Matteo Pasquinelli
Universidade Ca' Foscari

Tradução:
Bernardo Oliveira &
Bernardo Girauta

Máquinas que transformam a lógica: redes neurais e a automação distorcida da inteligência como inferência estatística

Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference

Máquinas que transforman la lógica: redes neuronales y la automatización distorsionada de la inteligência como inferencia estadística

ISSN: 2175-8689

RESUMO:
Tomando como ponto de partida o modelo do Perceptron, descrito por Frank Rosenblatt em 1957, o artigo apresenta uma breve história do desenvolvimento das redes neurais de indução estatística, forma hegemônica da dita "inteligência artificial" computacional atual, diferenciando-as das tentativas malsucedidas de automação da dedução simbólica. O autor enfatiza que a operação das redes neurais não corresponde a nenhuma definição exaustiva de inteligência, limitando-se à inferência estatística de certos conjuntos de dados de treinamento. Por fim, ele aponta que as formas de automação da indução estatística são parte de um processo mais amplo de automação do trabalho de modo geral.
PALAVRAS-CHAVE: *Perceptron; redes neurais; inteligência artificial; automação.*

ABSTRACT:
Taking as a starting point the Perceptron model, described by Frank Rosenblatt in 1957, the article presents a brief history of the development of statistical induction neural networks, a hegemonic form of the so-called current computational "artificial intelligence", differentiating them from unsuccessful attempts to automate symbolic deduction. The author emphasizes that the operation of neural networks does not correspond to any exhaustive definition of intelligence, being limited to statistical inference from certain training datasets. Finally, he points out that the forms of statistical induction automation are part of a broader process of labor automation.
KEYWORDS: *Perceptron; neural networks; artificial intelligence; automation.*

RESUMEN:
Tomando como punto de partida el modelo Perceptron, descrito por Frank Rosenblatt en 1957, el artículo presenta una breve historia del desarrollo de las redes neuronales de inducción estadística, forma hegemónica de la actual "inteligencia artificial" computacional, diferenciándolas de los fallidos intentos de automatización de la deducción simbólica. El autor destaca que el funcionamiento de las redes neuronales no corresponde a ninguna definición exhaustiva de inteligencia, sino que se limita a la inferencia estadística a partir de determinados conjuntos de datos de entrenamiento. Finalmente, señala que las formas de automatización de la inducción estadística son parte de un proceso más amplio de automatización del trabajo en general.
PALABRAS CLAVE: *Perceptron; Redes neuronales; inteligencia artificial; automatización.*

Submetido em 04 de agosto de 2023
Aceito em 11 de abril de 2024[1]

[1] Artigo Originalmente publicado em: https://www.glass-bead.org/article/machines-that-morph-logic/?lang=enview.

> *Perceptrons [redes neurais artificiais] não se destinam a servir como cópias detalhadas de nenhum sistema nervoso real. São redes simplificadas, projetadas para permitir o estudo de relações válidas entre a organização de uma rede nervosa, a organização de seu ambiente e as performances "psicológicas" das quais a rede é capaz. Perceptrons podem realmente corresponder a partes de redes mais amplas em sistemas biológicos... Mais provavelmente, eles representam simplificações extremas do sistema nervoso central, no qual algumas propriedades são exageradas, outras suprimidas.* - Frank Rosenblatt[2]

> *Não existe algoritmo para a metáfora, nem uma metáfora pode ser produzida por meio das instruções precisas de um computador, seja qual for o volume de informação organizada a alimentá-lo.* - Umberto Eco[3]

O termo Inteligência artificial é frequentemente citado na imprensa popular, bem como nos círculos de arte e filosofia, como um talismã alquímico cujo funcionamento raramente é explicado. O paradigma hegemônico até o momento (também crucial para a automação do trabalho) não é baseado no *GOFAI* (a Boa e Velha Inteligência Artificial[4] que nunca conseguiu automatizar a *dedução simbólica*), mas nas redes neurais projetadas por Frank Rosenblatt em 1958 para automatizar a *indução estatística*. Este texto destaca o papel das portas lógicas na arquitetura distribuída das redes neurais, na qual um loop de controle generalizado afeta cada nó de computação para realizar o reconhecimento de padrões. Nesta arquitetura distribuída e adaptativa de portas lógicas, em vez de a lógica ser aplicada à informação de cima para baixo, *a informação se transforma em lógica*, isto é, uma representação do mundo torna-se uma nova função na mesma descrição do mundo. Esta formulação básica é sugerida como uma definição mais precisa de aprendizagem para desafiar a definição idealista de inteligência (artificial). Se o reconhecimento de padrão por indução estatística é a descrição mais precisa do que é popularmente chamado de Inteligência artificial, os efeitos de

---

[2] Frank Rosenblatt. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Buffalo, NY: Cornell Aeronautical Laboratory, 1961. 28. Print.
[3] Umberto Eco. Semiotics and the Philosophy of Language. Bloomington: Indiana University Press, 1986. 127. Print.
[4] *Good Old-fashioned Artificial Intelligence (N.T.)*

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

distorção da indução estatística sobre a percepção coletiva, a inteligência e a governança (sobreajuste, apofenia, viés algorítmico, "*deep dreaming*", etc.) ainda não foram totalmente compreendidos.

