



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

RELAÇÕES TEMÁTICAS ENTRE ÉTICA E INTELIGÊNCIA ARTIFICIAL: UM ESTUDO HISTORIOGRÁFICO-BIBLIOMÉTRICO NAS BASES DE DADOS SCOPUS E WEB OF SCIENCE

THEMATIC RELATIONSHIPS BETWEEN ETHICS AND ARTIFICIAL INTELLIGENCE: A HISTORIOGRAPHIC-BIBLIOMETRIC STUDY IN THE SCOPUS AND WEB OF SCIENCE DATABASES

Natanael Vitor Sobral – Universidade Federal de Pernambuco (UFPE)

Eduardo Leite Alves – Universidade Federal de Pernambuco (UFPE)

Marcus Venicius Carvalho de Brito – Universidade Federal de Pernambuco (UFPE)

Modalidade: Trabalho Completo

Resumo: Objetiva analisar a produção de artigos sobre Inteligência Artificial relacionada à ética em bases de dados internacionais e interdisciplinares, enfocando aspectos de produção por ano, autoria e palavras-chave, destacando discussões seminais, clássicas e contemporâneas sobre o tema. Quanto aos procedimentos metodológicos, utiliza-se a historiografia bibliométrica, realizando-se as seguintes etapas: 1) Recuperação dos dados nas plataformas Web of Science e Scopus a partir do Bibliometrix, valendo-se de expressão de busca construída com apoio da literatura; 2) Download e armazenamento dos dados bibliométricos; 3) Tratamento dos dados no RStudio e Vantage Point; e 4) Representação gráfica via técnica de análise de redes sociais. Os principais resultados apontam: I) Produção crescente e ininterrupta desde 2015 com tendência ascendente; II) Consolidação de 26810 autores, sendo Luciano Floridi e Mariarosaria Taddeo os mais representativos, tanto na produtividade quanto nos relacionamentos; e III) Machine learning, deep learning, chatgpt e big data como os temas mais presentes na produção de artigos. Chama a atenção que a década foi a mais produtiva, mesmo possuindo apenas três anos de produção, o que demonstra o crescimento expressivo da produção.

Palavras-chave: ética; inteligência artificial; bibliometria; Web of Science; Scopus; Bibliometrix.

Abstract: It aims to analyze the production of articles on Artificial Intelligence related to ethics in international and interdisciplinary databases, focusing on aspects of production by year, authorship and keywords, highlighting seminal, classic and contemporary discussions on the topic. As for methodological procedures, bibliometric historiography is used, carrying out the following steps: 1) Data retrieval on the Web of Science and Scopus platforms from Bibliometrix, using a search expression constructed with support from the scientific literature; 2) Download and storage of bibliometric data; 3) Data processing in RStudio and Vantage Point; and 4) Graphical representation via social network analysis technique. The main results point out: I) Growing and uninterrupted production since 2015 with an upward trend; II) Consolidation of 26810 authors, with Luciano Floridi and Mariarosaria Taddeo

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

being the most representative, both in productivity and relationships; and III) Machine learning, deep learning, chatgpt and big data are the most present themes in the production of articles. It is noteworthy that 2020 was the most productive decade, even though it only had three years of production, which demonstrates the significant growth of the publications.

Keywords: ethics; artificial intelligence; bibliometrics; Web of Science; Scopus; Bibliometrix.

1 INTRODUÇÃO

Estudos recentes sobre as aplicações da inteligência artificial (IA) indicam que essa tecnologia pode trazer significativos benefícios para diversas áreas da sociedade. Aplicativos de rota geográfica, reconhecimento facial e casas inteligentes são apenas algumas das inúmeras contribuições da IA à vida cotidiana. De acordo com estudiosos, a combinação da IA e da robótica tem o potencial de livrar a humanidade do trabalho servil repetitivo, aumentar a produção de bens e de serviços e trazer uma era de prosperidade (Russell; Norvig, 2000).

Contudo, considerando o estágio atual de desenvolvimento destas tecnologias, contrapontos críticos vêm sendo discutidos enquanto sinal de alerta e superação da visão pueril de que a tecnologia pode solucionar todos os problemas da humanidade. Desta feita, entende-se que os algoritmos de IA devem ser melhor investigados por especialistas, sendo esta discussão consolidada a partir de diversos contributos, para além da Ciência da Computação, incluindo a Ciência da Informação (CI), Filosofia, Psicologia, entre outras áreas.

Neste sentido, a ética enquanto uma disciplina, surge como um componente fundamental para a compreensão das dinâmicas e implicações que envolvem a construção e o uso das IA. Inclusive, iniciativas nesta direção já vêm sendo desenvolvidas, merecendo menção o Grupo de Peritos de Alto Nível sobre IA, no âmbito da União Europeia. Este grupo apresentou o documento “Orientações éticas para uma IA de confiança” (Comissão Europeia, 2019), preconizando que uma IA só deve ser considerada confiável quando respeitar todas as disposições legislativas e regulamentares aplicáveis; observar os princípios e valores éticos; e alcançar robustez, tanto da perspectiva técnica quanto em relação ao ambiente social em que está inserida (Comissão Europeia, 2019).

Sobre as aplicações da IA na contemporaneidade, Lévy (2022) defende que estamos na era do decréscimo dos retornos cognitivos da IA neural. Isto significa que devido a problemas na qualidade dos dados usados para treinamento de IA, falta de resultados contextualizados que possam ser explicados e dificuldades em acumular e integrar o

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

conhecimento, o progresso no âmbito destas tecnologias parece ter parado nos últimos anos, sendo seu ápice na década 2010.

O autor chega a mencionar que os sistemas contemporâneos de aprendizagem de máquina não conseguem generalizar além dos limites dos dados de treinamento que lhes foram fornecidos, enquanto os humanos podem abstrair, conceituar e extrapolar o que aprendem. Numa perspectiva filosófica, as máquinas apesar de sua performance excepcional não entendem o significado dos textos que traduzem ou escrevem, assemelhando-se a papagaios mecânicos, capazes apenas de imitar, ou até superar o desempenho linguístico de um humano, sem entender nada do conteúdo dos textos que está traduzindo (Lévy, 2022).

