



Congresso Internacional  
de Administração  
ADM 2021

**24 a 28**  
**de outubro**  
Ponta Grossa - Paraná - Brasil

**SOBREVIVÊNCIA DAS ORGANIZAÇÕES  
EM TEMPOS INCERTOS:**

O papel dos gestores e do ambiente externo  
no sucesso e no fracasso organizacional.

## FATORES COGNITIVOS E EMOCIONAIS QUE INFLUENCIAM CONFIANÇA HUMANA EM INTELIGÊNCIA ARTIFICIAL: REVISÃO INTEGRATIVA

**ÁREA TEMÁTICA: INOVAÇÃO, TECNOLOGIA E EMPREENDEDORISMO**

Patrícia Valério Goulart, UFPR, Brasil, [patricia.goulart@ufpr.br](mailto:patricia.goulart@ufpr.br)

Farley Simon Mendes Nobre, UFPR, Brasil, [f.nobre@ufpr.br](mailto:f.nobre@ufpr.br)

### Resumo

A adoção de Inteligência Artificial (IA) se tornou cada vez mais presente nas organizações, contudo, nem todas estão preparadas para a disrupção causada pelo uso desta tecnologia que, de certa forma, pode assumir a função de um ser humano. A adoção de IA requer a confiança das pessoas que a utilizam e esta é ainda uma lacuna de pesquisa no campo teórico e em especial empírico. Este artigo tem como objetivo geral compreender os fatores (cognitivos e emocionais) que influenciam a confiança humana em IA nas organizações. Para atender ao objetivo de pesquisa, realizou-se uma revisão integrativa da literatura (RI) sobre confiança humana e inteligência artificial, utilizando-se de recursos como Web of Science e Scopus. Por se tratar de um tema emergente, o presente estudo buscou identificar quais são os fatores que afetam a confiança humana em IA nas organizações sob diferentes aspectos no que tange a confiança em IA. A partir da RI, elaborou-se um quadro teórico contendo fatores cognitivos e emocionais que influenciam a confiança humana em IA e sua relação com os diferentes tipos de IA nas dimensões de tangibilidade, antropomorfismo, confiabilidade, comportamentos imediatos, transparência e características da tarefa. Resultados e análises da RI permitiram a identificação de fatores que afetam a adoção de IA nas organizações, além de relacionar estes fatores aos diferentes tipos de IA (Robô, Agentes Virtuais e Bot/Embutida). Dessa forma, este artigo contribui para o avanço de pesquisas sobre confiança humana em IA abordando uma importante questão social e tecnológica.

**Palavras-chave:** Confiança humana; inteligência artificial; fatores cognitivos e emocionais; revisão integrativa; decisão.

### Abstract

The adoption of Artificial Intelligence (AI) has become increasingly present in organizations, however, not all are prepared for the disruption caused by the use of this technology that, in a way, can assume the role of a human being. The adoption of AI requires the trust of the people who use it, and this is still a research gap in the theoretical and especially empirical field. This article aims to understand the factors (cognitive and emotional) that influence human trust in AI in organizations. To meet the research objective, an integrative review of literature (IR) on human trust and artificial intelligence was conducted, using resources such as Web of Science and Scopus. As this is an emerging topic, the present study sought to identify the factors that affect human trust in AI in organizations under distinct aspects regarding trust in AI. Based on the IR, a theoretical framework was elaborated containing cognitive and emotional factors that influence human trust in AI and its relationship with the different types of AI in the dimensions of tangibility, anthropomorphism, reliability, immediate behaviors, transparency, and task characteristics. IR results and analysis allowed the identification of factors that affect the adoption of AI in organizations, in addition to relating these factors to selected types of AI (Robot, Virtual Agents and Bot/Embedded). In this way, this article contributes to the advancement of research on human trust in AI by addressing an important social and technological issue.

**Keywords:** Human trust; artificial intelligence; cognitive and emotional factors; integrative review; decision.

## 1. INTRODUÇÃO

Novas tecnologias como Inteligência Artificial (IA) passaram a ser mais amplamente conhecidas, difundidas e utilizadas principalmente com a crescente utilização de mecanismos digitais pelas organizações e seus usuários a partir da 4ª revolução industrial (Indústria 4.0).

