlphaGo: derrotou o maior campeão de Go do mundo.

2017 - Criação dos Transformers, arquitetura Deep Learning baseada em autoatenção.

2018 - OpenAI Five: derrotou um time amador de Dota 2.

2019 - AlphaStar: derrotou um profissional em Starcraft 2.

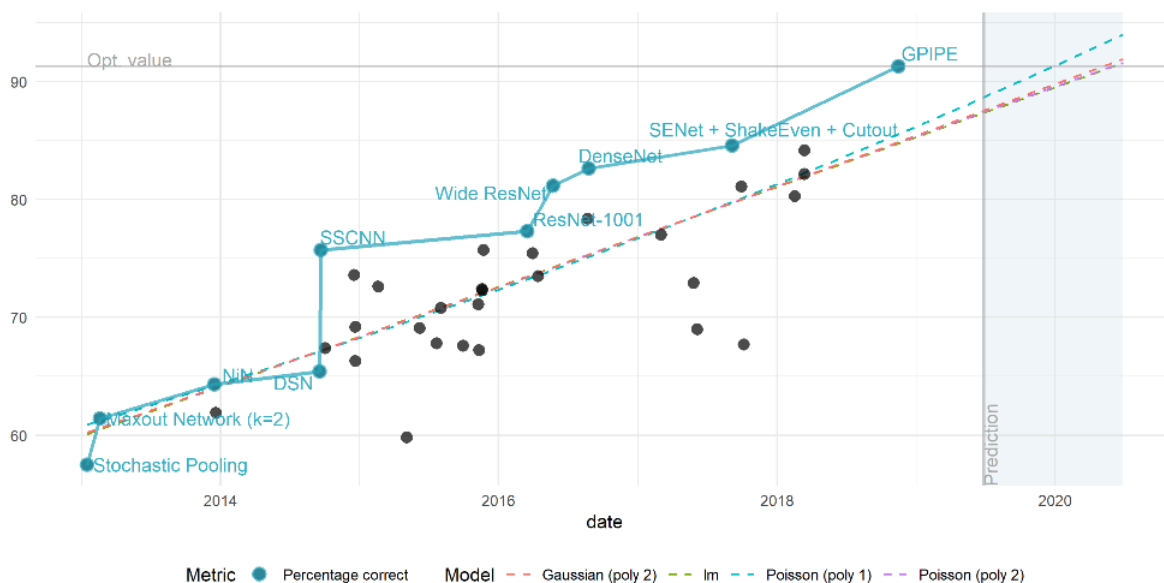
2019 - OpenAI criou o **Dactyl**, um robô humanóide que manipula objetos no mundo real, usando Reinforcement Learning.

2019 - GPT-2: lançado ao mercado.

Visivelmente a IA está ocupando cada vez mais todas as áreas de estudo humano, complicando nosso **controle** sobre sua **evolução**.

O site Papers With Code produz conteúdo aberto sobre os mais recentes estudos da IA no mundo

O AI Collaboratory é uma iniciativa colaborativa focada em analisar, avaliar, comparar e classificar algoritmos de Inteligência Artificial. Este projeto foi integrado ao AI WATCH, que é o serviço oficial da Comissão Europeia para o monitoramento e análise de tecnologias de IA. A imagem abaixo é uma pesquisa realizada pelo projeto:



Linha azul representa os avanços mais significativos dos sistemas, e as linhas coloridas representam o potencial de crescimento das mesmas. Data x Porcentagem de Acertos.

Sistemas de tradução avançaram conforme O **BLEU** (Bilingual Evaluation Understudy Score, métrica para avaliação da frase gerada em comparação com a frase de referência) foi melhorado. **SOTA**, sistema para tradução de inglês para francês e alemão.

Os modelos de linguagem receberam grande melhoria com a implementação de transformers na IA BERT e GPT-2

A AI Watch trouxe (página 17-18 do material original) opiniões de pessoas brilhantes e influentes sobre as IAs no futuro. Todas elas **focam** no alcance da **superinteligência**, dizendo que a quarta **revolução industrial** já está prevista, e que o cotidiano de todos será mudado, em todos os aspectos. Mas também mostram preocupações em relação ao poder que as máquinas vão possuir.