Diante disto, Lévy (2022) propõe a *Information Economy Metalanguage* (IEMML), uma “metalinguagem semântica computável” que conecta máquinas e humanos, de modo semelhante a uma linguagem natural, permitindo que as categorias e as relações sejam rotuladas com palavras ou com frases mais elaboradas, minimizando a ambiguidade semântica, com clareza tanto para os modeladores quanto para os usuários finais, alinhando-se, portanto, aos princípios contemporâneos da ética e da transparência (Lévy, 2022).

Conforme a discussão apresentada, a IA se apresenta atualmente como um grande paradoxo, que representa a promessa de um futuro mais produtivo e próspero, e, também, a supressão das formas de inteligência humana, sendo vetor de desigualdade social, concentração de riquezas e diminuição da participação humana na sociedade com forte implicação nas dinâmicas de construção e disseminação do conhecimento, nas afetividades e nos princípios éticos que norteiam as formas de relacionamento entre as pessoas. Assim, a ética é um componente indissociável na interpretação do advento da IA, permitindo uma reflexão teleológica e de construção de princípios capazes de distinguir o local da ferramenta na construção de uma sociedade mais justa, fraterna e crítica, devendo esta discussão estar presente nas produções científicas enquanto reflexo da preocupação dos pesquisadores sobre este tema.

Assim, o objeto desta pesquisa é a produção de artigos sobre IA relacionados à ética em âmbito internacional. Para a composição deste *corpus*, utilizaram-se as bases de dados interdisciplinares *Scopus* e *Web of Science* (WoS), haja vista reunirem importante fração da produção científica qualificada sobre o tema, permitindo a coleta, processamento e análise dos dados para a proposição de *insights*.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

Na base da problemática deste trabalho, tem-se o questionamento sobre as características do discurso científico, na publicação de artigos acerca da IA, superando a visão tecnicista ao incluir a ética como um elemento indispensável à produção de IA e reflexão sobre seus propósitos e consolidação como um recurso efetivo de progresso. Desta feita, a pergunta norteadora da pesquisa é: Quais as características da produção de artigos sobre IA relacionados à ética, em bases de dados internacionais e interdisciplinares (*Scopus* e *WoS*), considerando todo o período de indexação da base?

Neste desiderato, objetiva-se analisar a produção de artigos sobre IA relacionada à ética em bases de dados internacionais e interdisciplinares, enfocando aspectos de produção por ano, autoria e palavras-chave, destacando discussões seminais, clássicas e contemporâneas sobre o tema à luz da historiografia bibliométrica e da Análise de Redes Sociais (ARS).

A presente pesquisa se justifica em função da necessidade de reafirmar o papel da ética na construção e aplicação de algoritmos de IA, posicionando-os enquanto um fenômeno do mundo real, obrigatoriamente subordinado ao Direito e às leis, atendendo às expectativas da sociedade de privacidade, governança, transparência, respeito aos direitos individuais e coletivos etc. (Kaufman, 2021). Ademais, nota-se um conflito entre as grandes corporações de tecnologia, denominadas *big techs*, que estão desenvolvendo as novas tecnologias da IA com preocupações diferentes da sociedade, enfocando a lucratividade e praticando a não observância das consequências, almejando, inclusive, a obtenção de imunidade dos seus atos através do discurso fatídico da *Web* enquanto um ambiente de total liberdade. Citando Ferrarezi (2023) e Magnolo e Whitaker (2024), as *big techs*, nos últimos dez anos, têm evoluído para se tornarem entidades soberanas, com poder de influência para além da economia, adentrando no mundo político.

2 METODOLOGIA

O presente estudo adota como abordagem metodológica a historiografia bibliométrica e ARS, com o intuito de analisar artigos seminais e contemporâneos sobre as relações estabelecidas entre ética e IA. A historiografia bibliométrica é uma derivação do conceito de historiografia algorítmica (Leydesdorff, 2010), onde se busca através de redes de citações representar a história de pesquisadores, áreas do conhecimento e temas, sendo o registro bibliométrico a base constituinte das análises, privilegiando a perspectiva temporal. No

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

presente trabalho, utilizam-se as variáveis autoria e palavras-chave relacionadas ao ano para o mapeamento do surgimento e consolidação autoral e temática do assunto focado ao longo dos anos. Este método aparece em Santos *et al.* (2012), reconstruindo a historiografia de um tema na forma de “mapa de visualização cronológica”, a partir dos metadados.

Desta feita, a consecução da pesquisa envolveu as seguintes etapas:

1) Recuperação dos dados: Nas bases de dados *Scopus* e *WoS*, foram realizadas pesquisas para obtenção das informações necessárias à análise bibliométrica. A escolha das bases se deu pela relevância internacional, ampla cobertura, credibilidade e ferramentas de análise avançadas. Assim, com base na literatura de Picalho, Lucas e Amorim (2022), foram utilizados buscadores booleanos como mecanismos de filtragem do tema de interesse a fim de ordenar a expressão de busca de forma lógica e obter resultados precisos, visando recuperar informações compatíveis com o desiderato da pesquisa, resultando na sintaxe:

(“Artificial Intelligence” OR “deep Learning” OR “Neural Network” OR “Machine learning” OR “Natural Language Processing” OR “system Intelligent*”) AND (Ethic*)*.

A busca foi realizada no dia 20 de julho de 2024, obtendo 8.209 títulos na base *Web of Science* e 15.735 títulos na base *Scopus*, totalizando 23.944 registros.

2) Download e armazenamento dos dados bibliométricos: Após a obtenção dos resultados, fez-se o *download* dos arquivos com os registros a serem tratados e analisados. Em ambas as fontes, foi escolhido o formato *BibTeX* (.bib), pela compatibilidade com os pacotes disponíveis no ambiente R. Os arquivos foram armazenados nos computadores dos pesquisadores e seguiram para a fase de tratamento.

