

INTELIGÊNCIA ARTIFICIAL: SERIA O SENTIMENTO UMA CHAVE PARA INTERPRETÁ-LA?



ARTIFICIAL INTELLIGENCE: WOULD BE SENTIMENT
A KEY TO INTERPRETING IT?

Giovanni Moura de Holanda¹
gholanda@fitec.org.br



Este trabalho está licenciado sob uma Licença
Creative Commons Atribuição-NãoComercial-
SemDerivações 4.0 Internacional.

Data de Submissão: 22/09/2023
Data de Aprovação: 25/09/2023

RESUMO

Um roteiro para as reflexões propostas no livro “Sentimento da Inteligência Artificial: novas tecnologias, antigos conceitos”.

Palavras-chave: inteligência artificial; interpretabilidade algorítmica; inteligência artificial responsável; modelos explicáveis.

ABSTRACT

A script for the reflections proposed in the book “Sentiment of the Artificial Intelligence: new technologies, old concepts”.

Keywords *artificial intelligence; algorithmic interpretability; responsible artificial intelligence; explainable models.*

1 AFINAL, POR QUE SENTIMENTO?

Ao pensarmos como entender o funcionamento dos algoritmos, que determinam o modo como as máquinas inteligentes tomam decisões autonomamente, surge a indagação natural: qual o caminho, ou quais métodos e recursos poderiam nos revelar esse segredo? Não é nada trivial, mas talvez possa ser um pouco mais simples se nos desarmarmos de pressupostos mais rígidos que impedem de descobrirmos uma fenda capaz de revelar os meandros labirínticos dos algoritmos inteligentes. E dessa passagem –digamos, mais elucidadora–, buscar conhecer como são ponderados os parâmetros que dirigem as conexões neurais da inteligência artificial (IA).

Aqui, a alusão feita por Latour (2016) a um dos mitos fundadores da técnica é consideravelmente oportuna. Dédalo, encarnando esse símbolo, tem em sua etimologia (*dedalion*, em grego) o significado de desvio, que possivelmente estaria na motivação original para a criação do fabuloso labirinto em que transcorre o drama de Teseu e Ariadne. O labirinto, palco do trabalho de Teseu em busca do Minotauro, é percorrido dobra após dobra, em curvas que sucessivamente contornadas podem levar a desvios de rota corrigidos apenas pela orientação de um fio condutor –o fio de Ariadne¹ que conduziria o herói em sua saída do labirinto.

E qual seria então o fio de Ariadne para a interpretação algorítmica? Há um método executável em termos de explicar a passagem de uma camada a outra dos algoritmos complexos? As sinapses maquinais teriam um ponto de visão do seu funcionamento? E a partir dessa visão, qual seria a senha para interpretar o que foi entrevisto?

Talvez possamos empregar o sentimento para simbolizar esforços em torno dessa busca. A

palavra sentimento é polissêmica por excelência, traz em si vários significados. Pode ser entendida como um instinto, uma intuição, uma afinidade ou até mesmo uma convicção. E, o que é mais significativo para o que procuramos abordar no livro “Sentimento da Inteligência Artificial” (HOLANDA; PFEIFFER, 2023), traz a noção de sentido, de um caminho a ser trilhado para interpretar a inteligência das máquinas. Entendemos que esse caminho pode ser facilitado pela perspectiva linguística.

2 SUPERANDO DUALIDADES CONCEITUAIS

Um primeiro passo implica na revisão de conceitos basilares ao universo das ciências que aportam suas especificidades para a materialização da IA. Por exemplo, o que de fato podemos entender como inteligência, e qual o significado de suas adjetivações natural e artificial. Quem seria o sujeito e qual seria o objeto instaurado da inteligência? Essas dualidades clássicas parecem ter seus limites expirados e a separação que carregam precisa ser superada para alcançarmos uma visão mais ampla e compatível com os tempos atuais. O mesmo acontece para as polarizações entre realidade e ilusão, real *versus* virtual, humano e maquínico. Como argumentado no livro,

“O recorte que há muito a ciência e a filosofia tentam definir entre verdadeiro e falso, realidade e ilusão, parece agora mais longe de ser alcançado, a despeito de todos os desenvolvimentos tecnológicos e rupturas de paradigmas intelectuais. As ambiguidades e distorções que podem provocar, gerando por exemplo multirrealidades, convivem com toda uma promessa de abertura a novas formas de conhecimento que as novas

1 Na mitologia grega, Ariadne entregou um novelo a Teseu para que ele pudesse entrar e sair do labirinto criado por Dédalo, na ilha de Creta, de modo a vencer o Minotauro que lá habitava. Esse era o desafio imposto para que Teseu desposasse Ariadne, e o fio do novelo permitiu que ele retornasse do labirinto.

