O LADO “SOMBRIO” DA INTELIGÊNCIAARTIFICIAL: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Autoria

Henrique Vassali - henrique.vassali@gmail.com
Prog de Pós-Grad em Admin/Esc de Admin – PPGA/EA / UFRGS - Universidade Federal do Rio Grande do Sul

Raquel Janissek-Muniz - rjmuniz@ufrgs.br
Prog de Pós-Grad em Admin/Esc de Admin – PPGA/EA / UFRGS - Universidade Federal do Rio Grande do Sul

Resumo

É inegável o potencial transformador da Inteligência Artificial (IA), que impacta desde a capacidade inovativa e tecnológica das organizações até seu desempenho em termos de processos e economia. Sua conceitualização abrange uma grande variedade de aspectos tecnológicos e consequentes interações com as organizações e a sociedade em geral. No entanto, enquanto maior parte da atenção tem focado na perspectiva positiva do desenvolvimento e uso dessa tecnologia, existe uma preocupação crescente com potenciais aspectos negativos. Esta visão, de potenciais aspectos negativos ou não intencionais da IA – ou seu “lado sombrio” – têm provocado atenção de entidades públicas, privadas, e da comunidade científica, dado seu significativo impacto potencial, embora essa discussão seja recente e portanto fértil para aplicações e teorizações aprofundadas. Nesse sentido, e visando identificar estudos já realizados que contemplem o tema, este trabalho apresenta uma Revisão Sistemática da Literatura acerca do lado sombrio da IA. Uma busca em bases de dados resultou em 27 publicações que abordam, sob alguma perspectiva, o lado sombrio da IA. Após análise destas publicações, apresentam-se os principais aspectos já pesquisados e propõe-se temas proeminentes para agenda futura de pesquisa.

O LADO "SOMBRIO" DA INTELIGÊNCIA ARTIFICIAL: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Resumo: É inegável o potencial transformador da Inteligência Artificial (IA), que impacta desde a capacidade inovativa e tecnológica das organizações até seu desempenho em termos de processos e economia. Sua conceitualização abrange uma grande variedade de aspectos tecnológicos e consequentes interações com as organizações e a sociedade em geral. No entanto, enquanto maior parte da atenção tem focado na perspectiva positiva do desenvolvimento e uso dessa tecnologia, existe uma preocupação crescente com potenciais aspectos negativos. Esta visão, de potenciais aspectos negativos ou não intencionais da IA – ou seu "lado sombrio" – têm provocado atenção de entidades públicas, privadas, e da comunidade científica, dado seu significativo impacto potencial, embora essa discussão seja recente e portanto fértil para aplicações e teorizações aprofundadas. Nesse sentido, e visando identificar estudos já realizados que contemplem o tema, este trabalho apresenta uma Revisão Sistemática da Literatura acerca do lado sombrio da IA. Uma busca em bases de dados resultou em 27 publicações que abordam, sob alguma perspectiva, o lado sombrio da IA. Após análise destas publicações, apresentam-se os principais aspectos já pesquisados e propõe-se temas proeminentes para agenda futura de pesquisa. **Palavras-chave:** Inteligência Artificial; Lado sombrio; IA Responsável; IA Ética; Governança em IA.

1. Introdução

De acordo com González-Esteban e Calvo (2022), a Inteligência Artificial (IA) tornou-se uma das principais forças disruptivas da sociedade no século XXI em diferentes campos de atuação, mas nem sempre de maneira positiva. Segundo os autores, por um lado a IA oferece potencial de melhorias em diferentes processos organizacionais, como os produtivos, comunicativos, participativos, decisórios, e de inovação, em termos de sustentabilidade, capacidade, consistência, eficiência, entre outras características. Por outro lado, a IA também pode apresentar vieses menos positivos devido ao seu envolvimento - direto ou indireto - no aumento exponencial da complexidade subjacente, gerando níveis mais elevados de incerteza, desigualdade, instrumentalização, reificação, alienação, psicopatologias, etc. (Calvo, 2020b, 2021; González-Esteban e Calvo, 2022; Prunkl et al., 2021).

Na conjuntura dos constantes desenvolvimentos tecnológicos e sua contextualização social, Mikalef et al. (2022) observam que existe uma crescente e fraccionada tensão entre as capacidades tecnológicas e as estruturas sociais humanas nas quais residem essas tecnologias. Nessa intensa interação da tecnologia e dinâmicas sociais, Cheng et al. (2021) observam que existe uma crescente atenção da literatura de Sistemas de Informação (SI) para o lado sombrio da Tecnologia da Informação. Mais especificamente, a Inteligência Artificial, em particular, vem atraindo atenção em relação às implicações que seu lado "sombrio" pode potencialmente resultar (Mikalef et al, 2022). Segundo Devaraj et. al. (2019), embora a IA tenha se tornado um suporte quase onipresente no contexto tecnológico, em diversas aplicações, as conotações negativas e equívocos associados também vêm crescendo. Nessa perspectiva, questões como segurança, confiança, privacidade e justiça aparecem relacionadas ao lado sombrio da Inteligência Artificial, podendo ser inerente aos sistemas inteligentes (Jabbarpour, 2021).

Em relação à literatura desenvolvida neste contexto, Mikalef et al. (2022) observam que a pesquisa em Sistemas de Informação é dominada por estudos que se concentram no poder revolucionário e positivo da tecnologia. No entanto, os autores argumentam que,

recentemente, o campo de SI está começando a atentar para formas complexas e mesmo alarmantes de como o uso de tecnologia da informação (TI) afeta a vida organizacional e social, configurando o lado sombrio da tecnologia e seu uso. No campo da Inteligência Artificial, Salo et al. (2018) destacam que observar esse contexto com uma lente negativa poderia permitir visualizar circunstâncias ou cenários que foram até então ignorados pela pesquisa, favorecendo uma compreensão abrangente e diferenciada do fenômeno da IA.

Todavia, observa-se na literatura certo desequilíbrio entre o reconhecimento relativo ao lado sombrio da IA e estudos já desenvolvidos para cercar a questão. Embora admite-se que a IA tenha potencial de induzir riscos a nível individual, organizacional e social (Alt, 2018), Cheng et al. (2022) argumentam que extensiva atenção tem sido dada aos aspectos positivos do seu uso, enquanto o lado sombrio da IA recebe relativamente menos atenção, especialmente da comunidade acadêmica. Para Mikalef et al. (2022), este desequilíbrio pode ocorrer pelo fato de questões tipicamente marginalizadas, como anormais ou desviantes, raramente serem investigadas com rigor. Nesse sentido, considerando a importância, universalidade e as potenciais consequências trazidas pela IA, investigar esse fenômeno e possíveis impactos também negativos merece maior exploração (Cheng et al, 2022).

Para Rana et al. (2021), a IA gera diversos benefícios, mas inevitavelmente podem ocorrer consequências negativas. Zeng e Wu (2021) destacam que a exploração do lado sombrio da Inteligência Artificial ainda está em uma fase imatura. Assim, quanto antes se começar a pesquisar e contemplar quais seriam essas potenciais consequências, melhor preparado as organizações poderiam estar para mitigar e gerenciar tais perigos (Marr, 2021), em uma abordagem proativa e antecipativa de enfrentar a situação. Assim, assumindo a problemática relativa ao potencial "gap" de estudos do fenômeno sombrio da Inteligência Artificial e suas implicações, e o possível estado de "imaturidade" das pesquisas na temática, entende-se, como motivação deste trabalho, que se torna pertinente identificar e entender qual é o atual estado da arte relativo ao tópico e, em uma tentativa de estruturar as abordagens já realizadas, direcionar a investigação desta pesquisa para responder à seguinte questão: **Qual o estado da arte referente à perspectiva do lado sombrio da Inteligência Artificial?**

Dentro do objetivo de analisar qual é o estado da arte relativo ao lado sombrio da Inteligência Artificial, em uma tentativa de aprofundar o entendimento sobre esse potencial desequilíbrio levantado nesta problemática, optou-se pelo desenvolvimento de uma Revisão Sistemática de Literatura, a fim de identificar e analisar publicações existentes relativas ao tema. O presente trabalho está organizado da seguinte forma: a seção 2 apresenta uma revisão de literatura sobre o lado sombrio da IA; a seção 3 descreve o método utilizado no estudo; a seção 4 traz os resultados da análise; e na seção 5 são apresentadas as considerações finais com indicações de futuras pesquisas.