3) Tratamento dos dados: Utilizou-se o *software RStudio* para tratar os dados, aplicando processos de limpeza e remoção de duplicatas dos registros. Após isto, obteve-se um total de 17.497 registros. Ao término da limpeza dos dados, houve conversão para o formato .xlsx (*Excel Open XML Spreadsheet*) e *upload* do arquivo para o *Bibliometrix*, propiciando a realização das primeiras análises. Aplicado o filtro em artigos de periódicos, chegou-se ao quantitativo de 9.585 artigos, reduzidos a 7.451 registros ao eliminar o ano 2024 por encontrar-se em andamento. Com isto, chegou-se ao *corpus* definitivo da pesquisa.

Em paralelo, a planilha foi carregada no *software Vantage Point*[®] (Porter; Cunningham, 2005) para correção dos registros e realizar eventuais agrupamentos de dados, tais como junção de palavras-chave e autores sinônimos através de rotinas de mineração de textos, chamadas de “*clean up*” com o propósito de mitigar a dispersão dos dados. Em seguida,

elaboraram-se *rankings* e matrizes, essas últimas, determinam a ausência ou presença de relação entre categorias de dados. Assim, obtiveram-se informações sobre a produção por ano; autores mais produtivos; temas mais prolíficos relacionando IA e ética; e historiografia temática através das décadas.

4) Representação gráfica analítica dos dados: A apresentação dos dados pautou-se em um gráfico de colunas com aplicação de linha de tendência linear visando analisar o crescimento da produção científica relacionada ao tema, construído em planilha de cálculo; rede de autores mais produtivos e de temas mais representativos utilizando a ferramenta VOSviewer (Van Eck; Waltman, 2010), a partir da técnica de ARS; e tabela com as décadas associadas aos temas com base em matrizes. Nos grafos, consideraram-se a intensidade das relações, o quantitativo de produções e a clusterização, estabelecendo as sub-redes de autores e temas com padrão similar.

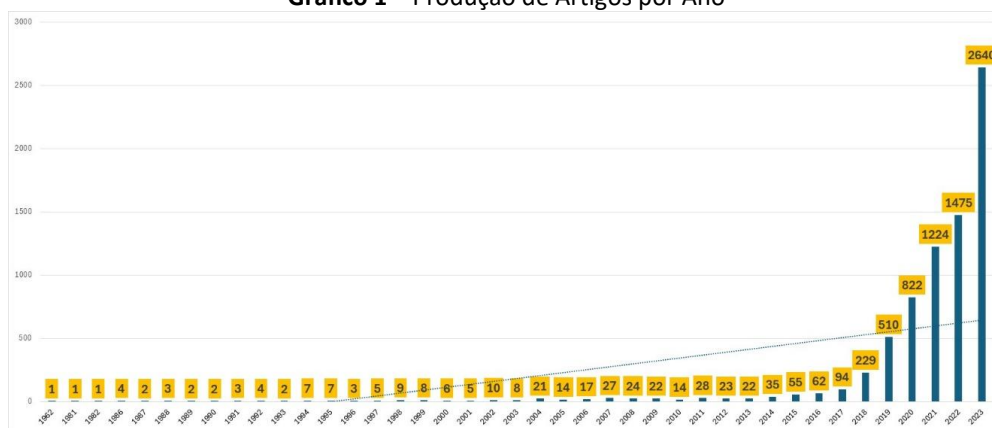
3 RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados da pesquisa e suas respectivas análises, em acordo com a diretriz proposta no objetivo. Desta feita, os dados estão dispostos na seguinte ordem: I) Gráfico de produção por ano, demonstrando o comportamento da publicação de artigos ao longo dos anos; II) Rede de autores mais produtivos, analisando-os frente as suas coautorias, ao seu perfil temático e aos seus indicadores bibliométricos; III) Rede de temas mais prolíficos relacionando ética e IA; e IV) Temas relacionados às décadas na busca pela construção de uma historiografia bibliométrica temática.

O gráfico 1, do tipo “colunas”, com linha de tendência do tipo “linear”, apresenta a distribuição dos artigos mapeados ao longo dos anos. Este indicador possibilita a identificação dos artigos seminais, permitindo a interpretação da gênese da produção de conhecimento no tema estudado. Assim, é possível observar tendências, obsolescências, estabilização ou crescimento do conhecimento produzido no assunto em questão.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

Gráfico 1 – Produção de Artigos por Ano



Fonte: Dados de pesquisa, 2024.

Conforme se pode observar, a produção é crescente e desde 2015 é ininterrupta e ascendente. Este comportamento foi objeto de debate na pesquisa “O cenário global das diretrizes éticas da IA” de Jobin, Ienca e Vayena (2019) e pode ser explicado através do aumento considerável da literatura científica acerca do tema visando as preocupações do meio acadêmico com ética em IA, especialmente, a partir do ano de 2016, quando as discussões sobre questões éticas se intensificaram devido ao aumento do uso das tecnologias de IA em diversas áreas. Para estes autores, há uma “convergência global emergindo em torno de cinco princípios éticos” (Jobin; Ienca; Vayena, 2019) relativos a IA, sendo eles, a transparência, justiça e imparcialidade, não maleficência, responsabilidade e privacidade.

Segundo Lee (2019), o uso da IA vem se intensificando em diversos segmentos da sociedade, fazendo parte da vida cotidiana das pessoas e das organizações, apresentando-se em sua quarta fase de desenvolvimento, ou “quarta onda”, as quais podem ser resumidas da seguinte forma: IA de internet, IA de negócios, IA de percepção e IA autônoma (Lee, 2019, p. 121). Desta feita, dilemas éticos em decorrência do seu uso são frequentes, pois estas tecnologias concebidas na forma de algoritmos vêm executando atividades tipicamente humanas, como a negociação de ações e diagnóstico de doenças, o que tem demandado reflexões acerca da responsabilidade e das consequências das decisões tomadas.