De maneira geral, este texto avança na hipótese de que novas máquinas enriquecem e desestabilizam as categorias matemáticas e lógicas que ajudaram a projetá-las. Qualquer máquina é sempre uma máquina de cognição, um produto do intelecto humano e componente indisciplinado das engrenagens da cognição estendida. Graças às máquinas, o intelecto humano atravessa novas paisagens da lógica de uma forma materialista, isto é, sob a influência de artefatos históricos em vez do Idealismo. Como, por exemplo, o motor térmico estimulou a ciência da termodinâmica (ao invés do contrário), Pode-se esperar que as máquinas de computação lancem uma nova luz sobre a filosofia da mente e a própria lógica. Quando Alan Turing teve a ideia de uma máquina de computação universal, ele buscou a mais simples maquinação para calcular todas as funções possíveis. A eficiência do computador universal catalisou em Turing o projeto alquímico para a automação da inteligência humana. No entanto, seria um doce paradoxo ver a máquina de Turing, que nasceu como *Gedankenexperiment,* demonstrar a incompletude da matemática que aspira a descrever exaustivamente um paradigma de inteligência (que é como o teste de Turing é frequentemente compreendido).

Uma unidade de informação é uma unidade lógica de decisão

Em vez de reiterar a GOFAI – isto é, a aplicação de cima para baixo da lógica às informações recuperadas do mundo –, este texto tenta enquadrar a transmutação de *informações externas em lógica interna* na maquinação de redes neurais. Dentro das redes neurais (como também de acordo com a estrutura da cibernética clássica), a informação torna-se controle; ou seja, uma entrada numérica recuperada do mundo transforma-se em uma função de controle do mesmo mundo. Mais filosoficamente, isso significa que uma representação do mundo (informação) torna-se uma nova regra no mesmo mundo (função), ainda que sob um grau

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

considerável de aproximação estatística. *Informação se tornando lógica* é uma formulação muito crua de inteligência que, no entanto, visa enfatizar a abertura para o mundo como um processo contínuo de aprendizagem.

A transformação da informação em funções superiores pode provavelmente ser detectada em diferentes estágios na história das máquinas inteligentes: este texto destaca apenas a definição inicial de informação e de *loops de feedback* antes de analisar sua ramificação em redes neurais. A metamorfose de um loop de informações em formas superiores de conhecimento do mundo era a preocupação da Cibernética de Segunda Ordem na década de 1970, mas já era exemplificada pelas redes neurais de Rosenblatt no final da década de 1950.[5] Para entender como as redes neurais transformam informações em lógica, pode ser útil desconstruirmos, então, a recepção tradicional de ambos os conceitos de informação e de feedback de informações. Normalmente, Claude Shannon é criticado pela redução da informação a uma medida matemática de acordo com o ruído do canal.[6] No mesmo período, de modo mais interessante, Norbert Wiener definiu informação como *decisão*.

> O que há na informação e como é medido? Uma das formas mais simples e unitárias de informação é o registro de uma escolha entre duas alternativas simples igualmente prováveis, estando uma ou a outra fadada a acontecer — uma escolha, por exemplo, entre cara e coroa no lançamento de uma moeda. Chamaremos uma única escolha desse tipo de *decisão*.[7]

Se cada unidade de informação é uma unidade de decisão, uma doutrina atômica do controle é encontrada dentro da informação. Se informação é decisão, qualquer bit de informação é um pequeno pedaço de uma lógica de controle. Bateson acrescentou que "a informação é uma diferença que faz diferença", preparando a cibernética para ordens superiores de

[5] Cf. Francis Heylighen e Cliff Joslyn. “Cybernetics and Second-Order Cybernetics” in Encyclopedia of Physical Science and Technology 19. Ed. R.A. Meyers. New York: Academic Press, 2001. Print.
[6] Claude Shannon. “A Mathematical Theory of Communication.” Bell System Technical Journal 27/3 (1948). Print.
[7] Norbert Wiener. Cybernetics: Or Control and Communication in the Animal and the Machine. Cambridge, MA: MIT Press, 1948. 61. Print. A formulação de Wiener passou a influenciar também a palestra de Jacques Lacan em 1955 sobre cibernética e psicanálise, na qual as portas lógicas [logic gates] são literalmente entendidas como “portas” que abrem ou fecham novos destinos dentro da ordem simbólica. Jacques Lacan. “Psychoanalysis and cybernetics, or on the nature of language.” The Seminar of Jacques Lacan 2. New York: Norton, 1988. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

organização.[8] Na verdade, a Cibernética de Segunda Ordem veio para quebrar o encanto do loop de feedback negativo e a obsessão da cibernética inicial em manter aspectos biológicos, técnicos e sistemas sociais constantemente em equilíbrio. Um loop de feedback negativo é definido como um loop de informação que é gradualmente ajustado para adaptar um sistema ao seu ambiente (regulando sua temperatura, energia, consumo, etc.). Um loop de feedback positivo, ao contrário, é um loop que cresce fora de controle e afasta um sistema do equilíbrio. A Cibernética de Segunda Ordem observou que apenas sistemas distantes do equilíbrio tornam possível a geração de novas estruturas, hábitos e ideias (foi o Prêmio Nobel Ilya Prigogine que mostrou que as formas de auto-organização, no entanto, ocorrem também em estados turbulentos e caóticos).[9] Se, já na formulação básica da cibernética inicial, o loop de feedback pode ser entendido como um modelo de informação que se transforma em lógica, que transforma a própria lógica para inventar novas regras e hábitos, é apenas a Cibernética de Segunda Ordem que parece sugerir que é a excessiva "pressão" do mundo externo que força a lógica maquínica a sofrer mutações.