tecnologias poderiam trazer nos terrenos epistemológico, sensorial e artístico.” (HOLANDA; PFEIFFER, 2023, p. 60)

Ademais, a confrontação entre o real e o virtual, aflorada com o uso das tecnologias digitais para produção e consumo de novas formas de conteúdo, tem sido frequentemente exercida a partir do uso equivocado do termo “virtual”. Virtual, em suas acepções possíveis, como bem contextualizadas por Pierre Lévy (2010 [1999]), não seria uma oposição ao real, mas uma complementariedade, uma conotação de potência, de algo que pode vir a ser e, inclusive, tornar-se real. A superação dessa dualidade, ambígua por uma questão semântica, já está na própria natureza da digitalização do mundo, em que as fronteiras entre o que é produzido para fruição no domínio do silício e o universo físico-temporal em que de fato vivemos são turvadas constantemente.

Essa condição contemporânea é retratada com mais recursos e argumentação no livro em questão. A narrativa proposta segue na direção de aprofundar a reflexão sobre a dimensão conceitual, que precisa aproximar-se da realidade semântica exigida pela introdução e uso massivo das novas tecnologias no nosso *modus vivendi*. Uma contextualização conceitual, que vá além da redução de ambiguidades, é igualmente cabível à terminologia que fundamenta os métodos e técnicas para interpretação e explicação algorítmica. É o caso dos termos interpretabilidade e explicabilidade, como destacado no livro:

“Após treinados, interpretar e explicar as decisões dos modelos de IA requer um conhecimento de suas estruturas internas. Métodos e técnicas capazes de interpretar o aprendizado de máquina (IML – *Interpretable Machine Learning*) ou de explicar as decisões de modelos de inteligência artificial (XAI – *explainable Artificial Intelligence*) têm sido

tema de interesse recente de pesquisas, motivadas sobretudo pelas questões éticas, de segurança, privacidade e equanimidade; e por consequências judiciais, envolvendo responsabilidade e rastreabilidade. Há muitos trabalhos na literatura levantando os esforços para definição desses termos e a criação de taxonomias que tentam inter-relacioná-los e agrupá-los junto a outros termos correlatos, e uma das questões que dificultam o estabelecimento de bases comuns de entendimento é o uso indevido dos conceitos de interpretabilidade e explicabilidade como sendo intercambiáveis (ARRIETA et al., 2020).” (HOLANDA; PFEIFFER, 2023, p. 37)

Reforçando tal preocupação em torno da profusão de definições e entendimentos sobre esses termos, inclusive alertando sobre a combinação confusa deles, Rudin *et al.* (2021) são convocados no livro para acrescentar seus pontos de vista e esforços de pesquisa que vêm empreendendo:

“... Um modelo interpretável é restrito, seguindo um conjunto de restrições específicas do domínio que tornam os processos de raciocínio compreensíveis, e destacam 10 conjuntos de desafios técnicos e metodológicos que servem como pontos de partida para um aprendizado de máquina interpretável. Descrevendo sem formalismos, os indicadores que podem refletir o grau de interpretabilidade medem a capacidade de um usuário do modelo de IA, inclusive um não especialista, de entender as relações de causa e efeito, ou de entrada e saída, no funcionamento interno desse modelo.” (HOLANDA; PFEIFFER, 2023, p. 38)

E mais, como colocado na página 60, as antinômias e dualidades clássicas podem ser superadas

a partir de um entendimento mais aprofundado sobre o modo com que os objetos e artefatos vêm à existência, e como eles evoluem em busca de seus *télos*. Daí, há no livro um capítulo dedicado aos modos de existência, sobretudo dos objetos técnicos, que põe em perspectiva um encontro entre as ideias de Simondon, Latour e Yuk Hui, envolvendo, claro, muitos outros autores que aportam visões instigantes sobre o tema e, em especial, apontam uma trilha que busca um diálogo entre a cibernética de Wiener e o tempo Bergsoniano.

Uma vez revisitados e postos em nova página, como esses conceitos podem contribuir para descortinar a complexidade algorítmica? O sentimento, como forma de interpretar, pode fornecer um caminho, traçado pela perspectiva linguística. A linguística vista como vetor de comunicacional e cultural, como mediador entre o eu subjetivo que sente e o político-histórico próprio ao sentido, e como meio para a construção do conhecimento. Linguística que também é chave para interpretar o mundo. Gadamer lembra que Humboldt compreende cada língua como uma determinada acepção de mundo.