2. Referencial Teórico

Segundo Cheng et al. (2022), contextualizada no século XXI, a Inteligência Artificial é uma inovação extremamente disruptiva que atraiu considerável atenção de profissionais e acadêmicos. Os autores argumentam que a IA oferece oportunidades sem precedentes para mudanças fundamentais e atualizações abrangentes em muitos setores. Na perspectiva de sua definição, Mikalef e Gupta (2021, p.2) trazem uma abordagem integrativa para a noção de IA, definindo-a como "a habilidade de um sistema identificar, interpretar, fazer inferências e aprender através de dados, para alcançar determinados objetivos organizacionais e sociais".

No entanto, para além dos numerosos potenciais positivos, Cheng et al. (2022) apontam que podem haver diversas consequências negativas relacionadas ao uso da Inteligência Artificial. Nesse sentido, Floridi et al. (2018) também defendem que não se pode apenas olhar para o potencial positivo e promissor da IA. Para Jia e Zhang (2021), existem desafios específicos na perspectiva dos riscos e do lado sombrio da IA, evidenciando três deles: a **complexidade técnica** – ou a "caixa preta" – comumente associada à IA, a qual pode não deixar claro os limites entre cada *stakeholder* quando a responsabilidade legal precisar ser determinada (Liu et al, 2019; Jia e Zhang, 2021); o **mecanismo de "auto reforço"** (*self-reinforcing*) da IA que pode potencialmente exacerbar problemas sociais existentes tais como preconceito e discriminação (Nelson, 2019; Jia e Zhang, 2021); e a **subjetividade da IA** que pode causar problemas éticos e legais, especialmente nos campos onde direitos eram previamente cedidos à humanos (Balkin, 2018). Assim, questões tais como "se algoritmos poderiam ser protegidos pela liberdade de expressão" ou "artefatos produzidos por IA serem protegidos por leis de direitos autorais" são exemplos comuns (Jia e Zhang, 2021).

Grewal et al. (2021) apontam que, quanto mais a solução de IA não for claramente explicada e parecer uma "caixa preta", maior a probabilidade de gerar níveis baixos de confiança e adesão para sua adoção e engajamento. Ainda na perspectiva do viés negativo que o desenvolvimento ou aplicação da IA pode assumir, observa-se que seu potencial impacto se estende a diversos contextos. Em um contexto social, Zanzotto (2019) defende que a revolução da IA pode ter um lado sombrio relacionado a crescentes níveis de desemprego que podem preceder uma transformação imprevisível no mercado de trabalho. Já Grewal et al. (2021) observam que o uso da IA também aumenta as preocupações com preconceitos. No ambiente da gestão das organizações, as atitudes e intenções gerenciais em utilizar a IA para tomada de decisão podem ser afetadas tanto pelos benefícios quanto pelos riscos associados ao seu uso (Cao et al., 2021), o que pode afetar a gestão estratégica das empresas já que, de acordo com Quehaja et al. (2017), a competição global baseada no conhecimento criou a necessidade de estratégias intencionais de processos de tomada de decisão eficazes.

Outro campo em que o lado sombrio da IA também começa a ser discutido é no ambiente de trabalho. Craig et al. (2019) observam que, considerando as interações sociais entre indivíduos e tecnologia, é importante entender o fenômeno da possibilidade da TI ameaçar a identidade de trabalhadores no ambiente de trabalho. Para Mirbabaie et al. (2021), a introdução da IA pode impactar negativamente a identificação dos trabalhadores com seus trabalhos, já que pode ser esperado que a IA possa mudar fundamentalmente os ambientes de trabalho e as profissões, alimentando o medo dos indivíduos de serem substituídos e podendo ainda afetar seu comportamento frente às suas atividades e obrigações

Os resultados adversos do uso da IA, em paralelo com a ameaça da perda de controle ou autonomia sobre entidades superiores de IA, desencadearam um debate contínuo sobre a necessidade de se estabelecer um conjunto de princípios para efetivamente governar a IA (Barredo Arrieta et al., 2020; Fjeld et al., 2020). Observa-se, na literatura, que enquanto alguns pesquisadores começam a entender as crescentes regulamentações relativas à IA já implementadas ou ainda em desenvolvimento, existe também uma perspectiva de se estabelecer uma noção mais conceitual através de diretrizes básicas para o desenvolvimento de IA que evite a manifestação de seu lado sombrio. Theodorou e Dignum (2020) observam que há um interesse crescente na noção de IA "Responsável" como um conjunto de proposições ou declarações normativas sobre como a IA geralmente deve ser desenvolvida,

implantada e governada. Garantir que consequências prejudiciais e/ou não intencionais sejam minimizadas, ou não ocorram durante a vida útil dos projetos de IA, requer uma compreensão abrangente do papel desses princípios responsáveis durante o projeto, implementação e manutenção de aplicativos de IA (Mikalef et al., 2022).

Considerando o potencial transformador de realidade da IA e as diferentes técnicas e tecnologias digitais envolvidas, vários órgãos públicos e empresas privadas lançaram iniciativas regulatórias ou "autorregulatórias" para o seu controle e melhorias (Hagendorff, 2020; Jobin et al., 2019). Segundo González-Esteban e Calvo (2022), essas propostas estão relacionadas ao desenvolvimento de quadros legislativos para reger a concepção e os impactos da IA; códigos de ética, conduta e boas práticas para orientar a prática profissional específica; comitês de ética para tratar da resolução de conflitos relacionados ao uso da IA por meio do diálogo e deliberação; e relatórios de prestação de contas para melhorar a transparência e a analisar impactos econômicos, sociais e ambientais da IA.

Em relação às regulamentações existentes, na inexistência de leis mais rígidas, que seriam regulamentos juridicamente embasados para definir condutas permitidas ou proibidas, os *stakeholders* públicos e privados evocam preocupações relativas ao lado sombrio da IA promovendo "*soft laws*" na forma de diretrizes éticas normativas, com o objetivo de restringir o lado negativo enquanto se preserva o incentivo à inovação (Jia e Zhang, 2021; Calo, 2017; Cath et al., 2018). Mais especificamente, Jia e Zhang (2021) identificaram, através do banco de dados "*AlgorithmWatch*", que existem mais de 160 diretrizes éticas para IA globalmente nos últimos 5 anos (em relação à data de publicação do estudo, em 2021).

Um fator a ser considerado ao se analisar o contexto do desenvolvimento e evolução da IA (e que pode afetar outras questões tecnológicas contemporâneas), é o quão dinâmico, intenso e veloz pode ser o processo. Para González-Esteban e Calvo (2022), são necessárias revisões do quadro jurídico-político, na tentativa de que as políticas e diretrizes regulamentadoras não se tornem obsoletas. O problema, na visão dos mesmos autores, é que por mais que se tente diminuir a "distância" entre a detecção das mudanças e seus impactos, e a revisão do quadro legislativo, sempre há um hiato temporal (entre as mudanças tecnológicas e a consequente adaptação normativa), com consequências difíceis de controlar.