Frank Rosenblatt e a invenção do Perceptron

Considerando que a evolução da inteligência artificial é feita de múltiplas linhagens, este texto lembra apenas o confronto crucial entre dois colegas da Bronx High School of Science, isto é, Marvin Minsky, fundador do Laboratório de Inteligência Artificial do MIT, e Frank Rosenblatt, o inventor da primeira rede neural operativa, o Perceptron. O confronto entre Minsky e Rosenblatt é muitas vezes descrito de modo simplificado como a disputa entre um paradigma baseado em regras de cima para baixo (IA simbólica) e computação paralela distribuída (conexionismo). Em vez de encarnar um algoritmo totalmente inteligente desde o início, no último modelo uma máquina *aprende* com o ambiente e gradualmente se torna

[8] Gregory Bateson. Steps to an Ecology of Mind. Chicago: University of Chicago Press, 1972. Print.
[9] Gregoire Nicolis and Ilya Prigogine. Self-Organization in Nonequilibrium Systems. New York: Wiley, 1977. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

parcialmente "inteligente". Em termos lógicos, aqui corre a tensão entre a *dedução simbólica* e a *indução estatística*.[10]

Em 1951, Minsky desenvolveu a primeira rede neural artificial SNARC (um solucionador de labirintos), mas então abandonou o projeto, convencido de que as redes neurais exigiriam um poder de computação excessivo.[11] Em 1957, Rosenblatt descreveu a primeira rede neural bem-sucedida em um relatório para o Cornell Aeronautical Laboratory, intitulado *The Perceptron:* A Perceiving and Recognizing Automaton. De modo semelhante a Minsky, Rosenblatt esboçou sua rede neural, dando uma estrutura ascendente e distribuída como a ideia de neurônio de Warren McCulloch e Walter Pitts, inspirada pelos neurônios do olho.[12] A primeira máquina neural, o Mark 1 Perceptron, nasceu de fato como uma máquina de visão.[13]

> Um requisito primário de tal sistema é que ele deve ser capaz de reconhecer padrões complexos de informações que são fenomenalmente semelhantes [...], um processo que corresponde aos fenômenos psicológicos de "associação" e "generalização de estímulos". O sistema deve reconhecer o "mesmo" objeto em diferentes orientações, tamanhos, cores ou transformações, e contra uma variedade de fundos diferentes. Deve ser viável construir um sistema eletrônico ou sistema eletromecânico que aprenderá a reconhecer semelhanças ou identidades entre padrões de informações ópticas, elétricas ou tonais, de uma maneira que pode ser bastante análoga a processos perceptivos de um cérebro biológico. O sistema proposto depende mais de princípios probabilísticos do que determinísticos para sua operação, e ganha sua confiabilidade a partir das propriedades de medidas estatísticas obtidas de grandes populações de elementos.[14]

[10] A Inteligência artificial geral (AGI) é a tentativa de encontrar um ponto entre as abordagens de cima para baixo (simbólicas) e as abordagens de baixo para cima (conexionistas), ou seja, de combinar dedução simbólica e indução estatística. Até o momento, no entanto, apenas o paradigma conexionista das redes neurais passou a ser automatizado com sucesso, lançando dúvidas sobre algumas premissas metafísicas e centralizadoras da AGI.

[11] Cf. Marvin Minsky. Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain Model Problem. Dissertation. Princeton University, 1954. Print.

[12] Warren McCulloch and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity." Bulletin of Mathematical Biophysics 5/4 (1943). Print. E Warren McCulloch and Walter Pitts. "How We Know Universals the Perception of Auditory and Visual Forms." Bulletin of Mathematical Biophysics 9/3 (1947). Print.

[13] Notadamente, o livro de Paul Virilio sobre a visão de máquina foi inspirado também pelo Perceptron (ainda que Virilio não pudesse prever que o Perceptron se tornaria o paradigma hegemônico da inteligência de máquina no início do século XXI). Cf. o capítulo 5 de Paul Virilio, La Machine de vision: essai sur les nouvelles techniques de représentation. Paris: Galilée, 1988. Print. English translation: The Vision Machine. Bloomington: Indiana University Press, 1994. Print.