3 LINGUÍSTICA E INTERPRETAÇÃO

Mas quem de fato interpreta?

“Motivados pela polissemia criativa e no empenho de perscrutar caminhos de compreensão que ocorre no interior dos algoritmos que investem as máquinas de inteligência, podemos chegar à visão nietzschiana de que quem interpreta são nossos sentimentos.” (HOLANDA; PFEIFFER, 2023, p. 14)

Para Nietzsche, não faz sentido pensar uma vida que não interpreta e ele defende a existência de uma relação indissociável entre interpretações e afetos, e que a diversidade de perspectivas

e interpretações afetivas devem ser utilizadas a favor do conhecimento. Claro, esses postulados não podem ser apresentados de forma tão sintética, e uma contextualização mais apropriada é tecida no livro, trazendo outros pensadores para encorpar a narrativa. Entre eles, Galimberti e António Damásio.

Galimberti, por exemplo, ao assinalar que em tempos do absoluto técnico, acentuamos o “desnível do sentimento”, como consequência do fato de a humanidade tornar-se “incapaz de perceber e mesmo imaginar onde a técnica conduzirá o ser humano e até que ponto o transformará” (página 91). E para refletir sobre esse desnível e alinhá-lo ao tema central do “Sentimento da Inteligência Artificial”, conjecturamos que a “interpretação de aparatos técnicos complexos, com autonomia e capacidade de decisão como vistas nos modelos de IA opacos, requer uma combinação de recursos interpretativos que superem a dualidade razão e emoção” (*ibidem*).

Damásio, com a ideia de que a consciência começa como um sentimento, um sentimento especial de conhecer relacionado às estruturas cerebrais que estão ligadas à consciência, e que essas estruturas funcionam com o vocabulário não verbal dos sentimentos, como tratado em (Damásio, 2015). Naturalmente, essa postulação suscita o questionamento de que as tecnologias modernas poderiam alcançar algum nível de consciência e Damásio apresenta duas respostas. Mas deixemos que a leitora ou o leitor descubra por si quais são elas.

Cumpramos ressaltar que esse tema da consciência possibilitada em algum nível pela inteligência artificial exige ainda muito terreno a ser percorrido, requerendo muitos estudos e debates interdisciplinares para desvendá-lo um pouco mais.

“Em nenhuma época anterior, mesmo em períodos recentes, imaginaríamos que máquinas poderiam aprender,

serem lógicas ou, o que ainda é fantástico, seguiriam princípios éticos e teriam preceitos morais. Por mais que fôssemos otimistas com os desdobramentos da IA, não imaginariamos o quanto a tecnologia poderia mudar o pensamento e a forma como apreendemos a realidade e o mundo. Longe de termos respostas elucidativas sobre os impactos que tudo isso traz, e trará em escala ampliada, podemos pelo menos buscar interpretar essa nova razão e tentar lidar com ela do ponto em que estamos, colocando coisas e eventos em relação.” (HOLANDA; PFEIFFER, 2023, p. 101)

funcionamento algorítmico que hoje decide e influência boa parte de nossas atividades e a forma como estamos no mundo, destaca-se que

“É importante que saibamos fazer as perguntas e interpretar o que acontece no interior de uma inusitada arquitetura de conhecimento. Lembrando que a intenção não é desocultar, uma vez que a transparência é sempre uma ilusão e que há sempre mais a compreender.” (HOLANDA; PFEIFFER, 2023, p. 102)

Diante do cenário atual e das mudanças em curso, somos convidados a revisitar o “Sentimento de Mundo” de Drummond, com o intuito de alçarmos formas mais efetivas de interpretar.

E no caminho para concluir o ensaio, em busca de interpretação e explicação sobre o

REFERÊNCIAS

ARRIETA, A. B., et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, v.58, p. 82–115, 2020.

DAMÁSIO, A. *O mistério da consciência*: Do corpo e das emoções aos

conhecimentos de si. (Motta, L. T., Trad.). [2ª ed.]. Companhia das Letras, 2015 [1999].

HOLANDA, G. M.; PFEIFFER, C. R. C. *Sentimento da Inteligência Artificial* – novas tecnologias, antigos conceitos. 1. ed., Campinas, SP: Pontes Editores, 2023.

LATOUR, B. *Cogitamus: Seis cartas sobre as humanidades científicas*. (Jamil P. Dias, Trad.). 1ª ed. Editora 34, 2016.

LÉVY, P. *Cibercultura*. (Carlos I. da Costa, Trad.). 1ª ed. Editora 34, 2010 [1999].