

GT-8 – Informação e Tecnologia

ISSN 2177-3688

DIÁLOGOS ENTRE A TEORIA MATEMÁTICA DA COMUNICAÇÃO, A TEORIA BOOLEANA E A INTELIGÊNCIA ARTIFICIAL NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO

DIALOGUES BETWEEN THE MATHEMATICAL THEORY OF COMMUNICATION, BOOLEAN THEORY AND ARTIFICIAL INTELLIGENCE IN THE CONTEXT OF INFORMATION SCIENCE

Mariana Silva Mota - Universidade Federal do Ceará (UFC)

Oswaldo de Souza - Universidade Federal do Ceará (UFC)

Modalidade: Trabalho Completo

Resumo: No presente texto discutimos diálogos entre as Teorias Matemáticas da Comunicação (TMC), a Inteligência Artificial (IA), mencionando o paradigma do Generative Pre-trained Transformer (GPT) e a Álgebra Booleana (TAB) no sentido de conhecermos as interações destas com objetos de estudo da Ciência da Informação (CI), particularmente na representação e recuperação da informação. Abordamos as propostas de Shannon para a TMC e a proposta de George Boole para a TAB. Apresentamos dados relativos a uma pesquisa bibliográfica, na qual foram selecionados textos de Revistas relacionadas a CI e qualificadas como Qualis A1 pela CAPES. Na análise de cada texto, visando estabelecer correlações e diálogos entre as teorias consideradas no estudo, foram evidenciados aspectos que aludem às teorias e demonstram a complementaridade entre elas no contexto da CI. Contextualizamos a discussão na historicidade das teorias até o momento atual. Abordamos e definimos questões relativas à informação moderna, posicionando-a como um fator com potencial de influência social, ao passo que evidenciamos problemas nos conjuntos de dados para a construção de sistemas de IA. Apontamos discussões sobre a informação moderna, dentre ela as *fake-news*, verdade, meia-verdade, mentira e desinformação, propondo abordagens para mitigar os impactos negativos desses atributos por meio de discussão sobre uma estruturação para os textos, através da qual seria possível atribuir um selo de autenticidade e assim permitir inferir as origens e reputação do conteúdo de um objeto digital.

Palavras-chave: chatGPT; desinformação; teoria matemática da comunicação; Inteligência Artificial; estruturação de textos.

Abstract: In this text we discuss dialogues between Mathematical Theories of Communication (MTC), Artificial Intelligence (AI) and Boolean Algebra (TBA) to know their interactions with objects of study of Information Science (IS), particularly in the representation and information retrieval. We discussed Shannon's proposals for MTC and Boole's proposal for TBA. We present data related to bibliographical research, in which texts from journals related to IS qualified as Qualis A1 by CAPES were selected. In the analysis of texts, aiming to establish correlations and dialogues between the theories considered in the study, aspects that allude to the theories and demonstrate the complementarity between them in the context of IS were highlighted. We contextualize the discussion in the historicity of theories up to the present moment. We address and define issues related to modern information, positioning it as a factor with potential for social influence, while highlighting problems in datasets for building AI systems. We point out discussions about modern information, including fake-news, and suggest attributes such as truth, half-truth, lie and disinformation and propose approaches to mitigate the negative impacts of these attributes through the proposal of a structure for the texts, through which

it would be possible to attribute a sign of authenticity and thus allow inferring the origins and reputation of the content of a digital object.

Keywords: chatGPT; disinformation; mathematical theory of communication; artificial intelligence; text structure.

1 INTRODUÇÃO

Há um comportamento humano de presumir que tudo que é moderno é uma visão norteada somente para frente, no futuro e no desconhecido, esquecendo-se que o passado tem influência no futuro tal qual o presente. O período de guerra no início da década de 40 foi marcado pela necessidade de comunicar a informação com velocidade, cientistas desenvolveram pesquisas e teorias matemáticas com o objetivo de buscar respostas aos problemas que enfrentavam, dentre os quais, podemos citar o que concerne ao processamento de informações. Resultados dos esforços realizados naquele período tem repercussão direta em tecnologias que estão presente atualmente, a exemplo, podemos citar as teorias matemáticas relacionadas ao processamento de similaridade entre informações e nos processos de comunicação, na aprendizagem de máquina e avanços em teorias matemáticas, como se viu com a lógica booleana, que até hoje é fundamental na construção de consultas que são formuladas para serem processadas e respondidas por softwares de gerenciamento de banco de dados, através do que normalmente são conhecidas pelo termo query da Ciência da Computação (CC). Também vimos o surgimento de teorias matemáticas seminais na área da inteligência artificial (IA), como a proposta do neurônio artificial de McCulloch e Pitts (MCCOLLOCH; PITTS, 1943, p. 115-133), trabalho bioinspirado no que se compreendia à época ser o funcionamento de um neurônio humano, o qual impactou fortemente em questões e possibilidades de processamento de dados. Portanto, no sentido de conhecer o impacto contemporâneo dessas teorias matemáticas, este trabalho concentra-se na tentativa de elucidar as relações e os diálogos entre a Teoria Matemática da Álgebra Booleana (TAB), a Teoria Matemática da Comunicação (TMC) e parte do arcabouço Teórico da Inteligência Artificial (IA), quando aplicadas ao campo da Ciência da Informação (CI) inclusive na Recuperação da Informação (RI).

Como ponto focal para as discussões iniciais, abordam-se a TAB de George Boole (BOOLE, 1854) e a relação desta com os trabalhos de Claude Shannon, na TMC (SHANNON, 1948) relacionados ao atual campo da Teoria da Informação e a paradigmas recentes de

sistemas baseados em inteligência artificial, particularmente no modelo teórico do Generative Pre-trained Transformer (GPT) em sua versão 3 (BROWN et al, 2020).