[14] Frank Rosenblatt. "The Perceptron a Perceiving and Recognizing Automaton." Technical Report 85/460/1 (1957). 1-2. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

Deve ser esclarecido que o Perceptron não era uma máquina para reconhecer formas simples como letras (o reconhecimento óptico de caracteres já existia na época), mas uma máquina que poderia *aprender* como reconhecer formas calculando um único arquivo estatístico em vez de salvar vários em sua memória. Especulando além do reconhecimento de imagens, Rosenblatt profeticamente acrescentou: "Espera-se de dispositivos deste tipo que, em última análise, sejam capazes de formação de conceitos, tradução de linguagem, comparação de inteligência militar e solução de problemas por meio da lógica indutiva."[15]

Em 1961, Rosenblatt publicou *Princípios de Neurodinâmica*: Perceptrons e a teoria dos mecanismos cerebrais, que influenciou a computação neural até hoje (o termo *Multi-Layer Perceptron*, por exemplo, já está presente). O livro parte de descobertas psicológicas e neurológicas sobre neuroplasticidade e as aplica ao projeto de redes neurais. O *Perceptron* era um modelo artificial do cérebro que se destinava a explicar alguns de seus mecanismos, mas sem ser confundido com o próprio cérebro. (Na verdade, as redes neurais foram concebidas pela imitação dos neurônios do olho em vez dos cerebrais, e sem saber como o córtex visual realmente elabora entradas visuais). Rosenblatt enfatizou que as redes neurais artificiais são uma simplificação e um exagero do sistema nervoso, e esta aproximação (que é o reconhecimento dos limites de um pensamento baseado em modelos) deve ser uma diretriz para qualquer filosofia da mente (artificial). Em última análise, Rosenblatt propôs a *neurodinâmica* como uma disciplina contra o *hype* da inteligência artificial.

> O programa perceptron *não* se preocupa prioritariamente com a invenção de dispositivos de "inteligência artificial", mas sim com a investigação das estruturas físicas e os princípios neurodinâmicos que fundamentam a "inteligência natural". Um perceptron é antes de tudo um modelo cerebral, não uma invenção para reconhecimento de padrões. Como um modelo cerebral, sua utilidade é nos permitir determinar as condições físicas para o surgimento de várias propriedades psicológicas. Não é de forma alguma um modelo "completo", e estamos plenamente cientes das simplificações que foram feitas a partir de sistemas biológicos; mas é, pelo menos, um modelo analisável.[16]

[15] Ibid. 30.
[16] Frank Rosenblatt. Op. cit. 1961. vii. Print

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

Em 1969, o livro de Marvin Minsky e Seymour Papert, intitulado *Perceptrons*, atacou o modelo de sistema neural em rede de Rosenblatt ao afirmar erroneamente que um Perceptron (embora de camada única simples) não poderia aprender a função XOR e resolver classificações em dimensões superiores. Este livro recalcitrante teve um impacto devastador, também por causa da morte prematura de Rosenblatt em 1971, e bloqueou fundos para pesquisa de redes neurais por décadas. O que é denominado como o primeiro "inverno da Inteligência artificial" seria melhor descrito como o "inverno das redes neurais", que durou até 1986, quando os dois volumes de *Parallel Distributed Processing* esclareceu que os Perceptrons (multicamadas) podem realmente aprender funções lógicas complexas.[17] Meio século e muitos neurônios depois, contrariando Minsky, Papert e os fundamentalistas da IA simbólica, Perceptrons multicamadas são capazes de reconhecer imagens melhor do que os humanos, e eles constituem o núcleo dos sistemas de aprendizado profundo, como os de tradução automática e os carros autônomos.[18]

Anatomia de uma rede neural

Em termos de uma arqueologia das mídias, a invenção da rede neural pode ser descrita como a composição de quatro formas tecno-lógicas: a escansão (discretização ou digitalização de entradas analógicas), a porta lógica (que pode ser concretizada como potenciômetro, válvula, transistor, etc.), o loop de feedback (a ideia básica da cibernética) e a rede (inspirada aqui pelo arranjo de neurônios e sinapses). No entanto, o propósito de uma rede neural é calcular um constructo estatístico-topológico que é mais complexo do que a disposição de tais formas. A função de uma rede neural é registrar padrões de entrada semelhantes (conjunto de dados de treinamento) como *estados internos* de seus nós. Depois que um estado interno

[17] David Rumelhart and PDP Research Group. Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1-2. Cambridge, MA: MIT Press, 1986. Print..

[18] As redes neurais continuam crescendo em direção a topologias cada vez mais complexas e inauguraram uma verdadeira *ars combinatoria* computacional (veja os diagramas de autoencoder, a máquina de Boltzmann, a rede neural recorrente e de memória de longo prazo, as redes generativas adversárias, etc.). As redes neurais são as máquinas mais articuladas e sofisticadas ao longo da tradição do conhecimento computável, como no antigo dispositivo árabe zairja ou no livro Ars Magna (1305), de Ramon Llull. Cf. David Link. "Scrambling T-R-U-T-H: Rotating Letters as a Material Form of Thought." Variantology 4. On Deep Time Relations of Arts, Sciences and Technologies in the Arabic–Islamic World. Eds. Siegfried Zielinski and Eckhard Fürlus. Cologne: König, 2010. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

foi calculado (ou seja, a rede neural foi "treinada" para o reconhecimento de um padrão específico), este constructo estatístico pode ser instalado em redes neurais com estrutura idêntica e usado para reconhecer padrões em novos dados.

As portas lógicas que normalmente fazem parte de estruturas lineares de computação adquirem novas propriedades na computação paralela das redes neurais. Nesse sentido, Rosenblatt fez provavelmente uma das primeiras descrições da inteligência de máquina como propriedade emergente: "É significativo que os elementos individuais, ou células, de uma rede nervosa nunca demonstraram possuir qualquer funções psicológicas, como 'memória', 'consciência' ou 'inteligência'. Tais propriedades, portanto, presumivelmente residem na organização e funcionamento da rede como um todo, em vez de nas partes elementares."[19] No entanto, as redes neurais não são horizontais, mas redes hierárquicas (em camadas).