Ademais, o crescimento na produção de artigos confunde-se com o aumento da cobertura de indexação em bases de dados, o que vêm repercutindo em tendência ascendente no estudo de várias temáticas em diversificados campos do conhecimento. Na trajetória recente da produção estudada, em termos percentuais, nota-se comportamento expressivo no período que compreende os anos 2022 e 2023, com crescimento de 78,98%.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

As razões que motivam o desenvolvimento acadêmico-científico acerca da ética no contexto da IA são atribuídas a inúmeros fatores, incluindo os defendidos por Jobin, Lenca e Vayena (2019), que destacam a ética como fundamental ao desenvolvimento tecnológico das IA e os de Mittelstadt *et al.* (2016), que apontam a imprescindibilidade da responsabilidade algorítmica, transparência e equidade no desenvolvimento da IA enquanto um campo.

Ressalta-se ainda que à medida que a tecnologia avança, novas preocupações éticas surgem, o que reflete diretamente na literatura científica (Mittelstadt *et al.*, 2016), pois conforme destacam Jobin, Lenca e Vayena (2019), a discussão da IA como ferramenta passível de regulamentação, deve ser condizente com a garantia dos direitos humanos, devendo alinhar-se aos órgãos legais com a proposição de estratégias modernas de Estado.

O artigo de partida da análise bibliométrica, datado de 1962, tem como título “*Sophistication in computers: A disagreement*” e foi escrito por J. Kelly, filiada ao laboratório de pesquisas em telecomunicações *Bell Telephone Laboratories*, em Murray, New Jersey, EUA e Oliver G. Selfridge, pesquisador do *Massachusetts Institute of Technology*, EUA, sendo publicado na revista *IRE Transactions on Information Theory*.

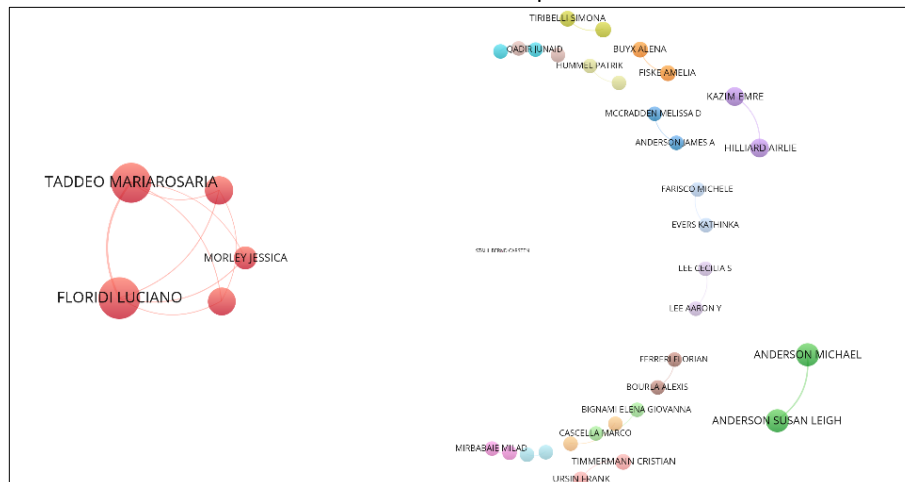
Na pesquisa supramencionada são abordadas discussões sobre IA e máquinas inteligentes e seu lugar na ciência a partir de dois pontos de vista propostos como “esquerda” e “direita”, objetivando identificar as duas escolas antípodas de pensamento sobre a filosofia básica das máquinas inteligentes. Os termos foram escolhidos em analogia ao contexto político para diferenciar perspectivas opostas no debate, mas sem implicar quaisquer conotações políticas partidárias. Em suma, o artigo procura fornecer uma compreensão mais profunda das complexidades envolvidas no debate em torno das capacidades das máquinas para pensar e da sofisticação dos computadores. No *Google Scholar*, possui 15 citações em diferentes anos.

Já o segundo artigo, de 1981, intitulado “*Managing the decision science technology: A case for ethical analysis?*”, foi elaborado por Harvey J. Brightman, associado à *Georgia State University*, EUA, e publicado na revista *Decision Sciences*. Reflete sobre a ciência da decisão e sua trajetória junto à tecnologia. Em tom crítico, argumenta que o crescimento da disciplina deve ser gerenciado em vez de ser ditado pelo imperativo tecnológico. Além disso, discute os aspectos éticos do cientista da decisão e das responsabilidades que a ele competem. No *Google Scholar*, possui sete citações em veículos diversificados.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

Para a análise de coautoria utilizaram-se os aportes da técnica de ARS, conectando no grafo os pesquisadores que publicaram artigos em conjunto (Gráfico 2). Ressalta-se que foi utilizado o campo com o nome completo do autor no padrão científico, pois o abreviado apresentou inúmeros problemas envolvendo pesquisadores orientais, que possuem nome e sobrenome parecidos, gerando dispersão no grafo.

Gráfico 2 – Rede de Autores mais representativos



Fonte: Dados de pesquisa, 2024.

Ao todo, foram mapeados 26957 autores, restando 26810 após a padronização dos nomes completos. Para o Gráfico 2, consideraram-se apenas os pesquisadores com mais de quatro relações de coautoria com algum par, chegando ao valor de 147 atores, 127 *clusters*, 24 *links* e *total link strenght* (TLS) de 304. Inicialmente, dada a característica quantitativa presente na rede, geralmente expressada pelo tamanho dos nós, inicia-se a análise pontuando os autores mais representativos na produção de artigos: Floridi Luciano (n=28), Taddeo Mariarosaria (n=21), Stahl Bernd Carsten (n=16), Anderson Michael (n=13), Ienca Marcello (n=13), Anderson Susan Leigh (n=12), Umbrello Steven (n=12), Mccradden Melissa D (n=11), Hagendorff Thilo (n=10), Kazim Emre (n=10) e Pagallo Ugo (n=10).

Em primeiro lugar, tem-se Luciano Floridi, pesquisador da *Yale University*, especialista em *AI Ethics*, *Digital Ethics*, *Information Ethics*, *Philosophy of Information* e *Philosophy of Technology*. Possui 56.210 citações no *Google Scholar* e Índice (n=115). Sua obra mais citada é *The fourth revolution: How the infosphere is reshaping human reality* (2014) e seu artigo mais citado é *The ethics of algorithms: Mapping the debate* (2016), publicado na *Big Data & Society*, em coautoria com Brent Daniel Mittelstadt; Patrick Allo; Mariarosaria Taddeo e Sandra Wachter. Este *paper* enfoca o *aboutness* do presente trabalho, sendo seu objetivo

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

esclarecer a importância ética da mediação algorítmica e fornecer um mapa prescritivo para organizar o debate.