2 MÉTODOS E MATERIAIS DA PESQUISA

Caracterizando-se como uma pesquisa de natureza qualitativa do tipo bibliográfica, os procedimentos foram iniciados com a leitura dos trabalhos seminais da TAB de George Boole (BOOLE, 1854) e no trabalho de Claude Shannon (SHANNON, 1948) sobre a TMC e no trabalho de definição do GPT em sua versão 3 (BROWN et al, 2020), na qual delineou-se a interseção e complementaridade entre esses trabalhos e uma vez consolidados, foram organizados no Quadro 1 na forma de postulados destas teorias. A interseção e complementaridade foram compreendidas com o apoio da análise do assunto (CESARINE; PINTO, 1980) na qual os textos basilares do presente estudo foram avaliados para determinar os conceitos presentes. Tais conceitos foram identificados e mapeados como os objetivos centrais dos textos e na sequência foram identificados os objetos de interesse que se relacionam com os objetivos. Esse mapeamento inicial foi organizado no Quadro 1, ao qual também foram incorporados os resultados teóricos que se produzem pelo alcance dos objetivos centrais dos textos em relação aos respectivos objetos de interesse. Os dados da coluna “Postulado” do Quadro 1 resumem os fundamentos propostos por cada uma das teorias. Posteriormente os postulados e demais dados do quadro apoiaram a definição de quais expressões de busca seriam apropriadas para uma pesquisa bibliográfica dentro do contexto do presente trabalho, as quais são introduzidas adiante, logo após um detalhamento dos Postulados presentes no quadro.

Quadro 1 - Postulados das Teorias

Teoria	Objetivo Central	Interseção de Objetos de Interesse	Resultados que produz	Postulado
Álgebra Booleana	Realizar a soma lógica de dados	Dados numéricos primários ou não numéricos através de abstrações	Decisão quando um ou mais parcelas da soma são verdadeiros ou diferente de zero	Operador disjuntivo OR (+) de soma lógica $0+0=0$ $0+1=1$ $1+0=1$ $1+1=1$
Álgebra Booleana	Realizar complementação lógica de dados	Dados numéricos primários ou não numéricos através de abstrações	Obter o complemento lógico de um determinado valor	Operador negação NOT ($\sim$) de soma lógica $\sim 0 = 1$ $\sim 1 = 0$
Álgebra Booleana	Realizar o produto lógica de dados	Dados numéricos primários ou não numéricos através de abstrações	Obter o complemento lógico de um determinado valor	Operador conjuntivo AND ($.$) de soma lógica $0.0=0$ $0.1=0$

XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB
Aracaju-SE – 06 a 10 de novembro de 2023

				1.0=0 1.1=0
Da Comunicação	Fonte de informação - Parte inicial do processo de comunicação visando a transmissão de dados	Dados numéricos primários ou não numéricos através de abstrações	O compartilhamento de um conjunto de dados	Transmissor $T(d)=\{datum,C\}$ C é o canal de Transmissão
Da Comunicação	Receptor de informação - Parte final do processo de comunicação visando a transmissão de dados	Dados numéricos primários ou não numéricos através de abstrações	Receber o compartilhamento de um conjunto de dados	Receptor $d=R(T(d))$
Da Comunicação	Prover um meio para a efetiva comunicação visando a transmissão de dados	Dados numéricos primários ou não numéricos através de abstrações	Um meio de comunicação para o efetivo envio e recepção	Canal $C = (F,r)$ F é o meio físico ou sem fio de transmissão r = meios para controlar fontes de ruído
GPT	Conjunto de dados de treinamento (Data Set)	Dados numéricos primários ou não numéricos através de abstrações	Fonte de dados para o estabelecimento de relações estatísticas sobre um vocabulário e práticas humanas sobre o texto escrito	$DS = \{datum\}$
GPT	Modelo de Linguagem (Language Model)	Dados numéricos primários ou não numéricos através de abstrações	Dados de uma linguagem representados através de equivalências de frases e dados probabilísticos de ocorrência de frases	$ML = Tr(DS)$
GPT	Conjunto de dados de teste ou ativação (Tests Data Set)	Dados numéricos primários ou não numéricos através de abstrações	Fonte de dados para testar a eficiência do aprendizado ou dados fornecidos para uso real do sistema	$TDS = \{datum\}$

Fonte: Dados da Pesquisa

Quanto ao Quadro 1 as operações booleanas OR, AND e NOT são fundamentais para quaisquer sistemas que visem a recuperação de informações ou de dados, já as operações de transmissão, recepção e suporte a canal de transmissão estão presentes mesmo em complexos sistemas distribuídos pela Internet. No quadro $T(d)=\{datum,C\}$ refere-se à operação de transmissão de um determinado conjunto de dados, expresso por datum, enquanto C refere-se ao necessário canal de transmissão, o qual pode ser provido por meio física, tais como cabos de redes de computadores ou cabos ópticos, ou ainda por recursos de comunicação sem fio. Na notação $C = (F,r)$, F representa o meio físico da transmissão e r representa algum mecanismo para controle de transmissão e correção de erros. A expressão $d=R(T(d))$ refere-se à operação de recepção, na qual $R(T(d))$ representa o processo efetivo da

recepção da transmissão realizada em $T(d)$. Na tabela a entrada relativa a GPT, com a operação $DS=\{datum\}$, define o necessário conjunto de dados que compõem o dataset para o treinamento da aplicação baseada na proposta de inteligência artificial do GPT, enquanto $ML=Tr(DS)$ define a abstração relativa ao modelo de representação da linguagem a ser “compreendida” e praticada pela aplicação. Por fim, a notação $TDS=\{datum\}$ representa o conjunto de dados de teste, os quais podem ser utilizados para verificação dos níveis de resposta correta da aplicação de inteligência artificial. Também representa as entradas de dados fornecidas diretamente pelos usuários através do teclado, por exemplo, quando se interage com uma aplicação como o chatGPT.