A rede neural é composta por três tipos de camadas de neurônios: camada de entrada, camadas ocultas (que podem ser muitas, daí o termo "aprendizado profundo") e camada de saída. Desde o primeiro Perceptron (e revelando a influência do paradigma visual), a camada de entrada é frequentemente chamada de retina, mesmo que não compute dados visuais. Os neurônios da primeira camada estão conectados aos neurônios da próxima, seguindo um fluxo de informações em que uma entrada complexa é codificada para corresponder a uma determinada saída. A estrutura que emerge não é realmente uma rede (ou um rizoma), mas uma *rede arborescente*, que cresce como um *cone hierárquico* no qual as informações são canalizadas e destiladas em formas superiores de abstração.[20]

Cada neurônio da rede é um nó de transmissão, mas também um nó computacional; é porta de informação e porta lógica. Cada nó tem então duas funções: transmitir informações e aplicar lógica. A rede neural "aprende" conforme a saída errada é redirecionada para ajustar

[19] Frank Rosenblatt. Op. cit. 1961. 9.
[20] Cf. Ethem Alpaydın. Introduction to Machine Learning. 2nd ed. Cambridge, MA: MIT Press, 2014. 260. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

o erro de cada nó de computação até que a saída desejada seja alcançada. As redes neurais são muito mais complexas do que sistemas cibernéticos tradicionais, uma vez que instanciam um *loop de feedback generalizado* que afeta uma infinidade de nós de computação. Nesse sentido, a rede neural é a arquitetura de computação mais adaptável projetada para aprendizado de máquina.

O feedback generalizado afeta a *função* de cada nó ou neurônio; isto é, a maneira como um nó calcula (seu "peso"). O feedback que controla o cálculo de cada nó (o que é denominado ajuste de peso, retropropagação do erro, etc.) pode ser uma equação, um algoritmo ou mesmo um operador humano. Em um exemplo específico de rede neural, ao se modificar um limiar de nó, o controle de feedback pode transformar uma porta OR ["ou"] em uma porta AND ["e"], por exemplo — o que significa que o controle de feedback muda a maneira como um nó "pensa".[21] As portas lógicas das redes neurais computam informações a fim de afetar a maneira como elas computarão as informações futuras. Desta forma, *a informação afeta a lógica*. O núcleo de negócios das principais empresas de TI hoje é encontrar a fórmula mais eficaz para feedbacks de controle neural.

Mais especificamente, a rede neural aprende como reconhecer uma imagem gravando as dependências ou relações entre os pixels e compondo estatisticamente uma representação interna. Em uma foto de uma maçã, por exemplo, um pixel vermelho pode estar rodeado por outros pixels vermelhos 80% do tempo, e assim por diante. Desta forma, relações incomuns também podem ser combinadas em recursos gráficos mais complexos (arestas, linhas, curvas, etc.). Assim como uma maçã deve ser reconhecida de ângulos diferentes, uma imagem real nunca é memorizada, apenas suas dependências estatísticas. O gráfico estatístico das dependências é registrado como uma representação interna multidimensional, que é então associada a uma saída legível por humanos (a palavra "maçã"). Este modelo de treinamento é chamado de aprendizado supervisionado, pois um

[21] Este é um caso específico para fins ilustrativos. As funções de ativação também operam de maneiras diferentes.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

ser humano decide se cada saída está correta. Aprendizado não-supervisionado é quando a própria rede neural tem que descobrir o que há de mais comum entre padrões de dependências em um conjunto de dados de treinamento sem seguir uma classificação anterior (dado um conjunto de dados de fotos de gatos, ele extrairá as características de um gato genérico).

Dependências e padrões podem ser rastreados nos mais diversos tipos de dados: conjuntos de dados visuais são os mais intuitivos de entender, mas os mesmos procedimentos são aplicados, por exemplo, aos dados sociais, médicos e econômicos. As técnicas atuais de inteligência artificial são claramente uma forma sofisticada de reconhecimento de padrões em vez de uma forma de inteligência, se a inteligência for entendida como a *descoberta e a invenção de novas regras*. Para ser preciso em termos de lógica, o que as redes neurais calculam é uma forma de *indução estatística*. Evidentemente, essa forma extraordinária de inferência automatizada pode ser um aliado precioso para a criatividade humana e a ciência (e é a melhor aproximação daquilo que é conhecido como a abdução fraca de Peirce), mas não representa *per se* a automação da inteligência enquanto invenção, precisamente enquanto permanecer dentro das categorias "demasiado humanas".[22]

Computação humana, demasiado humana

Peirce disse que "o homem é um signo externo".[23] Se essa intuição encorajou os filósofos a enfatizarem que a mente humana é um projeto artefactual que se estende à tecnologia, a imbricação real da mente humana com as máquinas externas de cognição, no entanto, raramente foi ilustrada *empiricamente*. Isto produziu posturas simplistas em que ideias como Inteligência Artificial Geral e Superinteligência são evocadas como talismãs alquímicos do pós-humanismo, mas com pouca explicação do funcionamento interno e dos postulados