Em seguida, encontra-se Mariarosaria Taddeo, professora de Ética Digital e Tecnologias de Defesa da *Oxford Internet Institute*, na *University of Oxford*, e especialista em *Ethics and AI*, *Military Ethics* e *Digital Ethics*. Possui 13.215 citações no *Google Scholar* e Índice h (n=47). Seu artigo mais citado é *The ethics of algorithms: Mapping the debate* (2016), em coautoria com Luciano Floridi e outros autores, conforme mencionado anteriormente.

Em terceiro lugar surge o nome de Bernd Carsten Stahl, professor de Pesquisa Crítica em Tecnologia na Escola de Ciência da Computação da *University of Nottingham*, e especialista em *Critical Theory*, *Information Systems*, *Computer Ethics*, *Digital Ethics* e *Responsible Innovation*. Possui 11.543 citações e índice h (n=54). Seu artigo mais citado é *Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice*, publicado no *International Journal of Information Management*, em coautoria com Yogesh K. Dwivedi et al. O paper traz 43 contribuições de especialistas em diferentes áreas para o debate sobre as capacidades do *ChatGPT* para o aumento da produtividade.

As relações mais profícuas identificadas envolvem: Luciano Floridi & Mariarosaria Taddeo (n=17); Michael Anderson & Susan Leigh Anderson (n=12); e Emre Kazim & Airlie Hilliard (n=8). A primeira denota forte afinidade temática no âmbito dos aspectos éticos em tecnologia, materializada na publicação *“What is data ethics?”* (2016), no periódico *Philosophical Transactions of the Royal Society: a Mathematical, Physical and Engineering Sciences*, onde os autores discutem a ética dos dados como um novo ramo da Ética que estuda e avalia problemas morais relacionados a dados, algoritmos, IA e práticas correspondentes, enfatizando a complexidade dos desafios éticos em *data science*.

A segunda relação mais intensa se dá entre os autores Michael Anderson e Susan Leigh Anderson, impulsionada pelo interesse comum em: ética na IA, computação ética e filosofia da IA. Em conjunto, desenvolveram a obra *“Machine Ethics”*, publicado na *Cambridge University Press*, no ano 2011. O livro discorre sobre o então novo campo da ética das máquinas, abordando a preocupação em atribuir às máquinas princípios éticos e a criação de procedimentos para resolver os dilemas éticos que as permitam funcionar de forma responsável através da sua própria tomada de decisões.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

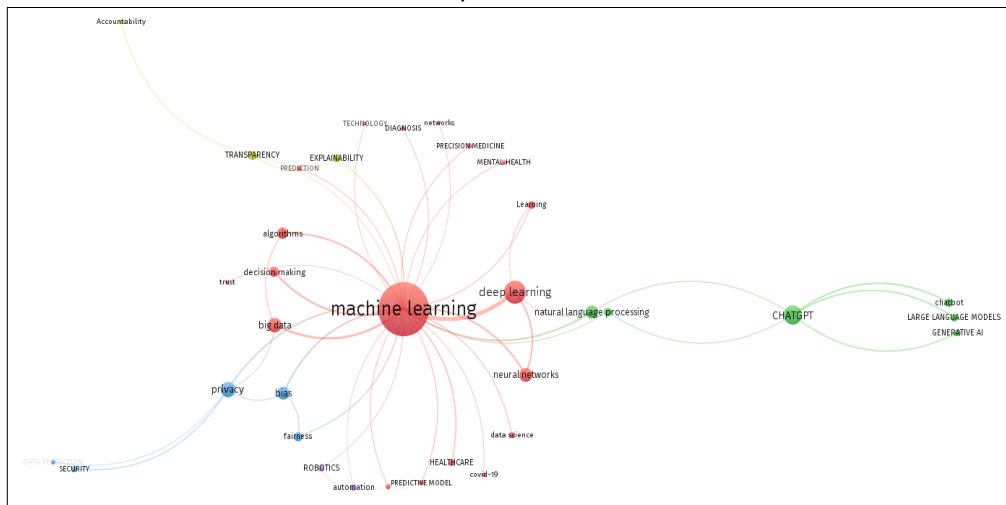
A terceira relação é entre os autores Emre Kazim e Airlie Hilliard que debatem temas relacionados a tecnologia, ética e IA, especificamente, aspectos sociais e de trabalho relacionados a automação de processos e atividades que antes necessitavam da ação exclusiva humana. Em conjunto com outros autores, produziram o artigo “*Systematizing Audit in Algorithmic Recruitment*”, de 2021, onde debatem os algoritmos empregados na identificação das personalidades compatíveis aos interesses de trabalho em recrutamentos, incluindo conversação com *chatbot* e discussões sobre como as avaliações de auditoria em sistemas baseados em IA podem ser usadas para garantir justiça, governança, análise de riscos, transparência e maneiras de medir preconceitos em algoritmos.

Na análise de coocorrência entre os autores mais produtivos, nota-se que para além da temática comum que os conecta, é relevante mencionar que tais ligações acontecem em razão da colaboração entre grupos de pesquisa institucionais e não necessariamente de uma iniciativa de agrupamento internacional que debruce esforços na pesquisa do tema.

O Gráfico 3 apresenta as relações de coocorrência entre palavras-chave. Inicialmente, identificaram-se 14541 palavras-chave, restando 12943 após a limpeza dos dados. Para a visualização, excluíram-se as palavras ética e IA com o propósito de focar elementos com maior representatividade semântica e interpretativa e pela necessidade de explorar a profundidade e diversidade dos temas adjacentes. Esta estratégia proporciona uma análise que detalha e possibilita a identificação de temas emergentes que orbitam os centrais.