A partir dos achados da análise dos textos basilares do estudo e dos elementos retratados no Quadro 1, que apontaram pistas quanto a natureza da interseção e complementaridade das teorias analisadas, foram eleitas as seguintes expressões de busca: “Claude Shannon”, “Teoria da Comunicação”, “Inteligência Artificial”, “Silêncio + Revocação + Ruído”, “Booleano”, “chatGPT”. As quais foram aplicadas em revistas científicas que publicam textos em português e que sejam classificadas com A1 pelo Qualis da CAPES no Quadriênio 2013-2016 na área de Comunicação e Informação, as quais totalizaram 55 revistas, no recorte temporal limitado a março de 2023. Do conjunto documental recuperado das 55 revistas, foram descartados todos os documentos que não se enquadravam no escopo da pesquisa do presente trabalho. Foram mantidos todos os documentos que expressavam relação direta ou indireta com a interseção e complementaridade percebidas neste trabalho em relação a TAB, TMC e ainda relacionados ao GPT ou paradigmas semelhantes na inteligência artificial, inclusive se a relação fosse desdobrada em aplicações atuais dos conceitos aludidos pelas expressões de busca usadas.

O Quadro 2 apresenta todos os documentos que foram mantidos na pesquisa, totalizando 9 documentos. Cada um dos 9 documentos foi analisado no sentido de apontar quais postulados constantes no Quadro 1 puderam ser percebidos em seus respectivos conteúdos e desta forma, compreender as possíveis relações entre a lógica Booleana e a Teoria Matemática da Comunicação e a Inteligência Artificial, quando aplicadas ao campo da Ciência da Informação (CI), o que resultou no que se vê na coluna “Principais teorias abordadas / mencionadas” do Quadro 2.

Quadro 2 - Documentos selecionados para o estudo

Título	Revista	Áreas/Sub-área do	Principais teorias	Ano
--------	---------	-------------------	--------------------	-----

XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB
Aracaju-SE – 06 a 10 de novembro de 2023

		conhecimento	abordadas / mencionadas	
Investigaciones en inteligencia artificial en España	Transinformação	CI	Inteligência Artificial	1989
Utilização de ontologias na Recuperação de Informação: principais abordagens, problemas e propostas de melhorias	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação	Álgebra Booleana Inteligência Artificial Da Comunicação	2021
Multi e interdisciplinaridade entre a Ciência da Informação e a Ciência da Computação no âmbito da Web Semântica.	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação / Representação da Informação	Álgebra Booleana Inteligência Artificial Da Comunicação	2021
Uso da IoT, Big Data e Inteligência Artificial nas capacidades dinâmicas: um estudo comparativo entre cidades do Brasil e de Portugal	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação	Álgebra Booleana Inteligência Artificial Da Comunicação	2019
Identificação e representação de conceitos por meio de termos.	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação / Repr. da Informação	Da Comunicação Álgebra Booleana	2021
A terminologia em Sistemas de Recuperação da Informação baseada na WORDNET.PT.	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação / Repr. da Informação	Álgebra Booleana Inteligência Artificial Da Comunicação	2020
Folksonomia e os conceitos de ordem, desordem e organização à luz de Edgar Morin	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação / Repr. da Informação	Álgebra Booleana Da Comunicação	2020
Linguagem natural ou linguagem controlada? Influência da palavra-chave na representação para indexação e recuperação de informações	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação / Repr. da Informação	Álgebra Booleana Inteligência Artificial Da Comunicação	2020
Estudo das similaridades entre a teoria matemática da comunicação e o ciclo documentário	Informação & Sociedade: Estudos	CC / CI / Rec. da Informação / Repr. da Informação	Álgebra Booleana Da Comunicação	2014

Fonte: Dados da Pesquisa

3 DISCUSSÕES PRELIMINARES

Como se vê no Quadro 1 todas as teorias consideradas no presente estudo possuem relevante nível de interseção e complementaridade. Pode-se ver que todas possuem os mesmos objetos de interesse. É importante ressaltar que imagens, sons, textos e demais elementos iconográficos são representados como dados numéricos primários em qualquer computador moderno. Elementos que excedem a complexidade de simples dados numéricos são mantidos coesos através de Tipos Abstratos de Dados (TAD), tais como: cadeias de caracteres, valores lógicos, datas, horas, etc. Os TAD são Modelos de Representação de Dados (MR) mais elaborados que são suportados por linguagens de programação (LISKOV, 1974) e desta forma, mesmo que de fato os computadores não processem textos, através desse mecanismo essa possibilidade é incorporada. Essa compreensão deve ser estendida para todos os tipos de documentos não textuais, por mais complexos que o sejam. De fato,