Com o avanço da Indústria 4.0, uma revolução baseada em tecnologia para alcançar maior eficiência e produtividade como uma estratégia de competitividade (Xu et al., 2021), as organizações passaram a ser cada vez mais desafiadas pela velocidade com que as mudanças ocorrem e requerem sua adaptação em ambientes complexos, dinâmicos e incertos. Raisch e Krakowski (2021, p.3) inferem que “avanços recentes em poder computacional, o aumento exponencial de dados e novas técnicas de aprendizado de máquina agora permitem que as organizações também usem soluções baseadas em IA” e outras tecnologias.

À medida que as empresas começaram a abraçar a Indústria 4.0, veio a 5ª Revolução Industrial (Indústria 5.0) para emparelhar o ser humano e a máquina a fim de utilizar ainda mais a capacidade intelectual e a criatividade humana para aumentar a eficiência dos processos, combinando os fluxos de trabalho com sistemas inteligentes (Nahavandi, 2019). A Indústria 5.0 é entendida como reconhecimento do poder da indústria para alcançar objetivos sociais além dos empregos e do crescimento desgovernado para se tornar um provedor resiliente de prosperidade, tornando produção respeite os limites do nosso planeta e colocando o bem-estar do trabalhador da indústria como fator central dos processos de produção. Neste sentido a Indústria 4.0 tem foco em tecnologia, enquanto a Indústria 5.0 tem foco no valor (Xu et al., 2021). Embora a principal preocupação da Indústria 4.0 seja sobre automação, a Indústria 5.0 será uma sinergia entre humanos e máquinas autônomas. A força de trabalho autônoma será perceptiva e informada sobre a intenção e o desejo humano (Nahavandi, 2019).

O desafio desta “nova geração de tecnologias capazes de interagir com o meio ambiente visando simular a inteligência humana”, conforme apontado por Glikson e Woolley (2020, p.628) é que para obter êxito na integração de IA nas organizações é preciso que haja confiança dos usuários nesta tecnologia. Mesmo considerando a existência de limitações, as máquinas, que faziam apenas trabalho mecânico, passaram a assumir trabalho cognitivo que antes era realizado apenas por seres humanos, promovendo uma evolução na relação homem-máquina e potencializando as capacidades humanas (Raisch & Krakowski, 2021).

“Por ser percebida como uma tecnologia que pode de certa forma assumir o papel de um humano, as organizações dependem criticamente da confiança dos usuários em IA e este é um fator crítico e determinante para as estas organizações sigam avançando”, destacam Glikson e Woolley (2020, p. 627). Embora “o quadro geral seja aquele em que tecnologias que competem com seres humanos criam ainda mais oportunidades - e geram novos tipos de trabalho”, conforme exposto por Curioni (2018, p.7), existe um temor pelas pessoas de perderem seus empregos para as máquinas ignorando fatos e estatísticas que comprovam que ao a revolução das máquinas “só mudou os tipos de trabalho que as pessoas fazem, ao invés de deixá-las desempregados”, constata Curioni (2018, p.7).

Asan, Bayrak e Choudury (2020, p. 1) destacam que “a confiança é um fator crucial que influencia as interações entre seres humanos, incluindo suas interações com IA”. Promover

confiança humana na utilização de IA é um fator chave para o sucesso da implantação desta tecnologia, principalmente porque os humanos temem serem substituídos pelas máquinas e este temor gera resistência criando barreiras no avanço desta tecnologia. Curioni (2018, p.7) afirma que “computação cognitiva é transformadora no mercado de trabalho, não destrutiva” e Curioni (2018, p.7) argumenta que “apesar das pessoas perderem empregos para as máquinas, os níveis de emprego aumentaram inexoravelmente com o tempo”. Humanos e máquinas são muitas vezes tidos como adversários em função dos efeitos da automação sobre o mercado de trabalho, porém, suas fortalezas combinadas elevam o potencial de aprendizagem melhorando suas capacidades e não o contrário (Curioni, 2018).