[22] Sobre as tentativas de automatizar a abdução fraca, cf "Automatic Abductive Scientists". In: Lorenzo Magnani. Abductive Cognition. Springer Science & Business Media, 2009. 112. Print.
[23] Charles S. Peirce. "Some Consequences of Four Incapacities" (1868). The Essential Peirce 1 (1867-1893). Eds. Nathan Houser and Christian Kloesel. Bloomington: Indiana University Press, 1992. 54. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

da computação. Um aspecto fascinante da computação neural é, na verdade, a forma como ela amplia as categorias do conhecimento humano, em vez de substituí-las em formas autônomas. Ao contrário da concepção ingênua de autonomia da inteligência artificial, na arquitetura das redes neurais muitos elementos ainda são profundamente afetados pela intervenção humana. Para entender o quanto a computação neural se estende ao "inumano", deve-se discernir o quanto ainda é "demasiado humana". O papel do humano (e também o *locus* de poder) é claramente visível (1) no design do conjunto de dados de treinamento e suas categorias, (2) na técnica de correção de erros e (3) na classificação da saída desejada. Por razões de espaço, apenas o primeiro ponto é discutido aqui.

O design do conjunto de dados de treinamento é o componente mais crítico e vulnerável da arquitetura de redes neurais. A rede neural é treinada para reconhecer padrões em dados anteriores com a esperança de estender essa capacidade em dados futuros. Mas, como já ocorreu várias vezes, se os dados de treinamento mostram um viés racial, de gênero e de classe, as redes neurais refletirão, amplificarão e distorcerão esse viés. Sistemas de reconhecimento facial que foram treinados em bancos de dados de rostos de pessoas brancas falharam miseravelmente em reconhecer os negros como humanos. Este é um problema chamado "sobreajuste": dada a abundância do poder de computação, uma rede neural mostrará a tendência a aprender demais, ou seja, fixar-se em um padrão superespecífico: é, portanto, necessário abandonar alguns de seus resultados para tornar seu ímpeto de reconhecimento mais relaxado. Semelhante ao sobreajuste é o caso de "apofenia", como nas paisagens psicodélicas do Google DeepDream, nas quais as redes neurais "veem" padrões que não existem ou, melhor, geram padrões contra um fundo ruidoso. Sobreajuste e apofenia são exemplos dos limites intrínsecos na computação neural: eles mostram como as redes neurais podem espiralar paranoicamente em torno de padrões em vez de ajudar a revelar novas correlações.

A questão do sobreajuste aponta para uma questão mais fundamental na constituição do conjunto de dados de treinamento: o limite das categorias dentro das quais a rede neural

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

opera. O modo pelo qual um conjunto de dados de treinamento representa uma amostra do mundo, marca, ao mesmo tempo, um universo fechado. Qual é a relação de tal universo fechado de dados com o exterior? Uma rede neural é considerada 'treinada' quando é capaz de generalizar seus resultados para dados desconhecidos com uma margem de erro muito baixa, mas tal generalização é possível devido à homogeneidade entre o conjunto de dados de treinamento e os de teste. Uma rede neural nunca é solicitada para ter um desempenho em categorias que não pertencem à sua "educação". A questão, portanto, é: o quanto uma rede neural (e a IA em geral) é capaz de escapar da ontologia categórica em que opera?[24]

Abdução do desconhecido

Normalmente, uma rede neural calcula a indução estatística de um conjunto homogêneo de dados; isto é, extrapola padrões que são consistentes com a natureza do conjunto de dados (um padrão visual de dados visuais, por exemplo), mas se o conjunto de dados não for homogêneo e contiver recursos multidimensionais (como em um exemplo básico, dados sociais que descrevem idade, sexo, renda, educação, condições de saúde do população, etc.), as redes neurais podem descobrir padrões entre os dados que a cognição humana não tende a correlacionar. Mesmo que as redes neurais mostrem correlações imprevistas para a mente humana, elas operam dentro da grade implícita de postulados e categorias (humanas) que estão no conjunto de dados de treinamento e, nesse sentido, não podem dar o salto necessário para a invenção de categorias radicalmente novas.

A distinção de Charles S. Peirce entre dedução, indução e abdução (hipótese) é a melhor maneira de enquadrar os limites e potencialidades da inteligência de máquina. Peirce notadamente observou que as formas lógicas clássicas de inferência — dedução e indução — nunca inventam novas ideias, mas apenas repetem fatos quantitativos. Apenas a abdução (hipótese) é capaz de introduzir novas visões de mundo e inventar novas regras.

[24] Ontologia é utilizada aqui no sentido das ciências da informação.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

> A única coisa que a indução consegue é determinar o valor de uma quantidade. Ela parte de uma teoria e mede o grau de concordância dessa teoria com o fato. Não pode criar qualquer ideia que seja. Tampouco a dedução. Todas as ideias da ciência vêm sob a forma de abdução. A abdução consiste em estudar fatos e conceber uma teoria para explicá-los.[25]

Especificamente, a distinção de Peirce entre abdução e indução pode iluminar a forma lógica das redes neurais, já que desde sua invenção por Rosenblatt elas foram projetadas para automatizar formas de indução.