esse é o mecanismo que explica a possibilidade de existirem softwares complexos como os baseados em inteligência artificial. Embora à época em que as teorias Booleana de George Boole e a Teoria da Comunicação de Claude Shannon foram desenvolvidas, os recursos computacionais disponíveis fossem incomparavelmente menores que os fornecidos por computadores atuais, suas respectivas propostas continuam em uso, pois os objetos de interesse das teorias continuam em uso e suas especificações técnicas não foram alteradas. O mecanismo dos TAD permitiu um isolamento da especificação dos dados primários, dando liberdade para que linguagens de programação que surgissem, incorporassem novos tipos de dados e novas facilidades para o desenvolvimento de sistemas. É justamente através desses novos tipos de dados incorporados que sistemas baseados em inteligência artificial puderam evoluir. Da proposta original do neurônio artificial de McCulloch e Pitts (MCCOLLOCH; PITTS, 1943, p. 115-133) até os sistemas de comportamentos sofisticados como os do chatGPT (BROWN et al, 2020) na qual é possível uma interação que lembra uma conversa com uma pessoa, todos dependem de complexas relações entre grupos de dados. No Quadro 1, a análise do que se vê na coluna “Principais teorias abordadas / mencionadas”, nos revela que, no escopo e recorte amostral desta pesquisa, apenas um texto, cerca de 11%, foi publicado na década de 1980, dois na década de 2010, cerca de 22% e seis textos, cerca de 67%, na década de 2020, percebe-se que o aumento de publicação de textos que se relacionam a temática deste estudo está indicando um crescimento de interesse e discussões sobre as questões pertinentes ao tema. Percebe-se também que entre os textos selecionados para o estudo, apenas um estudo abordou exclusivamente as teorias relacionadas à Inteligência Artificial. Três textos abordam duas teorias simultaneamente, TAB e TCM, e cinco textos abordam simultaneamente TAB, TCM e IA. Nesses cinco textos os respectivos autores transitam naturalmente entre as ideias, termos e conceitos dessas teorias, de maneira que se percebe que elas estão integradas aos trabalhos da CI. A interdisciplinaridade entre a CI e CC pode ser percebida, generalizando, nos contextos da Recuperação da Informação e ou da Representação da Informação. Essa interdisciplinaridade justifica um aprofundamento no estudo das respectivas teorias, no sentido de evidenciar as relações e os diálogos que possam ser percebidos no processo informacional.

4 A TEORIA MATEMÁTICA DA COMUNICAÇÃO NO PROCESSO INFORMACIONAL

O processo informacional e a comunicação na década de 2020 tem sido caracterizado por discussões sobre a desinformação - *fake-news*, (ANDREZA, FARIAS, VERAS, 2022, p. 1), sobre políticas para a manutenção da reputação de instrumentos de comunicação científica e práticas para a avaliação de textos científicos (MAIA, FARIAS, 2023, p. 122916). Vive-se um momento no qual há um excesso de informações, o qual refere-se a um volume que inviabiliza o seu consumo por parte dos potenciais interessados. Segundo dados da Sociedade Brasileira para o Progresso da Ciência (SBPC), em 2020 já se podia observar um volume 32% maior do que o volume anual publicado em 2015 e acumulados entre 2015 e 2020 há mais de 11 milhões de artigos indexados apenas na Web of Science (WoS) (SBPC, 2021). Esse, contudo, não era o cenário que se podia observar nas décadas de 1940 e 1950. Tratava-se de um cenário no qual o desafio não era vencer o volume de informação disponível, mas sim comunicá-la de maneira eficiente e segura. A preservação da integridade da informação e a velocidade de transmissão foram fatores que levaram vários cientistas a iniciarem pesquisas em torno do tema e dentre eles, como já mencionado anteriormente, o matemático Claude Elwood Shannon que, atuante na área da lógica e jogos mentais, pôde criar a TMC (SHANNON, 1948). Para Sá (SÁ, 2018, p.48-70) a TMC também pode ser chamada de Teoria da Informação, é um princípio que confere à informação a proporção quantificável e física dos emissores e receptores da informação independente do conteúdo a ser transmitido. Nesta teoria a informação adquire uma nova faceta, na qual a informação é um objeto que oferece liberdade de escolha do receptor sobre a seleção da mensagem e o que poderá ser dito como verdade advindo dessa informação. Os aspectos semânticos não são considerados, a informação adquire um formato fragmentado em dados para auxiliar a maior velocidade de transmissão, pois de acordo com Pelegrini (PELEGRINI, 2009, p. 11-23) um dos fatores cernes da TMC é que esta não está voltada ao “o quê” está sendo transmitido, mas “o quanto” está sendo transmitido através de bits, que é a medida matemática para representação de dados na TMC. O entendimento de Sá não é partilhado por Capurro e Hjørland que entendem que:

Comunicação é a união de oferta de significados, informação e compreensão. De acordo com esta teoria, nenhuma transmissão de informação ocorre entre o emissor e o receptor. Esta metáfora de coisas-orientadas, implica na idéia de que existe algo que o emissor possui e perde ao transmiti-la. (CAPURRO; HJORLAND, pág. 148, 2007).

No entendimento de Capurro e Hjørland a informação não é, de fato, fragmentada em dados quantificados e sequer transmitida, sendo um processo de compreensão dos signos pelo cognoscente, entendimento reforçado pelos autores ao afirmarem que “Esta metáfora de coisas-orientadas, implica na ideia de que existe algo que o emissor possui e perde ao transmiti-la” (CAPURRO; HJORLAND, pág. 148, 2007). Pois ao comunicarem, tanto emissor quanto receptor passam a possuir o que foi transmitido. Shannon não nos fala sobre significados, contextos ou compreensão, mas situa o processo da comunicação no espaço da entropia, sendo esse espaço o principal fator a tornar as comunicações complexas. Justamente uma das contribuições de Shannon é parcialmente afastar-se da entropia, minimizando-a e aproximando-se de estados bastante simples para representar a fragmentação da informação em dados e o faz a medida em que passa a representar todos dados a partir da variabilidade de possibilidades dentro de um sistema representado por bits, representando tudo a partir de 0 ou 1. Todavia, ainda nesta atenuação da entropia, quanto maior a quantidade e variedade, maior será a entropia e por conseguinte a complexidade. Trabalhar com a entropia acompanha vantagens e desvantagens, um exemplo disso é discutido por Pelegrini que nos diz:

Tal aspecto da entropia revela uma das grandes vantagens em se trabalhar com bits. O registro ou transmissão destes se torna mais resistente aos efeitos da distorção ou do ruído em função de sua entropia ser menor. (PELEGRINI, 2009, p. 11-12).