Considerando a dificuldade de acompanhamento e regulamentação do desenvolvimento e uso responsável de IA, González-Esteban e Calvo (2022) destacam que, dentre as iniciativas do mercado ou da sociedade civil, existem propostas de autorregulamentação com base em diretrizes e códigos de ética, conduta e melhores práticas para IA. No entanto, os autores defendem que a estrutura de autorregulamentação também sofre críticas, já que boa parte pode parecer não estar ligada à ética, mas sim a uma estratégia de evitar a regulamentação estatal da IA que limitasse os benefícios de sua aplicação prática.

Jia e Zhang (2021) argumentam que, apesar dos méritos das *soft laws,* em termos de agilidade e flexibilidade, críticas teóricas e empíricas têm sido propostas quanto à sua eficácia em orientar a conduta dos *stakeholders*. Alguns pesquisadores questionam as motivações por trás dessas diretrizes, especialmente aquelas apoiadas por *stakeholders* do setor privado, pois poderiam ser usadas como disfarce para tornar técnico um problema social, ou ainda para desencorajar os esforços de imposição de encargos regulatórios legais/jurídicos (Benkler, 2019; Jia e Zhang, 2021). De acordo com González-Esteban e Calvo (2022), apesar do crescente interesse das diversas iniciativas de órgãos governamentais e corporações, não há

dúvida de que ainda há um longo caminho a percorrer para alcançar o controle e a orientação adequados sobre o projeto e a aplicação da IA em diferentes campos de atividade.

3. Método de Pesquisa

Com a intenção de construir uma revisão integrada das publicações relativas à perspectiva de lado sombrio da Inteligência Artificial na literatura, optou-se pela estruturação de uma Revisão Sistemática da Literatura (RSL). O objetivo deste trabalho não é a indicação específica de um modelo ou abordagem para se tratar a questão sombria relativa a IA, mas sim identificar e revisar as publicações existentes para contextualizar a questão na literatura atual. Neste caso, uma RSL pode ser especialmente útil para identificar o conhecimento existente sobre um tópico, incluindo-se os possíveis "gaps", através da busca por estudos relevantes com uso de palavras chaves em base de dados científicas (Fink, 2013; Webster e Watson, 2002). Segundo Kraus et al. (2020), a RSL oferece a possibilidade de combinar diferentes de literatura criando sólidas fundações para desenvolvimento de pesquisas futuras se alinhando, assim, ao objetivo deste trabalho.

Com base nas diretrizes da formulação de uma RSL de Kraus et al. (2020), primeiramente foram definidos os objetivos da busca em relação à problemática e às questões levantadas no início deste trabalho, sintetizadas pelas palavras-chave que permitem conduzir a pesquisa nas bases de dados abrangendo as publicações relativas ao tema. A figura 1 sintetiza as etapas de busca e resultados da seleção de publicações.

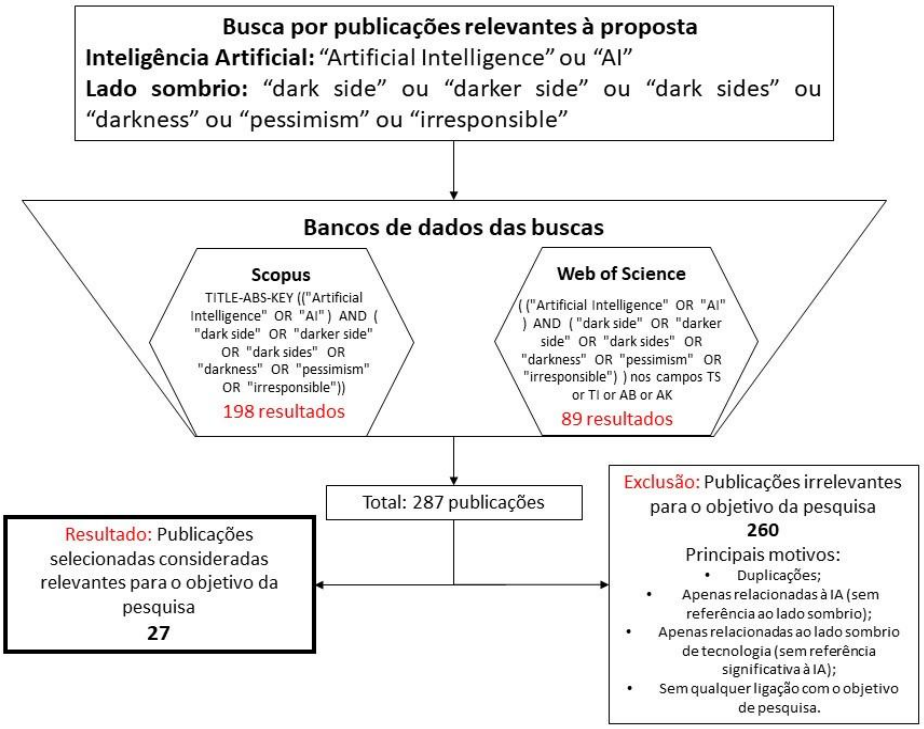

Figura 1- Fluxograma da RSL - busca e seleção de publicações

A busca original buscou filtrar trabalhos unindo um tema específico com uma perspectiva: a Inteligência Artificial e seu lado "sombrio". Assim, definiram-se as palavras-chave para encontrar publicações que unissem essas duas características. Em relação à Inteligência Artificial, dada sua difusão enquanto tema e conceito, utilizaram-se as palavras "Artificial Intelligence" e "AI" como modo de filtro das publicações que contemplassem o

tópico. Já em relação à visão sombria, considerando a possibilidade de que a temática fosse recente e a nominação específica, buscou-se a ligação do tema da IA e sua concatenação com terminologias que considerassem o possível viés negativo: “dark side”, “darker side”, “dark sides”, “darkness”, “pessimism” or "irresponsible".

O estudo utilizou, para a busca, as bases de dados de publicações científicas: *Scopus* e *Web of Science*. Na *Scopus*, considerando os termos anteriormente destacados, foram identificados 198 resultados para a busca, e na *Web of Science* 89 resultados, totalizando, assim, 287 publicações encontradas com características de interesse para esta pesquisa. Em termos de evolução da temática, percebe-se que mais da metade dos resultados foram publicados nos últimos 5 anos (aproximadamente 55% das publicações feitas a partir de 2017), com predomínio de artigos entre as publicações (aproximadamente 47%) e tendo os Estados Unidos e China como principal origem dos trabalhos.

Com base nos resultados observados, iniciou-se uma primeira análise das publicações a fim de selecionar as que se alinhavam ao objetivo deste trabalho. Esta etapa foi realizada através da abertura do documento e revisão básica de título, resumo e introdução, para uma filtragem prévia de resultados pertinentes. Nesta etapa foram realizadas exclusões com base em critérios recorrentes, como: duplicações (em relação às duas bases de dados utilizados); publicações que abordavam o tema da Inteligência Artificial mas que não entravam, de fato, na questão do lado sombrio da mesma; publicações que tratavam de alguma maneira do lado sombrio de aspectos tecnológicos mas sem referência significativa à IA; ou publicações sem qualquer ligação com o objetivo da pesquisa. Após a primeira análise, realizou-se uma segunda etapa considerando os resultados previamente filtrados. Foram aplicados os mesmos critérios de exclusão, mas desta vez aprofundando a leitura no texto da publicação, o que resultou em 27 publicações selecionadas para análise. Importante pontuar que não foi possível ter acesso ao texto integral de 2 publicações selecionadas (dentre as 27), devido ao acesso restrito das mesmas. As publicações tratavam-se, porém, de estudos ligados à visão sombria da IA, e suas temáticas e abordagens serão contextualizadas nos dados analisados.