- O Que Falta para a IA Geral (AGI)?

Cientistas e pensadores concordam que o Deep Learning (DL) **sozinho** pode **não** ser **suficiente**. As principais propostas para o próximo passo incluem:

Modelos Híbridos (Gary Marcus): A ideia de combinar o Deep Learning com abordagens mais antigas, como a manipulação de símbolos (IA simbólica), para dar às máquinas a capacidade de raciocinar de forma mais estruturada.

Raciocínio Causal (Judea Pearl): A crença de que a IA só atingirá o nível humano quando for capaz de entender causa e efeito (ex: saber que malária causa febre, não apenas que as duas coisas aparecem juntas) e fazer perguntas contrafactuais ("*o que aconteceria se...?*").

Sistema 1 vs. Sistema 2 (Daniel Kahneman): A IA atual é excelente em tarefas do Sistema 1 (pensamento rápido, intuitivo, reconhecimento de padrões), mas ainda falha em tarefas do Sistema 2 (pensamento lento, lógico e deliberado), que é crucial para a inteligência geral.

O Ciclo de Hype e os Riscos

Expectativas exageradas podem ser perigosas, e como visto nos invernos das IAs anteriores, são bem comuns no desenvolvimento de novas tecnologias. O que dizem especialistas:

Lei de Amara (citada por Rodney Brooks): "*Nós tendemos a superestimar o efeito de uma tecnologia no curto prazo e a subestimar o efeito no longo prazo.*" Isso explica por que a IA passa por ciclos de grande promessa, seguidos de decepção e, eventualmente, um progresso que supera as expectativas originais.

Ciclo de Hype da Gartner: A Lei de Amara é semelhante ao padrão gráfico do Ciclo de Hype, que mostra tecnologias passando por um pico de "*hype*", um vale de desilusão e, finalmente, um platô de produtividade.

Riscos Reais (Stuart J. Russell): Independentemente da velocidade do progresso, pensadores como Russell alertam que os riscos de uma IA avançada para a humanidade são reais e precisam ser levados a sério.

Investimentos

EUA: Liderança do Setor Privado

O investimento em IA nos EUA é impulsionado principalmente por grandes **empresas privadas** de tecnologia. Gigantes como Google (que comprou a DeepMind), Facebook, Microsoft (que investiu \$1 bilhão na OpenAI), Apple e Amazon lideram o caminho, contratando os principais cientistas e investindo bilhões de dólares. O capital de risco (Venture Capital) para IA também cresceu drasticamente, aumentando **350%** entre 2013 e 2017.

China: Ambição Estratégica do Governo

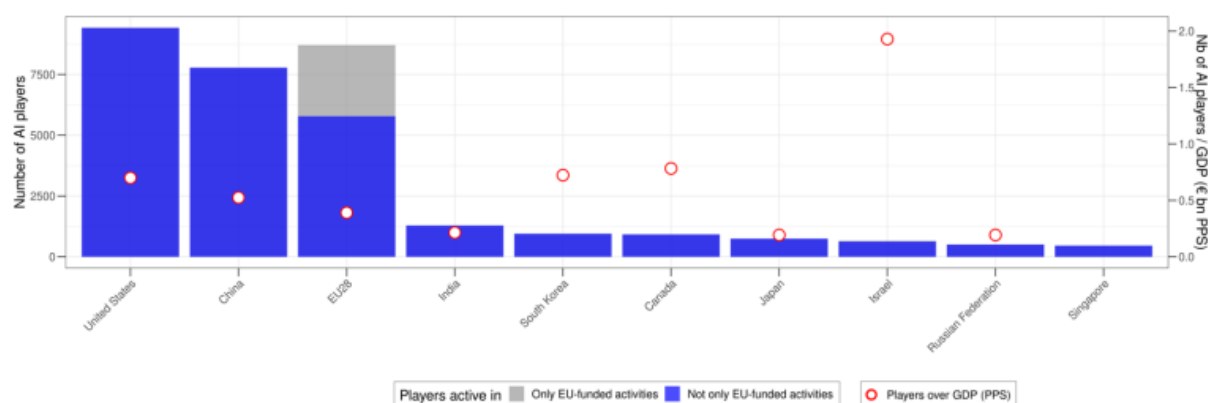
Na China, o ecossistema também é forte, com líderes como Tencent, Baidu e Alibaba. No entanto, o diferencial é um plano estratégico ambicioso do governo chinês, com metas claras em três etapas:

- Até 2020: Igualar-se aos países mais avançados.
- Até 2025: Atingir a liderança mundial em algumas áreas de IA.
- Até 2030: Tornar-se o principal centro de inovação em IA do mundo.

O plano inclui metas financeiras extremamente agressivas, projetando uma produção industrial relacionada à IA de mais de 1,35 trilhão de euros até 2030.

Europa: Esforços Coordenados e Fragmentados

A União Europeia (UE) está tentando aumentar seus investimentos, mas eles são baixos e fragmentados em comparação com EUA e China. A Comissão Europeia propôs dedicar pelo menos €1 bilhão por ano para IA através de programas como o Horizon Europe e o Digital Europe. A meta é aumentar progressivamente o investimento para €20 bilhões por ano na próxima década, combinando fundos da UE e dos estados-membros, como a França, que alocou €1,5 bilhão. A iniciativa AI Watch foi criada para monitorar esses esforços.



Número de players por país

Os Estados Unidos são especializados em Serviços com IA, Robótica e Automação e forte presença em várias outras áreas. Já a China é especializada em visão computacional, Machine Learning e veículos autônomos. Já a Europa lidera com sua rede de pesquisa e publicações, além de robótica e serviços com IA.

Mas claro, muitos outros países como Índia, Canadá, Japão... estão com suas pesquisas avançadas no assunto.

O artigo também traz alguns levantamentos interessantes sobre o tema:

Pesquisa Global (MIT Technology Review)

Esta pesquisa com executivos do mundo todo revela um **grande otimismo** sobre a IA:

- **Visão Positiva:** **81%** acreditam que a IA terá um impacto positivo e **83%** a consideram uma tecnologia transformadora ("game-changer").
- **Intenção de Adoção:** **78%** foram instruídos a explorar a IA e **64%** consideram investir na tecnologia.
- **Foco em Dados:** Mais de **85%** veem os dados como o recurso mais importante para o negócio.

Em resumo: Líderes globais estão muito otimistas e prontos para investir em IA.

Pesquisa na Europa (Comissão Europeia)

Esta pesquisa, focada na União Europeia, mostra um cenário mais complexo, com uma lacuna entre o conhecimento e a prática:

- **Conscientização vs. Adoção:** Embora a conscientização sobre a IA seja alta (**78%**), a adoção real é moderada (**42%**). Um número significativo de empresas (**40%**) não usa nem planeja usar a tecnologia.
- **Estratégia:** A maioria (**59%**) prefere comprar soluções de IA prontas em vez de desenvolvê-las internamente.
- **Principais Barreiras:** Os maiores obstáculos para a adoção são:
 - o **Internos:** Dificuldade para contratar talentos (**57%**) e o alto custo (**52%**).
 - o **Externos:** Incertezas sobre responsabilidade legal (**33%**) e falta de padronização de dados (**33%**).