Outro critério selecionado foi a inclusão de termos com mais de 14 relacionamentos com outro par coocorrente. Chegou-se a este número exercitando a visibilidade da rede, numa tentativa de aproximação com os parâmetros estabelecidos por Moreno (1992), proponente do sociograma. Ao final, 35 palavras foram identificadas, 5 *clusters*, 44 *links* com TLS (n=2608), denotando importante densidade.

Gráfico 3 – Temas mais prolíficos relacionando IA e ética



Fonte: Dados de pesquisa, 2024.

Os temas mais representativos na produção de artigos são: *machine learning* (n=1114); *deep learning* (n=336); *chatgpt* (n=256); *big data* (n=253); *robotics* (n=245); *decision making* (n=219); *algorithms* (n=217); *privacy* (n=216); *neural networks* (n=197); *natural language processing* (n=168); *bias* (n=163); *covid-19* (n=143); *learning* (n=140); *governance* (n=137); *trust* (n=135); *technology* (n=133); *healthcare* (n=126); *education* (n=124) e *chatbot* (n=117); *regulation* (n=103).

Em primeiro lugar, tem-se *machine learning*, subárea da IA, entendido por Mitchell (1997b), como campo que se concentra no desenvolvimento de algoritmos e modelos que permitem que os computadores aprendam a partir de dados disponíveis e façam previsões ou tomem decisões sem serem explicitamente programados para executar tarefas específicas e identificar padrões. Em essência, o objetivo do *machine learning* é capacitar os computadores a aprender e melhorar automaticamente com a experiência, sem intervenção humana direta (Mitchell, 1997b). Em seguida, encontra-se o *deep learning*, subárea do *machine learning* que utiliza redes neurais artificiais com várias camadas de redes neurais profundas (*deep neural networks*) para modelar e resolver problemas complexos de grandes volumes de dados (Goodfellow; Bengio; Courville, 2016).

Adiante, tem-se o *chatgpt*, um modelo de linguagem baseado na arquitetura GPT (*Generative Pre-trained Transformer*) com a finalidade de gerar textos de forma coerente e natural, similar à escrita humana (Božić; Poola, 2023). Seu funcionamento envolve um processo de treinamento extensivo utilizando grandes *corpora* de texto, onde o modelo aprende a prever a próxima palavra em uma sequência, capturando padrões linguísticos

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

complexos, além considerar múltiplos contextos no texto de entrada, resultando em respostas mais precisas e contextualizadas (Božić; Poola, 2023).

As relações mais intensas no que tange a pares de palavras contempla: *machine learning & deep learning* (n=132); *machine learning & big data* (n=70). A primeira relação, conforme descrito por Mitchell (1997b), é explicada pelo fato de *machine learning* ser um subcampo da IA focado em aprender a partir de dados e *deep learning* é sua subárea focada em redes neurais profundas para modelagem de padrões complexos. Ambas se alinham de forma hierárquica e complementar.

Na relação entre *machine learning* e *big data*, o primeiro tema se beneficia do segundo em virtude da necessidade de capacidade de processamento para lidar com grandes volumes e variedade de dados com velocidade, conforme preconiza o *big data*. Assim, os algoritmos de aprendizagem melhoram com mais dados disponíveis para treino, sendo o avanço do *big data* essencial para o progresso em *machine learning*, permitindo a análise de dados em escala sem precedentes.

A tabela 1 relaciona as palavras-chave dos autores com as décadas das publicações, consolidando dados acerca das opções temáticas dos pesquisadores ao longo dos anos, no escopo das relações entre ética e IA. Fez-se um recorte visando elencar os temas mais representativos de cada ano, sendo os anos 1960 de menor complexidade por possuir apenas um artigo, e a década 2020, mesmo com apenas quatro anos registrados (2020, 2021, 2022 e 2023), a mais complexa por abrigar muitos artigos e vasta quantidade de palavras-chave. Similarmente ao gráfico 3 foram excluídas as palavras IA.

Tabela 1 – Temas mais prolíficos relacionando IA e ética

Décadas / Artigos	Frequência das palavras-chave
1960 (n=1)	<i>machine intelligence</i> (n=1), <i>programming</i> , <i>educational institutions</i> (n=1), <i>telephony</i> (n=1), <i>brain modeling</i> (n=1), <i>biology computing</i> (n=1).
1970	Não houve artigos
1980 (n=13)	<i>education</i> (n=1), <i>copyright law</i> (n=1), <i>expert systems</i> (n=1), <i>information systems</i> (n=1), <i>computer</i> (n=1), <i>software</i> (n=1), <i>aristotle's syllogistics</i> (n=1), <i>policy</i> (n=1), <i>determination analysis</i> (n=1), <i>determination logic</i> (n=1), <i>logic of natural language</i> (n=1), <i>empirical humanitarian studies</i> (n=1), <i>data bases</i> (n=1), <i>social issues</i> (n=1), <i>information retrieval</i> (n=1), <i>empirical nonquantitative data processing</i> (n=1), <i>not-for-profit management</i> (n=1), <i>strategy</i> (n=1), <i>on-line information retrieval</i> (n=1).
1990 (n=50)	<i>neural networks</i> (n=4), <i>expert systems</i> (n=3), <i>decision making</i> (n=3), <i>closed-loop-control</i> (n=2), <i>concept of health</i> (n=2), <i>inflammation</i> (n=2), <i>medical knowledge-based systems</i> (n=2), <i>knowledge based-systems</i> (n=2), <i>mivacurium</i> (n=2), <i>muscle relaxants</i> (n=2), <i>rat</i> (n=2), <i>nucleus centralis of amygdala</i> (n=2), <i>urinary bladder</i> (n=2).