Compreendemos então que Shannon (SHANNON, 1948) se ocupa de instrumentalizar a comunicação, ele não se ocupa de estabelecer políticas sobre a compreensão daquilo que se transmite, sendo esse aspecto, o da coisa comunicada e compreendida pelo cognoscente, o foco no qual a fala de Capurro e Hjørland (CAPURRO; HJORLAND, 2007, pág. 148) encontra um sentido contemporâneo. Enquanto Shannon nos fala sobre a fragmentação do documento e sua posterior reconstrução, Capurro e Hjørland nos falam sobre o conteúdo semântico do documento. Não são visões opostas, mas complementares, situam-se em momentos diferentes entre os diálogos do emissor e do receptor nos processos informacionais contemporâneos.

5 A TEORIA BOOLEANA NO PROCESSO INFORMACIONAL

Amplamente conhecidos em sistemas de unidades de informação, assim como na pesquisa básica do Google, os operadores booleanos são chaves lógicas que ajudam na busca da informação de maneira mais assertiva, que viabilize uma melhor precisão no

processo de recuperação da informação. Este é apenas um exemplo de diálogo contemporâneo da TAB nos processos informacionais. De fato, Shannon desenvolveu a ideia de que era possível usar a TAB para simular o comportamento de bobinas elétricas de um sistema de comutação e assim conseguiu sintetizar circuitos de comutação baseados nas portas lógicas AND, OR e NOT, as mesmas que podem ser vistas no Quadro 1 deste texto. Embora não se perceba em um primeiro olhar, tanto a TMC quanto a TAB estão presentes em todo processo informacional contemporâneo, presentes em toda informação moderna. Todavia, para discutirmos a informação moderna precisamos antes, discutir o que venha a ser a modernidade. Para esta tarefa emprestamos inicialmente o entendimento de Giddens (GIDDENS, 1991) que nos diz que modernidade são os modos de vida que nos desvencilham das ordens sociais tradicionais. Aquilo que foi vivenciado há mais de vinte anos, no início dos anos 2000, ainda pode ser moderno. Portanto, seguindo este mesmo pensamento, pode-se inferir que o moderno pode arrastar-se por décadas até atingir um período do tempo em que as coisas sofram transformações, quer sejam de costumes, sociais ou até mesmo de novidades na tecnologia, que superem aquelas que até outrora se consideravam modernas. Assim, é correto afirmar que o moderno não pode ser visto em apenas uma amostra, um único dia ou em um único pensamento, é preciso olhar um período, uma faixa maior das coisas, para então poder se concluir que algo é novo e esse novo, agora é o moderno, tornando todas as outras coisas antigas. Um exemplo surpreendente, até contraditório, do uso cada vez mais evidente de ferramentas provenientes de métodos computacionais da área de inteligência artificial é a presença de chatbots e robôs que interagem em redes sociais, simulando a presença humana e influenciando interesses comerciais ou políticos (TSE, 2018), entre outros. A contradição a que nos referimos reside no fato de que o campo da inteligência artificial, relacionado ao aprendizado de máquina (LUND, BURGESS, 1996, p.203-208) aplicado a textos (ATSERIAS, et al, 1997), é tecnologia que já se via aplicada na década de 1990. Neste caso, tais ferramentas são modernas? Parece que não. Então, em se tratando de processos informacionais, o que é novo é a maneira como tais ferramentas estão sendo utilizadas e combinadas? O moderno, então, é o uso que se dá a ferramentas antigas? Parece que também não. Poderíamos então ir um tanto além e cogitar a possibilidade de que o moderno não esteja na ferramenta, mas naquilo que ela processa atualmente? Acreditamos que sim e neste caso o moderno é a informação, pois é ela que está a ser processada por essas ferramentas antigas.

A informação moderna, entretanto, apresenta-se com certas peculiaridades: ela é fugaz, ela é influenciável e manipulável, ela é disseminada em múltiplos suportes e em múltiplas formas. Ela transita do textual ao hipertexto e a hipermídia. Está presente no jornal, na televisão, na rádio e em todo o ecossistema informacional (DE SOUZA, FERNANDES, 2020). Então diferente do que se via à época de Shannon e Boole, o que se comunica na atualidade não são apenas dados, são verdades, meias-verdades e mentiras. Sendo os dois últimos, os ruídos e interferências previstas no modelo de Shannon, todavia, com a diferença de que Shannon estabeleceu mecanismos para minimizar ou detectar os ruídos prejudiciais à comunicação no contexto de dados. A comunicação contemporânea, no entanto, concordando e indo além de Capurro e Hjørland (CAPURRO; HJORLAND, 2007, pág. 148) comunica signos, significados e contextos, sendo essa comunicação afetada pelos ruídos provenientes da desinformação, da meia-verdade e da mentira. Sendo necessário aportar novos postulados que definam instrumentos para a mitigação destes novos tipos de ruídos. Tais postulados não se originam na TAB ou na TMC, mas nas possibilidades da IA.