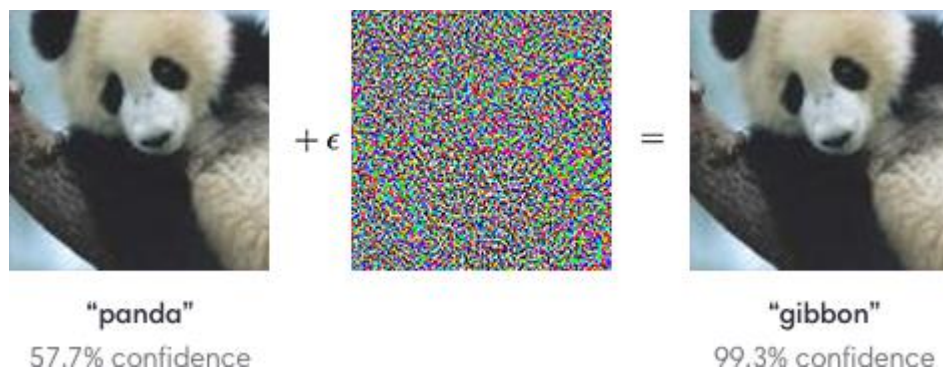
Problemas a se lidar

As IAs modernas trouxeram **praticidade** para nosso **cotidiano**, mas com ela, outros **problemas** surgiram:

Ataques adversários:

Técnica que tenta **explorar vulnerabilidades** da IA com dados corruptos/errôneos, fazendo-a cometer erros e até entregar informações importantes.

O exemplo dado aqui foi o da imagem de um panda, onde o adversário coloca uma pequena perturbação por cima, fazendo com que o sistema interprete a imagem como um gibão.



Perturbação quase invisível na imagem confunde as predições da IA

Deepfake:

A IA facilitou e muito o processo de criar **informações e campanhas falsas**, com sua capacidade rápida de criar falsas evidências. Com o passar dos anos, acreditar em fatos online será cada vez mais difícil.

Imagens e vídeos fakes, conteúdo de spam e phishing e roubo de voz são umas das possibilidades que adversários estão explorando.

Éticas do uso da IA:

Pelo seu imenso poder, cientistas passaram a se preocupar sobre **questões éticas** do **uso e desenvolvimento** das Inteligências Artificiais.

A Comissão Europeia definiu para a HLEG (grupo de experts que aconselham as estratégias do uso de IA) a tarefa de definir regras éticas para garantir confiança nas máquinas.

A conclusão foi, para que um sistema de IA seja confiável, ele deve **garantir** os três componentes a seguir durante todo o ciclo de vida do sistema:

Lícito, de acordo com todas as leis e regulamentos aplicáveis

Ética, garantindo adesão com princípios e valores éticos

Robusto, tanto do ponto de vista técnico quanto social, uma vez que, mesmo com boas intenções, um sistema de IA pode causar danos não intencionais

Agora os princípios éticos para o **desenvolvimento e implantação das IAs** de maneira **confiável** são:

- 1) Respeito pela autonomia humana,
- 2) Prevenção de danos,
- 3) Justiça,

4) Explicabilidade.

Conforme estabelecido acima, embora muitas obrigações legais reflitam princípios éticos, a adesão a princípios éticos **vai além do mero cumprimento formal das leis existentes**. Eles são especificados como imperativos éticos, aos quais os profissionais de IA devem sempre se esforçar para aderir.

Diferentes grupos de partes interessadas têm diferentes papéis a desempenhar para garantir que os requisitos da IA sejam atendidos: **(a) Desenvolvedores** devem **implementar** e **aplicar** os requisitos aos processos de design e desenvolvimento; **(b) Implantadores** (aqueles que usam a IA) devem garantir que os sistemas que utilizam e os produtos e serviços que oferecem **atendam aos requisitos**; e **(c) Usuários finais** e a sociedade em geral devem ser informados sobre esses requisitos e poder **solicitar** que sejam **cumpridos**.

Sejam os requisitos:

Agência e Supervisão Humana: Humanos devem ter o controle final sobre os sistemas de IA.

Robustez e Segurança Técnica: O sistema deve ser seguro, confiável e ter um "plano B" em caso de falhas.

Privacidade e Governança de Dados: Os dados dos usuários devem ser protegidos e usados de forma correta.

Transparência: Deve ser claro como o sistema funciona, como seus dados são usados e quais são suas limitações. Suas decisões devem ser rastreáveis.

Diversidade, Não Discriminação e Justiça: O sistema deve evitar preconceitos (bias), ser acessível a todos e incluir a participação de diversas partes interessadas.

Bem-estar Social e Ambiental: A IA deve ser usada para o bem da sociedade e do meio ambiente.

Responsabilização (Accountability): Deve haver mecanismos para responsabilizar as pessoas pelos resultados da IA, auditar suas decisões e oferecer reparação em caso de danos.