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

2000 (n=154)	<i>neural networks</i> (n=11), <i>robotics</i> (n=10), <i>decision making</i> (n=7), <i>computer ethics</i> (n=5), <i>natural language processing</i> (n=4), <i>morality</i> (n=4), <i>privacy</i> (n=4), <i>connectionism</i> (n=4), <i>information</i> (n=4), <i>artificial life</i> (n=4).
2010 (n=1072)	<i>machine learning</i> (n=139), <i>robotics</i> (n=82), <i>big data</i> (n=50), <i>neural networks</i> (n=41), <i>algorithms</i> (n=38), <i>deep learning</i> (n=32), <i>automation</i> (n=29), <i>technology</i> (n=27), <i>privacy</i> (n=25), <i>decision making</i> (n=24), <i>autonomy</i> (n=23), <i>governance</i> (n=20), <i>natural language processing</i> (n=16), <i>social media</i> (n=15), <i>accountability</i> (n=15), <i>robot ethics</i> (n=14), <i>innovation</i> (n=13), <i>law</i> (n=13), <i>autonomous vehicles</i> (n=12).
2020 (n=6161)	<i>machine learning</i> (n=972), <i>deep learning</i> (n=304), <i>chatgpt</i> (n=256), <i>big data</i> (n=203), <i>privacy</i> (n=187), <i>decision making</i> (n=185), <i>algorithms</i> (n=177), <i>bias</i> (n=154), <i>robotics</i> (n=153), <i>natural language processing</i> (n=148), <i>covid-19</i> (n=143), <i>neural networks</i> (n=141), <i>learning</i> (n=129), <i>trust</i> (n=126), <i>healthcare</i> (n=121).

Fonte: Dados de pesquisa, 2024.

Apesar de a primeira produção ter sido publicada em 1960, percebe-se que as pesquisas só se consolidam nos anos 1980 à medida que encontram-se presentes, de forma sequenciada, nos quatro últimos desta década. A característica da produção deste período é a dispersão, sendo marcada pela abordagem de temas distintos, que vão desde educação até recuperação da informação em bases de dados. Os anos 1990 introduzem de forma representativa as discussões sobre redes neurais, enfocando aspectos da tomada de decisão em contextos de aplicação da área Médica.

Os anos 2000 são caracterizados pela adição das discussões robótica, sendo a IA apresentada como a base da unidade de processamento de máquinas, agindo de forma inteligente e autônoma, similarmente à ação do cérebro sobre o restante do corpo humano. Diante deste avanço, a *computer ethics* apresenta importante crescimento, abordando questões sociais, políticas, jurídicas e éticas que envolvem a tecnologia da informação (Moor, 2001). Além disto, o NLP inicia seu processo de consolidação como objeto de interesse da Ética, fundando as bases de um dos aportes centrais da ética na IA: a linguagem.

Na década 2010, tem-se a difusão massiva, na produção de artigos, dos pilares contemporâneos da IA, envolvendo: *machine learning*, trazendo consigo as discussões sobre *big data* conforme já explicado no presente trabalho; *neural networks* e *deep learning*; e *natural language processing*, demonstrando um claro movimento dos estudiosos da Ética em direção aos aportes da IA e dos especialistas da IA rumo aos fundamentos da Ética. Além disto, com o advento das redes sociais e com o uso de dados dos usuários pelas *big techs*, questões de privacidade passaram a integrar frequentemente os assuntos no campo a partir das dimensões éticas e legais. Um contexto de aplicação razoavelmente representado nesta década foi o advento dos veículos autônomos, pois, segundo Cunneen, Mullins e Murphy

(2019), a sociedade terá de enfrentar conjunto de riscos inéditos, à medida que a IA incorporada nestes veículos tomarão decisões complexas que podem gerar consequências tangíveis de vida ou morte.

Por fim, tem-se os anos 2020, que apresentam como novidade o *chatgpt*, demonstrando a centralidade da linguagem enquanto *modus operandi* e *cogitandi* da IA, podendo ser fonte de virtude e conhecimento, mas, também, da expressão de vieses. E, a pandemia de covid-19, primeira grande crise da humanidade a dispor de recursos avançados de IA generativa.

O estudo historiográfico-bibliométrico da ética no contexto da IA é de fundamental importância para a compreensão da incorporação de princípios éticos nestas tecnologias, especialmente na atualidade, em que a IA vem se tornando cada vez mais autônoma e influente em áreas essenciais à vida humana, como saúde, justiça e segurança. Os algoritmos podem reproduzir preconceitos, violar direitos e causar danos, caso não haja uma supervisão humana ética, baseada na transparência e orientada ao bem-estar social. Para isso, é essencial o desenvolvimento de marcos regulatórios adequados, que estejam à altura das expectativas da sociedade civilizada, contrapondo-se ao modelo de barbárie já presente no contexto das mídias sociais digitais sob o rótulo de "liberdade".

4 CONSIDERAÇÕES FINAIS

Este trabalho buscou analisar a produção de artigos sobre IA relacionada à ética na WoS e Scopus à luz da historiografia bibliométrica e da ARS. Com isto, verificou-se que a produção é ininterrupta desde 2015, apresentando tendência ascendente, o que evidencia o interesse e a imprescindibilidade da ética no debate científico sobre IA. Atualmente, Luciano Floridi e Mariarosaria Taddeo, vêm liderando as discussões relacionadas ao tema, consolidando o lugar da Filosofia neste debate.

Os temas mais representativos, *machine learning*, *deep learning* e *chatgpt*, têm sua representatividade associada à centralidade que possuem no campo da IA, tornando-se objeto de pesquisa da ética, dadas as suas possibilidades de revolucionar e influenciar dinâmicas na sociedade e nas organizações.

Assim, o presente estudo contribui com a área de CI, à medida que a posiciona em lugar de protagonismo na construção de políticas de informação que possam incorporar a

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

dimensão ética aos fenômenos tecnológicos, que, em geral consolidam-se a partir de bases pragmáticas e tecnocratas, sem considerar as repercussões humanísticas implicadas. Ademais, através da metodologia de investigação bibliométrica, foi possível identificar a gênese, o desenvolvimento e as tendências emergentes do campo, permitindo visualizar quais autores e obras são estão moldando o debate entre ética e IA.