4. Análise e discussões

Com base na leitura detalhada das publicações selecionadas realizou-se uma análise de modo a estabelecer evidências para identificar qual o estado da arte referente à perspectiva do lado sombrio da Inteligência Artificial. Ao analisar as publicações selecionadas, observa-se que há um alinhamento com os argumentos levantados por outros autores na introdução deste trabalho, no que se refere ao estudo do lado sombrio da Inteligência Artificial ser um tópico recente na literatura acadêmica e que vem atraindo olhares nos últimos anos. Mesmo não havendo restrição de data inicial de busca nos filtros estabelecidos, observou-se que nenhuma das publicações selecionadas são anteriores a 2017.

Em relação às 27 publicações selecionadas - e desconsiderando-se o ano presente (2022) - é possível observar que o número de publicações aumentou ano a ano, tendo sido 3 publicações de 2019, 4 de 2020 e 16 publicações de 2021. Assim, quase 60% de todas as publicações selecionadas foram realizadas no último ano (Figura 2), o que reforça a contemporaneidade do tópico. O ano de 2022, ainda parcial pelo fato desta pesquisa estar sendo realizada no início do mês de abril, já demonstra-se igualmente representativo.

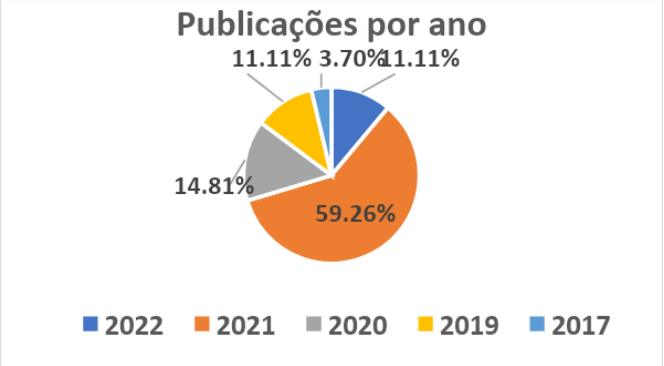

Figura 2 - Distribuição dos estudos selecionados por ano de publicação

Numa visão mais micro, relativa ao período dessas publicações, e com o uso da ferramenta *VOSviewer* para analisar as palavras-chave mais utilizadas nas 27 publicações selecionadas, é possível observar um aumento na intensidade das relações do tópico do "lado sombrio" no final do período, entre os anos com mais publicações (2020 e 2021). Reconhecendo que se trata de uma amostra significativamente pequena, porém, para uma análise de recorrência de palavras chave, a Figura 3 ilustra as relações mais recorrentes e recentes entre os tópicos que associaram o lado sombrio, para além da Inteligência Artificial, com termos como "*big data analytics*", "*advanced analytics*" e "*business-to-business*".

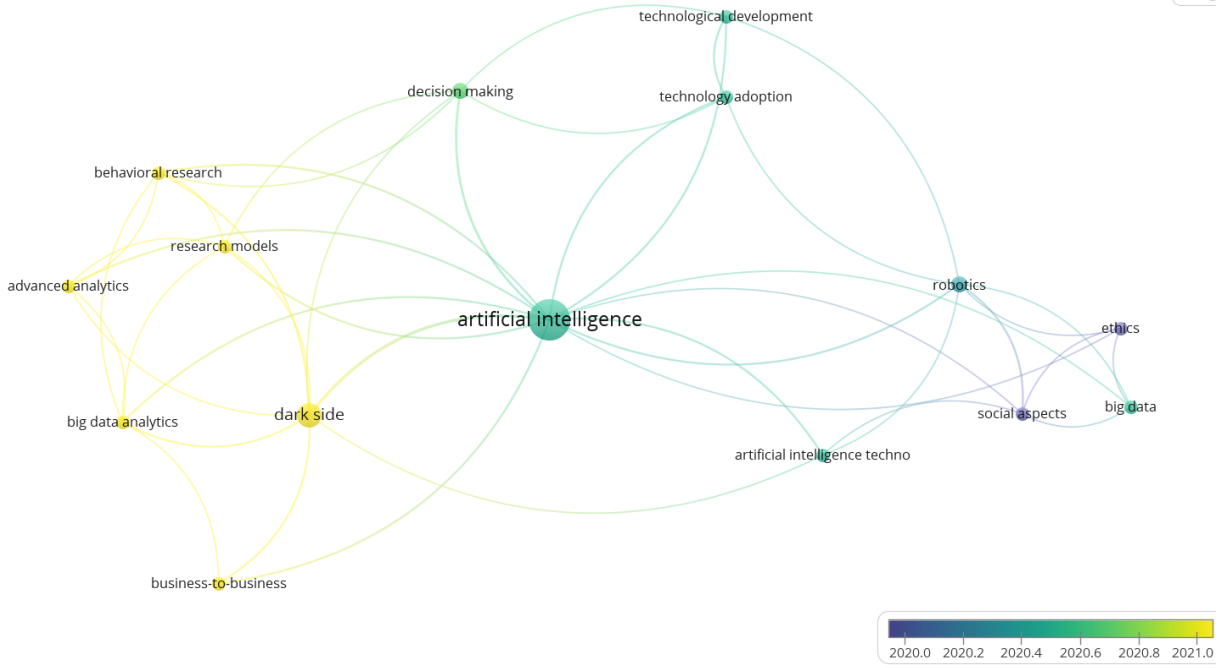

Figura 3 - Palavras-chave dos estudos selecionados no período de maior número de publicações

Os trabalhos selecionados também apresentam diversidade em relação ao país de origem, sendo as 27 publicações provenientes de 15 países diferentes (considera-se que uma mesma publicação pode ser escrita por autores de diferentes países). Os países responsáveis pelo maior número de publicações podem ser identificados conforme apresentado na Figura 4.

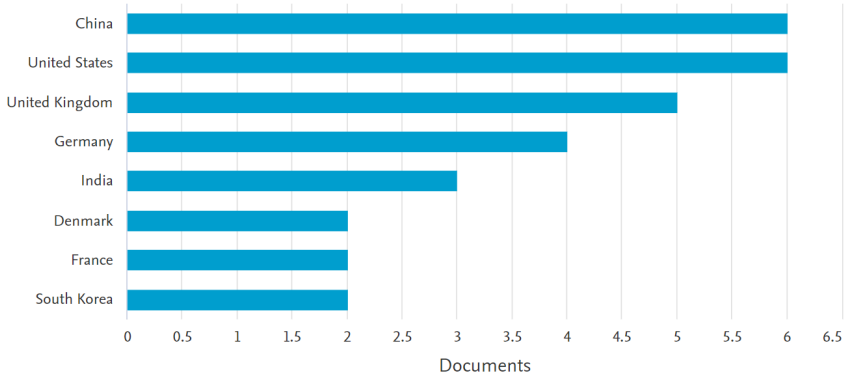

Figura 4- Número de publicações, dentre as selecionadas, provenientes de cada país

Embora a análise possibilite identificar que a visão do lado sombrio da IA é um tópico recente, e que possivelmente vem chamando a atenção de pesquisadores dado o número crescente de publicações a cada ano, pode-se dizer que ainda é uma questão incipiente e pouco abordada, considerando o número absoluto de publicações. Além disso, foram selecionadas publicações para além das que tratam apenas do lado sombrio da IA (em alguns dos estudos essa era uma questão presente, mas adjacente a outro assunto) e os trabalhos ainda apresentam uma fase inicial de amadurecimento, conforme evidenciado na Tabela 1.