Glikson e Woolley (2020 como citado em Prsuraman & Riley, 1997, p.629) destacam que “focar em questões de confiança nos permite abordar não apenas o desuso (uma rejeição) da tecnologia, mas também seu uso indevido (um excesso de confiança inadequada na tecnologia) ou seu abuso (uso prejudicial para obter um ganho individual”.

Confiança possui aspectos cognitivos (baseado e percepções) e emocionais (baseada em relacionamentos e conexões) além de ter diferentes níveis na relação do usuário com a tecnologia (baixa confiança, que pode levar ao desuso, perda de tempo e desperdício de recursos; alta confiança, que poderia levar ao uso indevido ou ocasionar violação de segurança, por exemplo) (Glikson & Woolley, 2020).

Neste contexto, para que as organizações possam potencializar o uso da IA, elas precisarão investir em forma de construir confiança em IA em seus processos produtivos e produtos (Glikson & Woolley, 2020).

## **2. FUNDAMENTOS TEÓRICOS**

Nesta seção são apresentados os conceitos sobre confiança humana, inteligência artificial e fatores cognitivos e emocionais com base em uma revisão da literatura, com intuito de proporcionar melhor compreensão das demais seções. Mais adiante, esses construtos são explorados de forma mais objetiva por meio da revisão integrativa da literatura a fim de possibilitar os resultados, as análises e as respostas para a pergunta de pesquisa.

### **2.1 Confiança Humana**

O termo confiança humana, está relacionado a tendência de um indivíduo em assumir o risco de acreditar em algo que não se pode confirmar ou validar o resultado de forma que este fica vulnerável às ações de outra parte com base na expectativa de que esta irá realizar uma ação esperada ou prometida (Sharan & Romano, 2020).

Nygaard et al. (2017) destaca em um dado contexto confiança pode ser usada como verbo transitivo, o que implica que existe um objeto (pessoa, máquina ou sistema) que é confiável. Isso poderia ser escrito na forma  $A \rightarrow B$ , onde A confia em B para cumprir algum propósito. Confiança é contextual, específica para algum domínio com objetivos e propósito pretendidos cujo objetivo pode ser acessar recursos e dados, fornecer algum serviço, tomar uma decisão ou ainda controlar ou monitorar um processo. Nesse sentido, Nygaard et al. (2017, p. 182) complementam ainda que “se uma pessoa confia em uma entidade que é outra pessoa ou em

uma máquina ou sistema específico, há confiança de que a entidade entregará o desempenho conforme o esperado. A confiança, então, pode ser um substantivo que captura uma crença”.

“A confiança é um elemento vital tanto do lado humano quanto da máquina”, afirma Nygard et al. (2017, p.178). Neste sentido, Zhang et al. (2021, p.3) complementa afirmando que “A confiança humana é crítica para a adoção de sistemas de IA, especialmente em contextos de alta criticidade”. As pessoas usam a confiança como forma de orientar seu comportamento para tomar uma decisão racional quando elas não conseguem confiar nas normas sociais e em seus recursos cognitivos para tomar uma decisão. Esta decisão estará sempre baseada na crença de que a outra parte, seja ela outra pessoa, uma empresa ou um sistema, fará o que se espera que seja feito (Sharan & Romano, 2020).

## **2.2 Inteligência artificial**

IA é uma forma de inteligência humana simulada por tecnologias complexas. Uma nova geração tecnológica capaz de interagir com o meio ambiente através da coleta e processamento de dados e interpretação de informações capazes de prever eventos, avaliar cenários, gera resultados, dar respostas ou formular instruções e se melhorar com base no conhecimento já adquirido (Glikson & Woolley, 2020).

Lapinska et al. (2021) apontam que IA é um conceito extremamente amplo por abranger inúmeras tecnologias e soluções modernas que contam com algoritmos complexos e, por isto, é difícil encontrar uma definição comumente aceita de IA.

Asan, Bayrak e Choudhury (2020, p.2) se referem a IA como “um processo de computador que toma decisões ótimas de maneira algorítmica com base em vários critérios, usando um ou mais modelos baseados em aprendizado de máquina”.

“A literatura futurística assume que a IA é um conjunto de algoritmos capazes de realizar todas as tarefas tão bem ou até melhor do que os humanos”, afirmam Glikson e Woolley (2020, p.628).