6 A IA COMO FONTE E FILTRO DE RUÍDOS NO PROCESSO INFORMACIONAL

A Inteligência Artificial tem ocupado lugar de destaque em atividades complexas (BRUNIALTI et al., 2015) e também no processo informacional relacionado a pesquisas de temas específicos e na criação de textos e conteúdos assistidos por software, a exemplo, cita-se o chatGPT (GPT) que em dois meses atingiu 100 milhões de usuários ativos (FORBES, 2023). Soluções baseadas em IA em geral necessitam de algum conjunto de exemplos para que possam realizar as tarefas de aprendizado. Esses conjuntos de exemplos, em geral, são produzidos pelas pessoas e obviamente, carregam as suas opiniões, tendência e preconceitos. Essa influência é geralmente denominada pelos termos bias ou viés, cuja existência tem causado problemas reais, como se vê no estudo (BOULAMWINI; GEBRU, 2018) intitulado *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender* que discute disparidades relevantes no reconhecimento facial com taxas de erros maiores para mulheres e para pessoas com pele mais escuras. Um conjunto de dados contaminado com vieses poderá afetar negativamente a imparcialidade e a igualdade de representação em sistemas autônomos. Se por um lado os sistemas baseados em IA podem aprender um viés negativo, também podem ser influenciados por um viés positivo. Isso é uma boa notícia, pois podemos usar a IA para identificar e rotular informação falsa, meias-verdades, mentiras e

desinformação. Se por um lado a IA pode ser usada na inserção de ruídos provenientes do bias dos dados, também pode ser usada como um filtro a combater a desinformação, da meia-verdade e da mentira. Se hoje temos ISSN, ISBN, e DOI para identificar os objetos digitais, podemos pensar em implementar uma solução baseada, por exemplo, no paradigma de inteligência artificial GPT, que atribua um número e um selo de verificação de conteúdo, atestando a originalidade do conteúdo do objeto digital, construindo uma estruturação dos textos, permitindo assim inferir suas origens e reputação. Considerando-se o volume de publicações disponíveis no ecossistema informacional, a IA pode ser utilizada como ferramenta em prol da redução do tempo necessário para que um potencial leitor possa decidir qual texto ler, ou até mesmo no esforço relacionado ao processo de seleção do conjunto bibliográfico a ser considerado em uma pesquisa. Segundo De Souza e Fernandes:

Há, portanto, um aumento do tempo necessário para que um potencial usuário de uma informação a encontre; (...) Acreditamos, contudo, que é possível alterar essa relação com o tempo através do uso de tecnologia que minimize o volume de informações a serem apreendidas, (...) Se um determinado documento puder ser transformado em outro tipo de documento, todavia preservando-se o valor semântico, (...), criando versões de denso valor semântico que requereriam menor tempo para seu consumo. (DE SOUZA; FERNANDES, 2020, p. 146).

Um trabalho direcionado à diminuição do tempo e esforço necessários ao contato com o material bibliográfico e seu potencial uso foi apresentado em (DE SOUZA et al, 2017) no qual é descrita uma solução baseada em processamento de linguagem natural (PLN) e inteligência computacional (IC) que produz automaticamente resumos de textos, a solução proposta foi testada e os resultados positivos dos testes podem ser vistos em (TABOSA et al, 2020). Todavia, ainda que existam estudos direcionados a diminuir o tempo e o esforço necessário no processo informacional, a ausência de uma estruturação dos textos gera um ambiente propício ao surgimento dos fenômenos das *fake-news* e discussões em busca de métodos para avaliação de um determinado texto ou informação quanto a suas características de verdade, meia-verdade, mentira e desinformação. Quando consideramos o ecossistema informacional (DE SOUZA; FERNANDES, 2020, p. 145) percebemos que o problema decorrente da ausência de uma estruturação dos textos, quanto a identificação de seus autores, origens, propriedade e principalmente seus antecedentes e ou descendentes semânticos, impacta negativamente na formação de opiniões e orientação correta para a sociedade quanto a assuntos diversos. Esse impacto negativo abrange todas as áreas da ação humana, tais como a educação, a saúde, a convivência social e a formação da cultura. Utilizar a IA como filtro de ruídos no processo informacional pode contribuir na minimização dos

problemas relativos as *fake-news*, tornando mais fácil identificar, analisar, quantificar e atribuir selos qualitativos aos textos e informações, por exemplo, classificando-os como: verdade, meia-verdade, mentira e desinformação. Para essa atribuição de classificação torna-se necessário haver algum esforço de estruturação, produzindo alguma noção de cadeia-produtiva dos textos, na qual fossem evidenciados, como dito anteriormente, os autores, origens, propriedade e conteúdo semântico que influenciaram ou que derivaram do texto. Obviamente, considerando-se o volume de textos e informações presentes no ecossistema informacional, não há como tal estruturação ser feita por processo manual.