Título	**Autores(as)**	**Ano**	**Revista/Conf**	**Contexto/Área**
Ethically governing artificial intelligence in the field of scientific research and innovation	González-Esteba, Calvo	2022	Heliyon Journal	IA na Pesquisa Científica e na Inovação
The Dark Sides of AI	Cheng, Lin, Shen, Zarifs, Mou	2022	Electronic Markets	Mercados eletrônicos
Thinking responsibly about responsible AI and 'the dark side' of AI	Mikalef, Conboy, Lundström & Popovič	2022	European Journal of Information Systems	Revisão teórica sobre o lado sombrio da IA
Artificial intelligence: The light and the darkness	Grewal, Guha, Satornino, Schweige	2021	Journal of Business Research	Revisão teórica sobre o lado sombrio da IA em empresas B2B e B2C
Theorizing the dark side of business-to-business relationships in the era of AI, big data, and blockchain	Gligor; Pillai; Golgeci.	2021	Journal of Business Research	Revisão teórica sobre o lado sombrio dos relacionamentos B2B na era de IA, big data e blockchain
Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making	Cao, Duan, Edwards, Dwivedi	2021	Technovation	Análise teórica sobre a motivação de gestores para adotar ou não IA na tomada de decisão organizacional
A two-dimensional research framework for analysing dark side of AI	Zeng ; Wu	2021	2nd International Conference on Computer Engineering and Application	Revisão teórica sobre o lado sombrio da IA

The bright and dark sides of artificial intelligence: A futures perspective on tourist destination experiences	Grundner; Neuhofer.	2021	Journal of Destination Marketing and Management	Turismo e viagens
A framework for component selection considering dark sides of artificial intelligence: a case study on autonomous vehicle	Jabbarpour, Saghiri, Sookhak	2021	Electronics (Journal)	Sistemas inteligentes
When more is less: The other side of artificial intelligence recommendation	Chen; Qiu; Zhao; Han; He; Siponen; Mou; Xiao.	2021	Journal of Management Science and Engineering	Recomendações automatizadas para consumidores baseadas em IA
The rise of artificial intelligence – understanding the AI identity threat at the workplace	Mirbabaie, Brünker, Möllmann, Stieglitz	2021	Eletronic Markets	Ameaça de identidade dos trabalhadores pela adoção de TI
AI invading the workplace: negative emotions towards the organizational use of personal virtual assistants	Hornung, Smolnik	2021	Electronic Markets	IA no suporte pessoal virtual ao usuário no ambiente de trabalho
The dark sides of AI personal assistant: effects of service failure on user continuance intention	Sun, Li, Yu	2021	Electronic Markets	IA no suporte pessoal virtual ao usuário
Understanding dark side of artificial intelligence (AI) integrated business analytics	Rana, Chatterjee, Dwivedi, Akter	2021	European Journal of Information Systems	Análises de negócios integrada por IA (AI integrated business analytics (AI-BA))
Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines	Jia, Zhang	2021	Electronic Markets	Analisa as diretrizes normativas destinadas a regular o desenvolvimento e a aplicação de IA
The dark sides of people analytics: reviewing the perils for organisations and employees	Giermindla, Strichb, Christa, Leicht-Deobald, Redzep	2021	Journal of Information Systems	Uso de IA e "analytics" na otimização da gestão de recursos humanos
The dark side of Industrial Revolution 4.0 - Implications and suggestions	Memon, Ooi	2021	Academy of Entrepreneurship Journal	Análise teórica sobre o lado sombrio da "Revolução Industrial 4.0"
The good, the bad, and the ugly: impact of analytics and artificial intelligence-enabled personal information collection on privacy and participation in ridesharing	Cheng, Su, Luo, Benitez, Cai	2021	European Journal of Information Systems	Plataformas de compartilhamento de carona
The dark side of AI-powered service interactions: exploring the process of co-destruction from the customer perspective	Castillo, Canhoto, Said	2021	The Service Industries Journal	Atendimento a cliente por "chatbots" operados por IA

From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence	Liu, Zawieska	2020	Ethics and Information Technology	Robótica
Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence	Jensen, Whyte, Cuomo	2020	International Studies Review	Militar/Guerra
The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration	Wirtz,Weyerer, Sturm	2020	International Journal of Public Administration	Administração pública
Politics of technology or technology of politics?	Scavlia; Stockman.	2020	European Conference on the Impact of AI and Robotics	Política/Democracia
Scary dark side of Artificial Intelligence: A perilous contrivance to mankind	Kumar, Singh, Vvek Bhatanagar, Jyoti	2019	Humanities and Social sciences Reviews	Lado sombrio da IA com base em experimentos com Super IA do MIT
On the Implications of Artificial Intelligence and its Responsible Growth	Devaraj, Makhija, Basak	2019	Journal of Schientometric Research	Contextualização teórica de IA que aborda, também, o lado sombrio
Viewpoint: Human-in-the-loop artificial intelligence	Zanzotto	2019	Journal of AI Research	Mercado de trabalho
Long-term trends in the public perception of AI	Fast; Horvitz	2017	31st AAAI Conference on AI	Percepção pública sobre IA

Tabela 1 - Publicações selecionadas e analisadas neste trabalho

Embora diversos autores venham teorizando sobre a crescente importância de se olhar para o possível lado sombrio da IA - conforme destacado na introdução e problematização deste trabalho - observa-se que os estudos dedicados prioritariamente ou exclusivamente a esse fim ainda são escassos, apesar de existirem indícios de aumento de publicações nesse contexto, como analisado anteriormente. A maior parte dos estudos existentes relativos ao tópico, e selecionados neste trabalho, abordam o lado sombrio de maneira mais adjacente, no contexto de algum outro estudo central.

Das 27 publicações analisadas, apenas 7 (aproximadamente 26%) têm o lado sombrio da IA como assunto central na pesquisa. Outras 3 abordam o lado sombrio em um paralelo ao lado "positivo" da Inteligência Artificial, enquanto que 17 delas tratam do lado sombrio da IA como um assunto adjacente ao conteúdo principal do estudo. Estes últimos - embora não tratem do assunto de maneira central – foram considerados nesta revisão por dois motivos: primeiro, pois abordam o lado sombrio da IA sob alguma perspectiva , mesmo em contexto específico, podendo trazer abordagens e perspectivas importantes para se levar a outras áreas ou formar um modelo geral; segundo, pois o presente estudo pode servir como uma referência futura de trabalhos que abordaram o tema – seja de maneira mais conceitual ou mais prática, em um nicho específico. O fato de as publicações mais focadas na abordagem do lado sombrio da IA serem mais recentes pode indicar uma convergência para construções teóricas baseadas na argumentação da problemática que muitos dos autores destacam em seus estudos.

Com base nas publicações analisadas, observa-se um "desequilíbrio" de intensidade entre a importância de se aprofundarem as pesquisas sobre o lado sombrio da Inteligência Artificial (descritas nas motivações e conceitualizações teóricas de diversos estudos) e as publicações existentes que abordam o tema. A maioria dos estudos analisados ainda aborda a questão de maneira adjacente em contextos específicos, indicando um esforço de diagnóstico e entendimento da importância do tema, mas ainda sem profundidade em seu tratamento. Dentre todos os estudos analisados, aproximadamente metade deles apresenta algum tipo de modelo – de diversas naturezas – conforme é evidenciado na Tabela 2.