IA está diretamente relacionada à capacidade dos sistemas de tomar decisões ou realizar tarefas específicas com pelo menos uma representação parcial da inteligência humana, bem como a capacidade de aprender e melhorar com base nas informações recolhidas. É a contraparte à robótica em empresas de manufatura que apostam em robôs nos processos produtivos para facilitar a execução física dos operários e para apoiar a área executiva no processo de tomada

de decisão cognitiva assumindo cada vez mais tarefas cognitivas complexas (Sassmannshausen et al., 2021).

Por ter a capacidade de traduzir a incerteza e complexidade em dados e decisões clínicas ou sugestões acionáveis - embora imperfeitas, IA pode ser encontrada em processos produtivos e produtos representada de várias formas, sendo as mais comuns:

- Robô físico: representações mecânicas físicas com similaridade humana contendo uma gama de funções e capacidades, que podem estar remotamente ou fisicamente presentes executando tarefas para as quais foram programados (Glikson & Woolley, 2020). Exemplo: Robô babá;
- Agentes virtuais: representação inteligente com características humanas sem presença física, uma identidade distinta como chat bot ou avatar (Glikson & Woolley, 2020). Exemplo: Lu, o chat bot premiado do magazine Luiza; e
- Embutida: é uma forma de IA invisível para os humanos, não possui representação visual e por ser totalmente incorporada a sistemas e pode ser incorporada a produtos e serviços tais como mecanismos de busca, programas e aplicativos (Glikson & Woolley, 2020). Exemplo: GPS.

### 2.3 Fatores cognitivos e emocionais

Confiança é uma construção e pode ser influenciada por fatores cognitivos (avaliação racional e características situacionais) ou emocionais (baseado no afeto, conexões sociais) (Glikson & Woolley, 2020).

Asan, Bayrak e Choudhury (2020, p. 3) afirmam que “IA pode ser influenciado por vários fatores humanos”, tais como:

- Educação;
- Experiências anteriores;
- Preconceitos do usuário;
- Percepção em relação a automação; e
- Propriedades do sistema de IA: controlabilidade, transparência e complexidade do modelo, riscos associados e outros.

Edmonds et al. (2019, p. 2) infere que “fatores cognitivos são uma crença ou uma avaliação com boas razões racionais”. Corroborando com esta afirmação Nygard et al. (2017, p.184) acrescenta que “em uma abordagem cognitiva, os estados de crença são a base da confiança”. Algumas características podem ajudar a fornecer base quantitativa para medir a confiança, tais como: disponibilidade, confiabilidade, segurança, usabilidade, acessibilidade, desempenho computacional e confiabilidade (Nygard et al., 2017). Complementarmente, Sharan e Romano (2020, p.3) destacam que estes fatores podem estar relacionados “ao meio ambiente (colaboração da equipe, tarefa / contexto), relacionado ao parceiro (recursos e capacidades do parceiro), relacionados a humanos (características, estados, fatores cognitivos e emocionais).

Abaixo são destacados alguns fatores emocionais e cognitivos:

- Fatores emocionais: competência, habilidades e experiência e afinidade digital; e
- Fatores cognitivos: percepção de qualidade, plausibilidade, compulsividade das explicações (Sassmannshausen et al., 2021).

O antropomorfismo “aumenta a resiliência da confiança em agentes cognitivos”, afirmam Sharan e Romano (2020, p.2). Aumentar a confiança humana em IA requer aumento da percepção de “cuidado”, características que incluem demonstrações de sentimentos e emoções ajudam neste sentido pois acredita-se que os robôs não têm estas características que ajudam a compreender o outro, como por exemplo, empatia (Yokoi et al., 2021).

### **3. METODOLOGIA: REVISÃO INTEGRATIVA**

Para cumprir o propósito de responder à pergunta de pesquisa “Como está a evolução dos conceitos de fatores cognitivos e emocionais, sobre confiança e inteligência artificial na literatura?”, foram construídos 2 quadros de análise: o primeiro quadro mostra como os conceitos de vêm sendo apresentados em conjunto pela academia, e o segundo quadro aponta em quais dimensões em que os fatores são encontrados de forma mais substancial.