> Por indução, concluímos que fatos, semelhantes aos fatos observados, são verdadeiros nos casos não examinados. Por hipótese, concluímos a existência de um fato bastante diferente de tudo o que foi observado, a partir de que, de acordo com as leis conhecidas, algo observado seria necessário resultar. O primeiro, é raciocínio do particular para a lei geral; o último, do efeito à causa. O anterior classifica, este último explica.[26]

A distinção entre indução como classificadora e a abdução como explicativa enquadra muito bem também a natureza dos resultados das redes neurais (e o problema central da inteligência artificial). A complexa indução estatística que é realizada por redes neurais chega perto de uma forma de abdução fraca, onde novas categorias e ideias surgem no horizonte, mas parece que a invenção e a criatividade estão longe de ser totalmente automatizadas. A invenção de novas regras (uma definição aceitável de inteligência) não é apenas uma questão de generalização de uma regra específica (como no caso de indução e abdução fraca), mas de romper planos semióticos que não estavam conectados ou concebíveis de antemão, como nas descobertas científicas ou na criação de metáforas (abdução forte).

Em sua crítica à inteligência artificial, Umberto Eco observou: “Nenhum algoritmo existe para a metáfora, nem uma metáfora pode ser produzida por meio de instruções precisas de um computador, não importa qual o volume de informação organizada a alimentá-lo.”[27] Eco enfatizou que os algoritmos não são capazes de escapar da camisa de força das categorias

[25] Charles S. Peirce. Collected Papers. Cambridge, MA: Belknap, 1965. 5, 145. Print.
[26] Charles S. Peirce. “Deduction, Induction, and Hypothesis” (1878). Op cit. 1992. 194.
[27] Umberto Eco. Op. cit. 127.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

que estão implícita ou explicitamente incorporadas pela "informação organizada" do conjunto de dados. Inventar uma nova metáfora é dar um salto e conectar categorias que nunca foram logicamente relacionadas. Quebrar uma regra linguística é a invenção de uma nova regra, mas apenas quando engloba a criação de uma ordem mais complexa em que a velha regra aparece como um caso simplificado e primitivo. Redes neurais podem computar metáforas *a posteriori*[28], mas não podem automatizar a invenção de novas metáforas *a priori* (sem cair em resultados cômicos como geração de textos aleatórios). A automação da abdução (forte) continua sendo a pedra filosofal da inteligência artificial.

Inteligência Artificial (Quase) Explicável

O debate atual sobre Inteligência artificial ainda está basicamente elaborando os traumas epistêmicos provocados pela ascensão da computação neural. Alega-se que a inteligência de máquina abre novas perspectivas de conhecimento que devem ser reconhecidas como patrimônio pós-humano (ver a noção de Lyotard do inumano), mas há pouca atenção às formas simbólicas de reconhecimento de padrões, inferências estatísticas e abdução fraca que constituem essa mudança pós-humana. Além disso, alega-se que essas novas escalas de computação constituem uma caixa preta que está para além do controle humano (e político), sem que se perceba que a arquitetura dessa caixa preta pode sofrer uma engenharia reversa. As passagens seguintes nos mostram que o humano ainda pode entrar no abismo "inumano" da computação profunda e que a influência humana ainda é reconhecível em uma boa parte dos resultados "inumanos" da computação.

É verdade que camadas e mais camadas de neurônios artificiais complicam tanto a computação que é difícil *olhar para trás* em tal estrutura e descobrir onde e como uma "decisão" específica foi computada. As redes neurais artificiais são consideradas como *caixas pretas* porque têm pouca ou nenhuma capacidade de explicar causalidades, ou quais características são importantes na geração de inferências como a classificação. O

[28] Cf. Word2vec, uma estrutura para o mapeamento de incorporação de palavras em espaço vetorial.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

programador muitas vezes não tem controle sobre quais recursos são extraídos, pois estes são deduzidos pelo sistema neural por conta própria.[29]

O problema é mais uma vez claramente percebido pelos militares. A DARPA (a agência de pesquisa da Defesa dos EUA) está estudando uma solução para o efeito caixa preta no programa *Explainable Artificial Intelligence* (XAI).[30] O cenário a ser abordado é, por exemplo, um tanque de guerra autônomo que toma uma direção incomum, ou a detecção inesperada de armas inimigas em uma paisagem neural. A ideia da XAI é que as redes neurais devem fornecer não apenas uma saída inequívoca, mas também uma justificativa (parte do contexto computacional para essa saída). Se, por exemplo, a figura de um inimigo é reconhecida ("esta imagem é um soldado com uma arma"), o sistema dirá por que pensa assim – isto é, de acordo com quais recursos. Sistemas semelhantes podem ser aplicados também ao monitoramento de e-mails para detectar potenciais terroristas, traidores e agentes duplos. O sistema tentará não apenas detectar anomalias de comportamento em relação a um padrão social normal, mas também dar explicações de que contexto de elementos descreve uma pessoa como suspeita. Como a automação da detecção de anomalias já gerou suas vítimas (veja o caso Skynet, no Paquistão)[31], está claro que a XAI deve supostamente também prevenir novos desastres algorítmicos no contexto de policiamento preditivo.