Autor/ Ano	**Principais achados/contribuições**	**Modelo proposto?**
González-Esteban e Calvo (2022)	Analisa o contexto atual do uso de IA na pesquisa científica e propõe diretrizes para desenvolvimento e implementação de um sistema de governança de IA	Propõe um sistema de governança ético e responsável para IA na Pesquisa Científica, centros de inovação e agências de financiamento da ciência
Cheng et al. (2022)	Breve revisão e contextualização em Prefácio de publicação de revista em edição especial sobre o lado sombrio da IA (apresenta 6 artigos sobre o tema)	
Mikalef et al. (2022)	Com a adoção de uma lente do "lado escuro" da IA, o estudo teoriza sobre a noção de "IA Responsável", identificando as possíveis maneiras em que IA pode produzir consequências não esperadas e sugere caminhos futuros de pesquisas em Sistemas de Informação para aprimorar o conhecimento na área	Define princípios básicos do que configuraria a "Inteligência Artificial Responsável" e propõe diversos potenciais caminhos para seu aprofundamento
Grewal et al. (2021)	Para além do lado "positivo" da IA, os autores identificam os impulsionadores do lado sombrio da IA, contextualizando em empresas B2B e B2C	Propõe um modelo para entender tanto o lado positivo quanto negativo de IA em empresas com configuração B2C e B2B
Gligor et al. (2021)	Examina como as novas tecnologias emergentes potencialmente transformam os relacionamentos B2B e podem conduzir a consequências de lado sombrio - ilustra teorias que podem fornecer novos insights ao explorar o lado sombrio dos relacionamentos B2B em sua relação com o contexto tecnológico atual	
Guangming et al. (2021)	Embora IA apareça no estudo como adjacente no seu papel de suporte gestão e tomada de decisão, as hipóteses levantadas trazem considerações sobre consequências inesperadas de sua aplicação	
Zeng e Wu (2021)	Revisão de literatura do lado sombrio da IA, com proposta de um modelo bidimensional para abordar o lado sombrio de aplicações emergentes de tecnologia com IA; destaca áreas para futuras pesquisas	Propõe um modelo bidimensional para abordar o lado sombrio de aplicações emergentes de tecnologia com IA
Grundner e Neuhofer (2021)	Sob a lente teórica da lógica "Service-dominant" (S-D), o estudo analisa os aspectos positivos e sombrios da IA contextualizando-os no setor do turismo e propõe um modelo teórico	Propõe um modelo teórico de avaliação dos aspectos positivos e sombrios da aplicação de IA no contexto do Turismo
Jabbarpour et al. (2021)	O estudo analisa o lado sombrio da IA no processo de seleção de componentes de sistemas inteligentes e propõe um novo modelo para esse processo	Proposta de um novo modelo de quatro etapas para a seleção de componentes de sistemas inteligentes que considera os potenciais lados sombrios da IA
Chen et al. (2021)	O estudo analisa as possibilidades de lado sombrio da IA na recomendação de conteúdo para consumidores e propõe evidências empíricas para a possível regulamentação de IA	

	nesse contexto	
Mirbabaie et al. (2021)	O estudo traz o entendimento da possível sensação de "ameaça" de identidade pela IA no ambiente de trabalho, levanta maneiras de identificar o acontecimento do fenômeno e aprofunda a discussão da colaboração com a IA no local de trabalho	O estudo propõe um modelo teórico para identificar preditores da ameaça de identidade pela IA.
Hornung e Smolnik (2021)	O estudo identifica e categoriza emoções de funcionários associadas ao uso de assistentes/suporte virtuais (Personal Virtual Assistants -PVAs) baseados em IA no ambiente de trabalho	
Sun et al. (2021)	Através de aplicação e análise empírica de um questionário, o estudo propõe demonstrar como a experiência negativa do usuário com a Assistência Pessoal por meio de IA pode levar a respostas psicológicas negativas dos consumidores por meio da exaustão e carga cognitiva	Da perspectiva de "technostress", o estudo propõe um modelo teórico para os consumidores lidarem com fontes de pressão por falha de serviço
Nripendra et al. (2021)	Com base na visão baseada em recursos, visão de capacidade dinâmica e teoria de contingência, o modelo proposto captura os componentes e efeitos de uma opacidade AI-BA no ambiente de risco das empresas e o impacto nas suas vantagens competitivas. O estudo indica a falta de governança, pobreza de qualidade de dados e ineficiência de treinamento de pessoal como fatores chave que levariam a opacidade de AI-BA.	Propõe um modelo de pesquisa, baseado em teorias, para o contexto do trabalho (opacidade de AI-BA)
Jia e Zhang (2021)	Usando um modelo baseado na gestão de riscos, o estudo analisa os riscos de IA e as correspondentes medidas de governança (baseado nas diretrizes normativas existentes). Conclui que os riscos seguem bastante subestimados e propõe futuras melhorias	Desenvolve um modelo teórico baseado na gestão de riscos para analisar as diretrizes normativas existentes de regulamentação de IA
Giermindla et al. (2021)	O estudo analisa o contexto da gestão de pessoas, o uso de "people analytics" e as tecnologias atuais e futuras para identificar seis perigos dessa interação, e sugere direções para futuras pesquisas	
Memon e Ooi (2021)	O estudo analisa o fenômeno da "Revolução industrial 4.0" e teoriza sobre o lado sombrio de seus principais componentes (dentre eles, a IA)	
Cheng et al. (2021)	Através de pesquisa e entrevistas com usuários, o estudo identifica o lado sombrio da IA e BDA em plataformas de compartilhamento de carona (incertezas e invasão de privacidade, por exemplo).	
Castillo et al. (2021)	O estudo aborda o lado negativo da relação de clientes/usuários com atendimento automatizado por IA, considerando problemas de autenticidade, desafios cognitivos, conflitos de integração, entre outros.	Modelo conceitual de "co-destruição" relativo ao atendimento automatizado de clientes por tecnologias suportadas por IA
Liu e Zawieska (2020)	O artigo analisa o impulso da robótica responsável propondo um conjunto complementar de direitos humanos direcionados especialmente para os possíveis efeitos negativos (lado sombrio) decorrentes das tecnologias de Robótica e Inteligência Artificial	
Jensen et al. (2020)	Artigo não disponível	Artigo não disponível
Wirtz et al. (2020)	Artigo não disponível	Artigo não disponível

Scavlia e Stockman (2020)	Aborda a interação de política com tecnologia e teoriza sobre a possibilidade de a IA, nesse contexto, ser corrompida e ameaçar as estruturas democráticas	
Kumar et al. (2019)	O estudo analisa as possibilidade de aplicação do lado sombrio de IA com base em experimentos do MIT com "super AI", argumentando que a IA pode ser "programada" em um viés negativo e discorre sobre possíveis consequências pessimistas	
Devaraj et al. (2019)	O estudo faz uma contextualização teórica do histórico da IA e e aborda parcialmente o lado sombrio especialmente no que se refere a ética e responsabilidade de uso	
Zanzotto (2019)	O estudo aborda o futuro com o crescimento da IA de maneira pessimista em relação ao mercado de trabalho, sugerindo que o uso da tecnologia associada banco de dados estaria "roubando" o conhecimento de trabalhadores para automação, e que dessa forma eles estariam dando um "tiro no próprio pé"	Reconhecendo que sistemas de IA contemplam humanos, o estudo propõe o "Human-in-the-loop Artificial Intelligence (HitAI)" como um paradigma mais justo para sistemas de IA.
Fast e Horvitz (2017)	O estudo analisa as visões relativas a IA expressas em um grande jornal norte-americano durante 30 anos para sintetizar a perspectiva dessas visões, sendo otimistas ou pessimistas. Identifica um grande crescimento na discussão sobre o tópico desde 2009, predominância de uma visão otimista, mas um crescimento da visão pessimista nos últimos anos (descreve os tópicos mais recorrentes da visão sombria, principalmente a perda de controle sobre IA e preocupações éticas)	

Tabela 2 - Detalhamento das publicações selecionadas e analisadas neste trabalho

A maior parte dos modelos contidos nos estudos analisados se referem a contextos específicos relativos a algum negócio ou área de estudo. Dentre os diversos contextos ou áreas de atuação dos estudos sem foco específico no lado sombrio da IA, é possível citar: lado sombrio da IA na pesquisa científica e inovação, nos mercados eletrônicos, nas organizações B2B, no turismo, em sistemas inteligentes, no ambiente de trabalho, na gestão de recursos humanos, no suporte virtual a usuários, na análise de negócios integradas por IA, na robótica, no contexto militar, na administração pública, entre outros. Em relação a modelos mais conceituais e de visão abrangente do lado sombrio da IA, o trabalho de Jia e Zhang (2021) propõe um modelo teórico baseado na gestão de riscos para analisar as diretrizes normativas existentes de regulamentação de IA, e Mikalef et al. (2022) destacam um modelo que define princípios básicos associados à "Inteligência Artificial Responsável".