Para responder à pergunta de pesquisa, foi realizada uma RI pois o método auxilia na coleta, análise e correlação de informações de pesquisas anteriores “integrando descobertas e perspectivas de muitos resultados empíricos, uma revisão da literatura pode abordar questões de pesquisa com um poder que nenhum estudo individual tem”, conforme apontado por Snyder (2019, p. 333). Por se tratar de um tema emergente, a abordagem escolhida foi a RI, pois a “maioria das revisões integrativas da literatura se destina a abordar dois tipos de tópicos - tópicos maduros ou tópicos novos e emergentes”, destaca Torracco (2005, p. 357). Além disso, considerando o exposto por Snyder (2019, p. 334), “uma revisão integrativa é abordagem que pode ser útil quando o objetivo da revisão não é cobrir todos os artigos já publicados sobre o tema, mas sim para combinar perspectivas para criar novos modelos teóricos”.

#### **3.1 Análise e seleção dos artigos**

Por se adequar melhor aos objetivos deste trabalho a técnica escolhida foi a RI, pois permite a realização de análise, crítica e síntese da literatura de forma a integrar as visões, criando perspectivas novas (Torraco, 2005).

Uma RI é um método específico que resume o passado da literatura empírica ou teórica com o objetivo de traçar uma análise sobre o conhecimento já construído anteriormente sobre um dado tema a fim de fornecer uma compreensão mais abrangente de um fenômeno particular possibilitando a síntese de vários estudos já publicados, gerando novos conhecimentos (Botelho, Cunha & Macedo, 2011). A finalidade da RI é integrar perspectivas através do

mapeamento e visão geral das publicações e não necessariamente cobrir todas as publicações existentes (Snyder, 2019; Torraco, 2016).

O primeiro passo foi realizar o planejamento da RI definindo as bases Scopus e Web of Science. Já os critérios de busca foram estabelecidos considerando os campos de busca, título e resumo, e as chaves de busca "human trust" OR "people trust" AND "artificial intelligence" OR "AI". A autora também reduziu a busca a "artigos", "publicados na língua inglesa" e sem delimitação de tempo. No dia 19/02/2022 foi realizada a busca que resultou em 22 artigos encontrados na base Scopus e 12 artigos encontrados na base WOS, totalizando 34 artigos. Destes, identificou-se 9 artigos repetidos, que foram removidos, totalizando 25 artigos encontrados.

Por fim, com objetivo de responder à questão: Como está a evolução dos conceitos e fatores sobre confiança humana em IA? Os 15 artigos selecionados foram utilizados para construção de 2 quadros:

- Evolução dos conceitos de confiança, IA e fatores cognitivos e emocionais; e
- Fatores cognitivos e emocionais que influenciam confiança humana em IA.

As fases adotadas para RI podem também ser observadas na Figura 1, que ilustra o fluxograma Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009).