A visão da UE é que isso deve ser uma **responsabilidade compartilhada**: desenvolvedores precisam **construir** a IA da forma correta, empresas que a utilizam precisam **garantir** que ela seja confiável, e a sociedade deve ser informada e ter o poder de **exigir** que esses princípios sejam cumpridos.

Esses são os princípios éticos definidos pela União Europeia, mas existem outros como os princípios de Beijing, Google, Asilomar (feito por pesquisadores acadêmicos e da indústria de IA), OCDE...

No final, cada princípio leva, mais ou menos, os mesmos objetivos.

5. Resumo Geral:

O texto compara os três grandes períodos da IA, notando que os dois primeiros seguiram um padrão cíclico: **avanço científico** → **previsões ousadas** → **investimento massivo** → **decepção** → **"Inverno da IA"**.

A era atual, no entanto, apresenta diferenças cruciais:

- **Investimento:** Antes, os governos eram os principais investidores. Hoje, o investimento é liderado por **empresas privadas** gigantes (Google, Microsoft, etc.).
- **Pesquisa:** A pesquisa básica, antes domínio da academia, agora também é liderada por essas grandes empresas, que atraem os melhores talentos com recursos computacionais de ponta e vastos conjuntos de dados.
- **Paradigma:** A IA evoluiu de sistemas simbólicos e "baseados em modelos" para a abordagem atual, que é "orientada a dados" (**data-driven**), e isso foi o que tornou a Deep Learning viável..

Capacidades Atuais vs. Limitações Profundas

A IA moderna alcançou um desempenho sobre-humano em tarefas específicas e de grande escala:

- Reconhecimento de objetos em bilhões de imagens.
- Tradução instantânea de mais de 100 idiomas.
- Filtragem de spam em bilhões de e-mails.
- Domínio de jogos complexos como xadrez e Go.

Apesar disso, o texto enfatiza uma grande limitação: **a IA mais avançada de hoje ainda não possui as capacidades de bom senso de uma criança de quatro anos** para se mover, falar, entender o mundo, a causalidade ou discutir conceitos abstratos.

Riscos, Ética e o Ciclo do "Hype"

O documento alerta para os aspectos negativos e os riscos da IA:

- **Riscos:** A IA pode levar a resultados enviesados e não transparentes, além de poder ser usada para fins maliciosos, como **fake news, deep fakes e ciberataques automatizados**.
- **Ética e Regulação:** Para mitigar esses riscos, há um grande esforço em criar frameworks de ética focados em **justiça, responsabilidade e transparência**, que servirão de base para futuras regulações governamentais.

- **Hype vs. Realidade:** O texto adverte sobre o perigo de promessas exageradas, que no passado levaram aos "Invernos da IA". É preciso ter uma visão equilibrada, reconhecendo tanto as oportunidades quanto as limitações da tecnologia para evitar uma nova onda de desilusão.

Soberania e o Futuro Incerto

- **Estratégia Europeia:** A Europa está focada em criar sua própria estratégia de dados e IA, buscando **soberania tecnológica e de dados** para não depender de EUA e China. A iniciativa **AI Watch** foi criada para monitorar esse progresso.
- **Dois Caminhos Possíveis:** O futuro da IA é incerto e pode seguir dois caminhos:
 1. Um **novo "Inverno da IA"**, causado por decepções e promessas não cumpridas.
 2. Um **"Verão da IA" ainda maior**, com avanços ainda mais transformadores.

Dada essa incerteza, é crucial ter iniciativas como o AI Watch para monitorar a evolução da IA e seus impactos na sociedade.

Matéria Original:

Este documento é um resumo e uma análise do relatório 'Historical Evolution of Artificial Intelligence', publicado originalmente pela iniciativa AI Watch do Joint Research Centre (JRC) da Comissão Europeia. Todo o crédito pelo trabalho original pertence aos seus autores. Fonte original disponível em:

Historical Evolution of Artificial Intelligence -
<https://publications.jrc.ec.europa.eu/repository/handle/JRC120469>