Nesse contexto, a partir da análise das publicações sobre o lado sombrio da IA, e também na proposição de modelos, existem ao menos duas distintas perspectivas no processo: alguns autores focam seus estudos nas regulamentações normativas específicas, enquanto outros direcionam seus estudos à construção de um conceito mais abrangente e teórico das diretrizes para a IA responsável, íntegra e ética. Observa-se que ainda não existe maturidade teórica ou consenso na abordagem da questão do lado sombrio da IA, tanto no que se refere à estruturação teórica e conceitual, como também na construção de modelos de qualquer natureza (sejam de diagnóstico do fenômeno, tratamento ou prevenção). Os poucos estudos relativos ao tema ainda são recentes e sem definição de modelos que possam ser aplicados e/ou reproduzidos em diferentes contextos, sem uma consolidação das abordagens usadas em áreas específicas em um modelo mais genérico e potencialmente reproduzível.

5. Conclusões

Com o objetivo de analisar qual é o estado da arte relativo ao lado sombrio da Inteligência Artificial, este estudo realizou uma pesquisa utilizando as bases de dados Scopus e Web of Science como fonte de busca por publicações pertinentes relativas ao tópico. Foi desenvolvida uma Revisão Sistemática da Literatura com objetivo de, seguindo às associações de Loureiro et al. (2019), identificar, escolher e analisar criticamente pesquisas relevantes e, assim, buscar por *insights* baseados em resultados de diversos pesquisadores sobre o conhecimento já estabelecido em pesquisas anteriores.

A busca inicial – a qual resultou em 287 publicações – passou por diferentes etapas de filtragem e seleção, seguindo critérios específicos de inclusão e exclusão, resultando em um conjunto final de 27 publicações, as quais foram organizadas e estruturadas (Tabelas 1 e 2). Com a análise das características principais da amostra, e dentro do objetivo de consolidar os dados, estas tabelas apresentam os principais autores, *journals/conferences*, contextos de aplicação e modelos, o que pode favorecer futuros trabalhos que abordem o tema.

A condução da revisão e sua análise foram balizadas pela pergunta de pesquisa deste estudo resultando em aspectos genéricos importantes para o entendimento da temática pesquisada. Considerando que o lado sombrio dos aspectos de Inteligência Artificial é um tema relativamente recente, e que aos poucos vem recebendo maior atenção dos pesquisadores, buscou-se contextualizar o que já se sabe e se produziu em termos de publicações sobre o assunto, dada essa contemporaneidade do tema.

Observou-se que, de fato, o tema é recente, e que mais de 70% das publicações analisadas foram desenvolvidas nos dois últimos anos, o que sugere um crescimento de interesse relacionado ao tópico pela crescente disposição e atenção por parte dos pesquisadores. Também foi possível analisar que, embora crescente em termos quantitativos, ainda existe um baixo volume de publicações que abordam o lado sombrio da IA de maneira direta e central. De fato, a maior parte dos estudos analisados contemplam o tema de maneira adjacente a outro assunto ou negócio específico, e como esses elementos se integram. A análise também evidencia que, conforme argumentado pelos autores da revisão teórica, os estudos sobre o tema ainda se encontram em estágio inicial de maturidade, sendo poucos os modelos propostos para diagnóstico, prevenção ou tratamento desse fenômeno até o momento

Por fim, este trabalho contribui de maneira preambular para organizar os estudos realizados relativos ao lado sombrio da IA, identificando como se estruturam e o que existem proposto até o momento. Com isto, os achados podem servir de ponto de partida para estudos futuros que busquem preencher os *gaps* ainda existentes, bem como identificar novas problemáticas e aprofundar os tópicos que começaram a ser analisados, integrando às abordagens realizadas alguns desses diferentes trabalhos. Observa-se uma variedade de possibilidades para futuras pesquisas em relação ao tema: abordar a perspectiva da lente sombria em relação à IA em diferentes tipos de negócio ou áreas de atuação para identificar especificidades de cada negócio; pode ser interessante olhar para as pesquisas existentes buscando integrá-las em frameworks conceituais para entender como ocorre e se seria possível propor modelos e "testá-los"; entender como o lado sombrio da IA afeta e se relaciona com a gestão estratégica das empresas, mais especificamente, no sentido da motivação ou resistência gerencial em adotar a tecnologia, por exemplo, também é um contexto de estudo que poderia ser aprofundado; Enfim, observa-se que, embora ainda em estágio inicial, baseado na análise dos estudos realizada identificam-se duas linhas conceituais

para "tratar" o fenômeno, sendo o estabelecimento de um modelo teórico integrativo como a "IA Responsável" – que possui alguns pilares fundamentais – e também a análise das regulamentações públicas e privadas, que ainda estão em estágio inicial, e que buscam regulamentar o desenvolvimento e o uso da IA. Analisar como essas abordagens se relacionam e potencialmente se integram pode ser um caminho importante a se contemplar no aprofundamento do estudo do tema.

6. Referências

ALT, Rainer. Electronic markets and current general research. **Electronic Markets**, 28(2), p. 123–128. 2018.

BALKIN, Jack. M. Free Speech is a Triangle. **Columbia Law Review**, 118(7), p. 2011–2056. 2018.

BENKLER, Yochai. Don't let industry write the rules for AI. **Nature**, 569, p. 161. 2019.

CALO, Ryan. Artifcial Intelligence policy: a primer and roadmap. **UCDL Review**, 51, p. 399. 2017.

CALVO, Patrici. The ethics of Smart City (EoSC): moral implications of hyperconnectivity, algorithmization and the datafication of urban digital society. **Ethics Inf. Technol**. (22), p. 141–149. 2020.

CAO, Guangming; DUAN, Yanqing; EDWAARDS, John S; DWIVEDI, Yogesh K. Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making. **Technovation**, 106, art. no. 102312. 2021.

CASTILLO, Daniela; CANHOTO, Ana I; SAID, Emanuel. The dark side of AI-powered service interactions: Exploring the process of co-destruction from the customer perspective. **Service Industries Journal**. 2020.

CATH, Corinne, WALTCHER, Sandra, MITTELSTADT, Brent, TADDEO, Mariarosaria, & FLORIDI, Luciano Artifcial intelligence and the 'good society': the US, EU, and UK approach. **Science and Engineering Ethics**, 24(2), p. 505–528. 2018.

CHEN, Sihua; QIU, Han; ZHAO, Shifei; HAN, Yuyu; HE, Wei; SIPONEN, Mikko; MOU, Jian., XIAO, Hua. When more is less: The other side of artificial intelligence recommendation. **Journal of Management Science and Engineering**. 2021.

CHENG, Xusen; SU, Linlin; LUO, Xin R; BENITEZ, Jose, CAI, Shun. The good, the bad, and the ugly: impact of analytics and artificial intelligence-enabled personal information collection on privacy and participation in ridesharing. **European Journal of Information Systems**. 2021.