Embora considerando-se que a IA, ao dialogar com a TAB e a TMC, beneficia-se das operações de conjunção, negação e disjunção, constantes no Quadro 1, para elaborar as complexas relações entre textos, determinando suas equivalências e divergências, ainda assim há uma incompletude perceptível (GÖDEL, 1931). Essa incompletude é materializada justamente pela questão do viés humano a influenciar positivamente ou negativamente os textos e informações. Há, contudo, momentos em que o viés apenas esteja presente como uma mera manifestação de preferência, na qual não estejam abaladas as questões da verdade, meia-verdade, mentira ou desinformação. De fato, considerando-se o teorema da incompletude de Gödel, estabelece-se uma base sólida para a aceitação de que, mesmo que as soluções tecnológicas, baseadas na inteligência computacional, na inteligência artificial, e nas teorias TAB e TMC, possam ser perfeitas ao primeiro olhar, haverá algo a ser acrescido. E talvez o que se possa acrescentar seja justamente uma estruturação que estabeleça um “fio” condutor que permita ao potencial interessado, conhecer toda a antecendência e descendência do texto ou da informação de seu interesse. Os resultados dessas complexas relações podem ser comunicados através dos mecanismos discutidos na TMC, especialmente quando consideramos os valores semânticos dos textos, transmitidos aos cognoscentes por mecanismos nos quais sejam consideradas as particularidades da informação moderna. Nos processos informacionais contemporâneos o “o quê” está sendo transmitido é relevante, sendo esse um dos principais desafios social, epistemológico e tecnológico a ser discutido na CI e áreas relacionadas aos temas deste estudo.

7 CONSIDERAÇÕES FINAIS

Os diálogos entre a TCM, a TAB e a IA abordados neste texto revelam um campo fértil para estudos e discussões na CI, são questões que afetam a representação e a recuperação

da informação. Sinalizam a necessidade de discussão em torno de um aporte teórico próprio da área, para nortear, por exemplo, trabalhos de curadoria de dados, visando mitigar as questões do viés humano nas informações, de modo a estabelecer ferramentas que disseminem informações corretas, desprovidas de tendências que possam corromper o valor semântico do que se pretende disseminar, ou então que permitam ao leitor saber de antemão, que a informação que lhe foi fornecida possui tendências teóricas, sociais, entre outras, permitindo ao leitor compreender que a informação pode ser tendenciosa.

REFERÊNCIAS

- ANDREZA, Pereira Batista, FARIAS, Gabriela Belmont De, NUNES, Jefferson Veras. Popularização Científica E Desinformação: Reflexões a Partir Das Percepções Públicas Da Ciência. *Encontros Bibli* 27.1 (2022): Encontros Bibli, 2022, Vol.27 (1). Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/85326>. Acesso em 19 mar. 2023.
- ATSERIAS, J.; CLIMENT, S.; FARRERES, X.; RIGAU, G.; RODRÍGUES, H. *Combining multiple methods for the automatic construction of multilingual WordNets*. In: ***Proceedings of the International Conference on Recent Advances in Natural Language Processing***, Tzigov Chark. 1997.
- BOOLE, GEORGE. ***Investigation of the Laws of Thought***. Cambridge University Press. 1854.
- BROWN, Tom B.; MANN, Benjamin; RYDER, Christopher; MCCANDLISH, Sam; RADFORD, Alec; SUTSKEVER, Ilya; AMODEI, Dario. ***Language Models are Few-Shot Learners***. OPENIA. 2020. Disponível em: <https://arxiv.org/pdf/2005.14165.pdf>. Acesso em 19 mar. 2023.
- BUOLAMWINI, John, GEBRU, Torn. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. in ***Proceedings of Machine Learning Research***, 81, 77-91. Disponível em <https://www.bit.ly/2ErqQPK>. Acesso em 19 mar. 2023
- BRUNIALTI, L. et al. **Aprendizado de Máquina em Sistemas de Recomendação Baseados em Conteúdo Textual: Uma Revisão Sistemática**. Disponível em: <https://sol.sbc.org.br/index.php/sbsi/article/view/5818/5716>. Acesso em: 18 jun. 2022.
- CAPURRO, Rafael.; HJORLAND, Birger. **O conceito de informação. Perspectivas em Ciência da Informação** [online]. 2007, v. 12, n. 1, pp. 148-207. Disponível em: <https://www.scielo.br/j/pci/a/j7936SHkZJkpHGH5ZNYQXnC/?lang=pt#ModalArticles>. Acesso em 08 jul. 2022.

CERVANTES, Brigida Maria Nogueira. Identificação e representação de conceitos por meio de termos. **Informação & Sociedade: Estudos**, [S. l.], v. 30, n. 4, p. 1–14, 2021. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/57261>. Acesso em 23 jun. 2023.

CESARINO, Maria Augusta Nobrega., PINTO, Maria Cristina Ferreira. Análise de assunto. *Revista de Biblioteconomia de Brasília*, Brasília, v. 8, n. 1, p. 32-43, jan./jun. 1980.

CONEGLIAN, Caio Saraiva.; VALENTIM, Marta Ligia Pontin.; SANTAREM Segundo, José Eduardo . Muli e interdisciplinaridade entre a Ciência da Informação e a Ciência da Computação no âmbito da Web Semântica. **Informação & Sociedade: Estudos**, [S. l.], v. 31, p. 1–18, 2021. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/52120>. Acesso em: 23 jun. 2023.

CURRÁS, Emília. *Investigaciones en inteligencia artificial en España*. **Transinformação**. 1989. Disponível em: <https://periodicos.puc-campinas.edu.br/transinfo/article/view/1699>. Acesso em: 19 mar 2023.

DE SOUZA, Osvaldo de; FERNANDES, Joana D’Arc Páscoa Bezerra. Discutindo a acessibilidade informacional no ecossistema digital. In: **Ciência da Informação em contextos**. Fortaleza: Imprensa Universitária, 2020. p. 141-163.

DE SOUZA, Osvaldo; TABOSA, Hamilton. Rodrigues; OLIVEIRA, David Martins; OLIVEIRA, Mayra Helena Um método de sumarização automática de textos através de dados estatísticos e processamento de linguagem natural. **Informação & Sociedade: Estudos**, João Pessoa, v. 27, n. 3, p. 307-320, set./dez. 2017.