A *Inteligência artificial explicável* (para ser denominada mais corretamente, *Aprendizado profundo explicável*) adiciona mais um loop de controle no topo da arquitetura das redes neurais, e está preparando uma nova geração de mediadores epistêmicos. Isso já faz parte de um interesse empresarial multibilionário, dado que as empresas de seguro, por exemplo, cobrirão apenas os carros autônomos que forneçam a caixa preta computacional, apresentando não apenas gravações de vídeo e áudio, mas também a justificativa para suas decisões de direção (imagine o caso do primeiro acidente entre dois veículos autônomos).

[29] Isso se aplica tanto ao aprendizado supervisionado quanto ao não supervisionado. Agradeço a Anil Bawa-Cavia pelo esclarecimento.
[30] Cf. www.darpa.mil/program/explainable-artificial-intelligence
[31] Matteo Pasquinelli. "Arcana Mathematica Imperii: The Evolution of Western Computational Norms." Former West. Eds. Maria Hlavajova, et al. Cambridge, MA: MIT Press, 2017. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

Escalas inumanas de computação e a estética da *nova idade das trevas* já encontraram seus representantes legais.

Conclusão

A fim de compreender o impacto histórico da Inteligência artificial, este texto destaca que seu paradigma hegemônico e dominante até o momento não é simbólico (GOFAI), mas conexionista, a saber, as redes neurais que constituem também sistemas de aprendizado profundo. O que a grande mídia chama de inteligência artificial é uma forma folclórica de se referir a redes neurais de reconhecimento de padrões (uma tarefa específica dentro da definição mais ampla de inteligência e, com certeza, uma não exaustiva). Reconhecimentos de padrões são possíveis graças ao cálculo do estado interno de uma rede neural que incorpora a forma lógica de indução estatística. A "inteligência" das redes neurais é, portanto, apenas uma inferência estatística das correlações de um conjunto de dados de treinamento. Os limites intrínsecos da indução estatística se encontram entre o superajuste e a apofenia, cujos efeitos estão emergindo gradualmente na percepção coletiva e na governança. Os limites extrínsecos da indução estatística podem ser ilustrados graças à distinção de Peirce entre indução, dedução e abdução (hipótese). É sugerido que a indução estatística se aproxima de formas de abdução fraca (por exemplo, diagnóstico médico), mas não é capaz de automatizar a abdução forte, como acontece na descoberta de leis científicas ou na invenção de metáforas linguísticas. Isto porque redes neurais não podem escapar do limite das categorias que estão implicitamente incorporadas no conjunto de dados de treinamento. As redes neurais apresentam um grau relativo de autonomia em sua computação: elas são ainda dirigidas por fatores humanos e são componentes de um sistema de poder humano. Com certeza, elas não mostram sinais de "inteligência autônoma" ou consciência. Escalas super-humanas de conhecimento são adquiridas apenas em colaboração com o observador humano, sugerindo que *Inteligência aumentada* seria um termo mais preciso do que *Inteligência artificial*.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304

A inferência estatística via redes neurais permitiu que o capitalismo computacional imitasse e automatizasse tanto o trabalho de baixa como o de alta qualificação.[32] Ninguém esperava que até mesmo um motorista de ônibus pudesse se tornar uma fonte de trabalho cognitivo a ser automatizado por redes neurais em veículos autônomos. A automação de inteligência via inferência estatística é o novo olho que o capital lança sobre o oceano de dados do trabalho global, da logística e dos mercados com novos efeitos de *anormalização* – isto é, de distorção da percepção coletiva e das representações sociais, como ocorre na ampliação algorítmica dos vieses de classe, raça e gênero.[33] A inferência estatística é o novo olho distorcido do Mestre do capital.[34]

Matteo Pasquinelli - Universidade Ca' Foscari
Professor Associado de Filosofia da Ciência no Departamento de Filosofia e Patrimônio Cultural da Universidade Ca' Foscari em Veneza, onde coordena o projeto ERC AIMODELS, com duração prevista para cinco anos. Entre outras publicações, escreveu o livro *The Eye of the Master: A Social History of Artificial Intelligence* (Londres: Verso, 2023).
ORCID: https://orcid.org/0000-0001-9626-3371

O autor deseja agradecer a Anil Bawa-Cavia, Nina Franz e Nikos Patelis pelos seus comentários.

[32] Existem diferentes abordagens para a inteligência da máquina, mas a hegemonia do conexionismo na automação é manifesta. Para uma introdução acessível às diferentes famílias de aprendizado de máquina, cf. Pedro Domingos. The Master Algorithm. New York: Basic Books, 2015. Print.
[33] Matteo Pasquinelli. "Abnormal Encephalization in the Age of Machine Learning." e-flux 75 (September 2016). Web.
[34] "A habilidade especial de cada operador de máquina individual, que agora foi privado de todo significado, desaparece como uma quantidade infinitesimal em face da ciência, das gigantescas forças naturais e da massa do trabalho social incorporado no sistema de máquinas , que, juntamente com essas três forças, constitui o poder do mestre". Karl Marx. Capital 1 (1867). London: Penguin, 1982. 549. Print.

ISSN 2175-8689 – v. 27, n. 1, 2024
DOI: 10.29146/eco-ps.v27i1.28304