CHENG, Xusen, LIN, Xiao, SHEN, Xiao-Leng, ZAFIRIS, Alex. MOU, Jian. The dark sides of AI. **Electron Markets**. 2022.

CRAIG, Kevin, THATCHER, Jason B.., GROVER, Varun. The IT Identity Threat: A Conceptual Definition and Operational Measure. **Journal of Management Information Systems**, 36(1), p. 259–288. 2019.

DEVARAJ, Harsha, MAKHIJA, Simran, BASAK, Suryoday. On the Implications of Artificial Intelligence and its Responsible Growth. **Journal of Scientometric Research**,8,2s,s2-s6. 2019.

FAST, Ethan, HORVITZ, Eric. Long-term trends in the public perception of artificial intelligence. AAAI'17: Proceedings of the **Thirty-First AAAI Conference on Artificial Intelligence**. p. 963–969. 2017.

FINK, Arlene. Conducting Research Literature Reviews: from the Internet to Paper. Sage Publications. 2013.

FLORIDI, Luciano, COWLS, Josh, BELTRAMETTI, Monica, CHALITA, Raja, CHAZERAND, Patrice, DIGNUM, Virginia, LUETGE, Christoph, MADELIN, Robert, PAGALLO, Ugo, ROSSI, Francesca, SCHAFER, Burkhard, VALCKE, Peggy, VAYENA, Effy. An ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. **Minds and Machines**, 28(4), p. 689–707. 2018.

GIERMINDL, Lisa M., STRICH Franz, CHRIST Oliver, LEICHT-DEOBALD Ulrich, REDZEPI, Abdullah. The dark sides of people analytics: reviewing the perils for organizations and employees. **European Journal of Information Systems**. 2021.

GLIGOR, David M; PILLAI, Kishore Gopalakrishna; GOLGECI, Ismail. Theorizing the dark side of business-to-business relationships in the era of AI, big data, and blockchain. **Journal of Business Research**, Elsevier, vol. 133(C), p. 79-88. 2021.

GONZÁLEZ-ESTEBAN, Elsan, CALVO, Patrici. Ethically governing artificial intelligence in the field of scientific research and innovation. **Heliyon**. 2022.

GREWAL, Dhruv, GUHA, Abhijit, SATORNINO, Cinthia B., SCHWEIGER, Elisa B. Artificial intelligence: The light and the darkness. **Journal of Business Research**, Elsevier, vol. 136(C), pages 229-236. 2021.

GRUNDNER, Lukas, NEUHOFER, Barbara. The bright and dark sides of artificial intelligence: A futures perspective on tourist destination experiences. **Journal of Destination Marketing & Managemen**t, 19. 2019.

HAGENDORFF, Thilo. The ethics of AI ethics: an evaluation of guidelines. **Minds Mach.** 30 (1), 2020.

HORNUNG, Olivia, SMOLNIK, Stefan. AI invading the workplace: negative emotions towards the organizational use of personal virtual assistants. **Electron Markets**. 2021.

JABBARPOUR, Mohammad R.; SAGHIRI, Ali M; SOOKHAK, Mehdi. A Framework for Component Selection Considering Dark Sides of Artificial Intelligence: A Case Study on Autonomous Vehicles. **Electronics**, 10, p. 384. 2021.

JENSEN Benjamin M., WHYTE Christopher, CUOMO, Scott. Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence, **International Studies Review**, Volume 22, Issue 3, p. 526–550. 2020.

JIA, Kai, ZHANG, Nan. Categorization and eccentricity of AI risks: a comparative study of the global AI guidelines. **Electron Markets**. 2021.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. The global landscape of AI ethics guidelines. **Nature Machine Intell**. (1), p. 389–399. 2019.

KRAUS, Sascha. BREIER, Matthias. DASÍ-RODRÍGUEZ, Sonia. The art of crafting a systematic literature review in entrepreneurship research. **International Entrepreneurship and Management Journa**l, p. 1–20. 2020.

KUMAR, Gautam. SINGH, Gulbir. BHATANAGAR, Vivek. Scary dark side of artificial intelligence: A perilous contrivance to mankind. **Humanities and Social Sciences Reviews**,7, p. 1097–1103. 2019.

LIU, Han-Wei. LIN, Ching-Fu. CHEN, Yu-Jie. Beyond State v Loomis: artifcial intelligence, government algorithmization and accountability. **International Journal of Law and Information Technology**, 27(2), p. 122–141. 2019.

LIU, Hin-Yan; ZAWIESKA, Karolina. From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. **Ethics Inf. Technology**, 22, p. 321–333. 2020.

LOUREIRO, Sandra M. C; ROMERO, Jaime; BILRO, Ricardo G. Stakeholder engagement in cocreation processes for innovation: A systematic literature review and case study. **Journal of Business Research**. 2019.

MARR, Bernard. What Are The Negative Impacts Of Artificial Intelligence (AI)? Bernard Marr & Co. USA. 2021.

MEMOM, Khalid R; OOI, Say K. The dark side of Industrial Revolution 4.0 - Implications and suggestions. **Academy of Enterpreneuship Journal**, Volume 27, Issue Special Issue 2, p. 1 – 18. 2021.

MIKALEF, Patrick; GRUPTA, Manjul. Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. **Information & Management**, 58(3). 2021.

MIKALEF, Patrick; CONBOY, Kieran; LUNDSTRÖM, Jenny E; POPOVIC, Ales. Thinking responsibly about responsible AI and "the dark side" of AI. **European Journal of Information Systems**. 2022.

MIRBABAIE, Milad; BRÜNKER, Felix; MÖLLMANN FRICK, Nicholas .R.J. STIEGLITZ, Stefan. The rise of artificial intelligence – understanding the AI identity threat at the workplace. **Electronic Markets**. 2021.

NELSON, Gregory S. Bias in Artifcial Intelligence. **North Carolina Medical Journal**, 80(4), p. 220–222. 2019.

PRUNKL, Carina.E.A; ASHUURTS, Carolyn; ANDERLJUNG, Markus; WEB, Helena; LEIKE, Jan; DAFOE, Allan. Institucionalizar la etica en la IA atraves de requisitos de impacto mas amplios. **Nat Mach Intel**l (3), p. 104–110. 2021.

QUEHAJA, Albana B; KUTLLOVCI, Enver; PULA, Justina S. Strategic Management Tools and Techniques: A Comparative Analysis of Empirical Studies. **Croatian Economic Survey,** v.19, n.1. p.67-99. 2017.

RANA, Nripendra P; CHATTERJEE, Sheshadri; DWIVEDI, Yogesh K; AKTER, Shahriar. Understanding dark side of artificial intelligence (AI) integrated business analytics: Assessing firm's operational inefficiency and competitiveness. **European Journal of Information Systems**, p. 1–24. 2021.

SCALIA, Vicenzo; STOCKMAN, Caroline. Politics of technology or technology of politics? **2nd European Conference on the Impact of Artificial Intelligence and Robotics**, ECIAIR 2020. p. 146 – 152. 2020.

SUN, Yi; LI, Shihui; YU, Lingling. The dark sides of AI personal assistant: effects of service failure on user continuance intention. **Electron Markets**. 2021.

WEBSTER, Jane; WATSON, Richard. T. Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26(2). 2002.

WIRTZ, Brend W; WEYERER, Jan C; STURM, Benjamin J. The dark sides of artificial intelligence: An integrated AI governance framework for public administration. **International Journal of Public Administration**, 43(9), p. 818–829. 2020.

ZANZOTTO, Fabio M. Viewpoint: Human-in-the-loop artificial intelligence. **Journal of Artificial Intelligence Research**. 2019.

ZENG Lifan; WU Jun. A two-dimensional research framework for analysing dark side of AI. International Conference on Computer Engineering and Application (ICCEA), 2021, p. 291-294. 2021.