FORBES. **ChatGPT tem recorde de crescimento na base de usuários**. Disponível em: <https://forbes.com.br/forbes-tech/2023/02/chatgpt-tem-recorde-de-crescimento-da-base-de-usuarios/>. Acesso em: 20 jun 2023.

FUJITA, Mariângela Spotti Lopes. Linguagem natural ou linguagem controlada? Influência da palavra-chave na representação para indexação e recuperação de informações. **Informação & Sociedade: Estudos**, [S. l.], v. 30, n. 4, p. 1–29, 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/57041>. Acesso em: 23 jun. 2023.

GUEDES, Willian.; ARAÚJO Junior, Rogerio Henrique. Estudo das similaridades entre a teoria matemática da comunicação e o ciclo documentário. **Informação & Sociedade: Estudos**, [S. l.], v. 24, n. 2, 2014. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/16498>. Acesso em: 23 jun. 2023.

GIDDENS, Anthony. **As consequências da modernidade**. Paulo: Unesp, 1991.

GÖDEL, Kurt. *On Formally Undecidable Propositions in Principia Mathematica and Related Systems I*. **Monatshefte für Mathematik und Physik**. Vol. 38: p. 173-98. 1931.

LISKOV, Barbara; Zilles, Stephen (1974). Programming with abstract data types. ***Proceedings of the ACM SIGPLAN Symposium on Very High Level Languages***. SIGPLAN Notices. 9. pp. 50–59.

LUND, K.; BURGESS, C. *Producing high-dimensional semantic spaces from lexical co-occurrence*. ***Behavior Research Methods, Instruments, & Computers***. 1996. V. 28(2) p.203–208.

MAIA, Francisca Clotilde De Andrade, FARIAS, Maria Giovanna Guedes Farias. "Contribuição Das Dimensões Dialógica E ética Da Mediação Da Informação Para a Revisão Por Pares Aberta." **Em Questão 29**. : p. 122916. 2023.

MCCULLOCH, Warren S.; PITTS, Walter. *A logical calculus of the ideas immanent in nervous activity*. ***The bulletin of mathematical biophysics***. 115–133. ISSN 1522-9602. 1943.

MENDONÇA, Claudio Marcio Campos; ANDRADE, Antônio Manuel Valente. Uso da IoT, Big Data e Inteligência Artificial nas capacidades dinâmicas: um estudo comparativo entre cidades do Brasil e de Portugal. **Informação & Sociedade: Estudos**, [S. l.], v. 29, n. 4, p. 37–60, 2019. Disponível em: <https://periodicos.ufpb.br/ojs2/47755>. Acesso em: 23 jun. 2023.

NHACUONGUE, Januário Albino; DUTRA, Moises Lima. A terminologia em Sistemas de Recuperação da Informação baseada na WORDNET.PT. **Informação & Sociedade: Estudos**, [S. l.], v. 30, n. 2, 2020. DOI: 10.22478/ufpb.1809-4783.2020v30n2.50756. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/50756>. Acesso em: 23 jun. 2023.

PELEGRINI, Cristiano Hugo. O Significado Contemporâneo da Teoria Matemática na Comunicação. *caderno.com*, v. 4, n. 2, p. 11–23, 2009. Disponível em: <https://www.researchgate.net/publication/268505957> Acesso em 11 jul. 2022.

PINHO, Fábio Assis; OLIVEIRA, Jéssica Pereira. Folksonomia e os conceitos de ordem, desordem e organização à luz de Edgar Morin. **Informação & Sociedade: Estudos**, [S. l.], v. 30, n. 4, p. 1–15, 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/57040>. Acesso em: 23 jun. 2023.

REIS, Márcia Cristina; FERNEDA, Edberto. Utilização de ontologias na Recuperação de Informação: principais abordagens, problemas e propostas de melhorias. **Informação & Sociedade: Estudos**, [S. l.], v. 31, p. 1–36, 2021. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/50273>. Acesso em: 23 jun. 2023.

SÁ, Alzira Tude. Uma abordagem matemática da informação: a teoria de Shannon e Weaver – possíveis leituras. *Logeion: Filosofia da Informação*, v. 5, n. 1, p. 48–70, 31 ago. 2018. Disponível em: <https://revista.ibict.br/fiinf/article/view/4245>. Acesso em 05 jul. 2022.

SHANNON, Claude. WEAVER, Warren. *The Mathematical Theory of Communication*.

University of Illinois Press, 1948. Disponível em:

<https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>.

Acesso em 19 mar. 2023.

TABOSA, Hamilton. Rodrigues; DE SOUZA, Osvaldo; CÂNDIDO, José. Carlos. S.; MELO, Ana Cristina Ursulino; REIS, Keila Braga. Avaliação do desempenho de um software de sumarização automática de textos. **INF. INF.** Londrina. v. 25. p. 189-210. 2020. Disp. em

https://www.researchgate.net/publication/342727034_Avaliacao_do_desempenho_de_um_software_de_sumarizacao_automatica_de_textos Acesso em 20 mar. 2023.

TSE, Tribunal Superior Eleitoral. Eleições 2018: **seminário vai discutir fake news e uso de robôs**. TSE. 2018. Disponível em <https://www.tse.jus.br/comunicacao/noticias/2017/Dezembro/eleicoes-2018-seminario-vai-discutir-fake-news-e-uso-de-robos>.

Acesso em 20 out. 2022.

SBPC. **Panorama da Ciência Brasileira 2015-2020**. Disp. em: <http://portal.sbpnet.org.br>

/noticias /producao- brasileira-de-artigos-cresce-32-em-2020-em-relacao-a-2015/. Acesso em 19 mar. 2023.