Acerca das limitações, espera-se em pesquisas futuras ampliar o repertório das plataformas utilizadas, incluindo bases de dados com predominância de artigos em língua portuguesa, a exemplo da Brapci e do *Scielo*. Enquanto perspectivas de trabalhos futuros, há duas dissertações sendo desenvolvidas por autores deste trabalho. A primeira objetiva investigar a IA sob a ótica da CI, buscando compreender as influências práticas e teóricas nas relações de organização, disseminação, acesso e uso da informação frente aos desafios éticos. A segunda tem o propósito de realizar uma análise bibliométrica sobre a presença da literatura *sci-fi* na produção de artigos científicos relacionados a AI no domínio da ética. Assim, a elaboração do presente trabalho constitui-se em importante contribuição metodológica para a obtenção de conhecimento sobre o domínio da IA numa perspectiva das humanidades; e acrescenta conhecimento sobre as literaturas mais representativas na temática das pesquisas do Programa de Pós-Graduação a qual está associado.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001; e da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (Facepe) – Bolsa de Mestrado.

REFERÊNCIAS

BOŽIĆ, Velibor; POOLA, Indrasen. Chat GPT and education. United States: **Preprint**, 2023. DOI 10.13140/RG.2.2.18837.40168

COMISSÃO EUROPEIA. **Orientações éticas para uma IA de confiança**. Bruxelas: Comissão Europeia, 2018. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Acesso em: 11 jun. 2024.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

CUNNEEN, Martin; MULLINS, Martin; MURPHY, Finbarr. Autonomous vehicles and embedded artificial intelligence: the challenges of framing machine driving decisions.

Applied Artificial Intelligence, London, v. 33, n. 8, p. 706-731, 2019. DOI:

<https://doi.org/10.1080/08839514.2019.1600301>

FERRAREZI, Thiago. Equilibrando o poder da inteligência artificial: uma reflexão sobre governança estatal e paradoxos futuros. **Migalhas**, Ribeirão Preto, ago. 2023. Disponível em:

<https://www.migalhas.com.br/depeso/392554/inteligencia-artificial-reflexao-sobregovernanca-estatal-e-paradoxos>. Acesso em: 8 dez. 2023.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. Cambridge: MIT press, 2016.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. The global landscape of AI ethics guidelines.

Nature machine intelligence, London, v. 1, n. 9, p. 389-399, 2019. DOI:

<https://doi.org/10.1038/s42256-019-0088-2>

KAUFMAN, Dora. Inteligência artificial e os desafios éticos: a restrita aplicabilidade dos princípios gerais para nortear o ecossistema de IA. **PAULUS: Revista de Comunicação da FAPCOM**, São Paulo, v. 5, n. 9, 2021. Disponível em:

<https://revista.fapcom.edu.br/index.php/revista-paulus/article/view/453/427>. Acesso: 26 jun.2024.

LEE, Kai-Fu. Inteligência artificial recurso eletrônico: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos. Tradução Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019. Disponível em:

[https://www.kufunda.net/publicdocs/Intelig%C3%Aancia%20artificial%20\(Kai-Fu%20Lee\).pdf](https://www.kufunda.net/publicdocs/Intelig%C3%Aancia%20artificial%20(Kai-Fu%20Lee).pdf). Acesso em: 16 set. 2024.

LÉVY, Pierre. IEML: rumo a uma mudança de paradigma na inteligência artificial. **Matrizes**, São Paulo, v. 16, n. 1, p. 11-34, 2022. Disponível em:

<https://www.redalyc.org/articulo.oa?id=143071289015>. Acesso em: 6 jun. 2024

LEYDESDORFF, Loet. Eugene Garfield and algorithmic historiography: Co-words, co-authors, and journal names. **Annals of Library and Information Studies**, Amsterdam, v. 57, n. 3, p. 248–260, 2010.

MAGNOLO, Talita Souza; WHITAKER André Machado Coelho. As relações de poder com a IA: perspectivas corporativistas no campo da Comunicação. **Organicom**, São Paulo, Brasil, v. 21, n. 44, p. 55–70, 2024. Disponível em:

<https://www.revistas.usp.br/organicom/article/view/220292>. Acesso em: 9 jul. 2024.

MITCHELL, T. M. Artificial neural networks. **Machine learning**, v. 45, n. 81, p. 127, 1997a.

MITCHELL, Tom M. **Machine learning**. New York: McGraw-hill, 1997b.

MITTELSTADT, Brent Daniel; ALLO, Patrick; TADDEO, Mariarosaria; WACHTER, Sandra; FLORIDI, Luciano. The ethics of algorithms: mapping the debate. **Big Data & Society**, Los Angeles, v. 3, n. 2, p. 1-21, 2016. DOI: <https://doi.org/10.1177/2053951716679679>

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

MOOR, James H. The future of computer ethics: you ain't seen nothin'yet!. **Ethics and Information Technology**, Berlim, v. 3, p. 89-91, 2001.

MORENO, Jacob Levy. **Quem sobreviverá?** fundamentos da sociometria, psicoterapia de grupo e sociodrama. Goiânia: Dimensão, 1992.

PICALHO, Antonio Carlos; LUCAS, Elaine Rosangela de Oliveira; AMORIM, Igor Soares. Lógica booleana aplicada na construção de expressões de busca. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 11, p. 1-12, 2022. DOI: <https://doi.org/10.5380/atoz.v11i0.81838>

PORTER, Alan L.; CUNNINGHAM, Scott W. **Tech mining**: exploiting new technologies for competitive advantage. Nova Jersey: Wiley Online Library, 2005. 384 p.

RUSSELL, Stuart; NORVIG, Peter. **Artificial intelligence**: a modern approach. 4. ed. London: Pearson.

SANTOS, R. N. M. dos.; HOLANDA, C. M. S.; SILVA, F. M. E.; SILVEIRA, M. A. A. Historiografia da atividade científica: reflexões sobre o papel da teoria "vis-à-vis" da prática. *In*: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 3., 2012, Gramado. **Anais [...]** Gramado: EBBC, 2012. p. 1-7.

VAN ECK, Nees Jan; WALTMAN, Ludo. Software survey: vosviewer, a computer program for bibliometric mapping. **Scientometrics**, Amsterdam, v. 84, n. 2, p. 523-538, 2010.