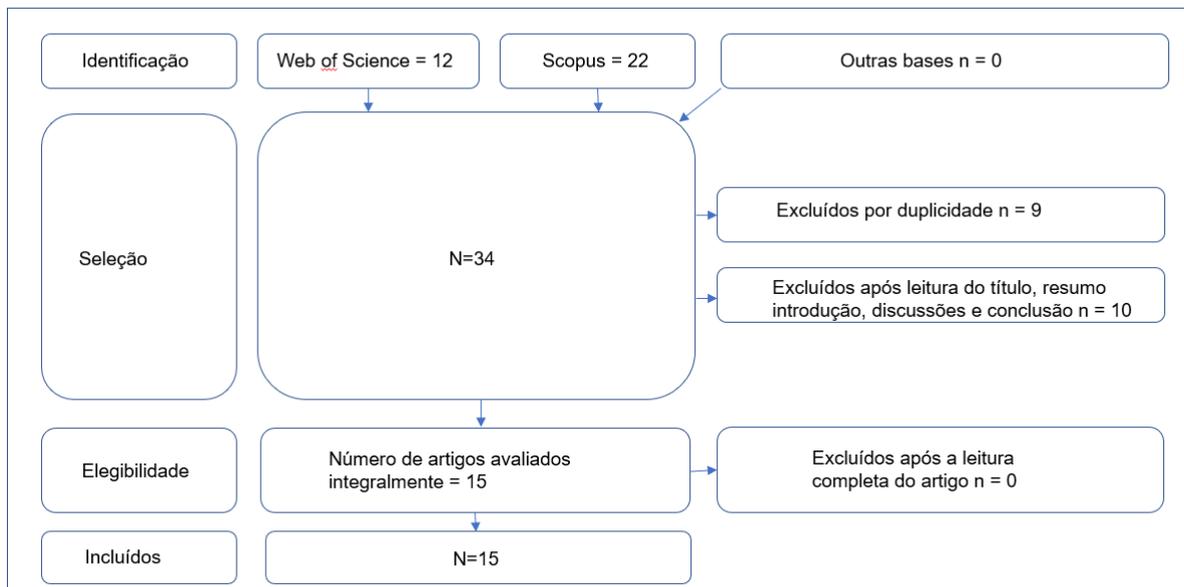


Figura 1 – Fluxograma Prisma

### 3.2 Tabulação dos dados

Considerando a recomendação de Snyder (2019), em que o pesquisador deve utilizar uma abordagem qualitativa para avaliar os achados após seguir um processo de revisão sistemática

para reunir as publicações, com isto, foi realizada a tabulação dos 15 artigos elegidos utilizando como ferramenta para registro, organização e análise de dados o software Microsoft Excel, estruturando os seguintes campos descritivos: identificador, ano de publicação título, autores, breve resumo, identificação dos conceitos de confiança, IA e fatores cognitivos e emocionais.

A análise do conteúdo dos 15 artigos selecionados, considerou 3 etapas (Bardin, 2011): 1. pré-análise, com a primeira leitura dos 15 artigos selecionados; 2. codificação e categorização dos dados; e 3. Interpretação dos dados e inferência.

#### 4. RESULTADOS E ANÁLISES

##### 4.1 Quadros de análise

Com base na revisão integrativa e na análise de conteúdo dos 15 artigos selecionados, foram elaborados os quadros 1 e 2.

O quadro 1 abaixo apresenta os conceitos de confiança humana, inteligência emocional e fatores cognitivos e emocionais encontrados nos artigos referenciados na terceira coluna.

CONSTRUTOS	CONCEITOS	REFERÊNCIA
Confiança humana	Confiança humana é um aspecto essencial das relações interpessoais e se refere a um estado psicológico, uma crença, um mecanismo para lidar com a incerteza e uma tendência a lidar com riscos diante do que é conhecido e desconhecido. Representa comportamentos gerados a partir de perspectivas diversas cujo objeto não são apenas pessoas, mas também, instituições, sistemas computacionais e outros objetos.	Song e Zhao (2022); Lewis e Marsh (2022); Yokoi et al. (2021); Lapinska et al. (2021); Sassmannshausen et al. (2021); Vereschak, Bailly e Caramiaux (2021); Asan, Bayrak, e Choudhury (2020); Glikson e Woolley (2020); Sharan e Romano (2020); Edmonds et al. (2019); Nygard et al. (2017).
Inteligência artificial	IA é uma tecnologia moderna altamente de alta capacidade de processamento de dados baseada em conceitos computacionais complexos e abrangentes com a finalidade de melhorar a tomada de decisão em situações envolvendo incerteza. Baseada em sistemas computacionais, IA tende a imitar a inteligência humana na forma de robôs físicos, virtuais ou embutidos e com capacidade de aprendizagem interativa das funções cognitivas humanas para processar informações gerando resultados	Song e Zhao (2022); Lewis e Marsh (2022); Yokoi et al. (2021); Lapinska et al. (2021); Sassmannshausen et al. (2021); Vereschak, Bailly e Caramiaux (2021); Asan, Bayrak, e Choudhury (2020); Glikson e Woolley (2020); Sharan e Romano (2020); Edmonds et al. (2019); Nygard et al. (2017).
Fatores cognitivos e emocionais	São fatores característicos do comportamento humano, relacionados a avaliações cognitivas ou racionais/situacionais ou	Song e Zhao (2022); Zhang et al. (2021); Yokoi et al. (2021); Lapinska et al. (2021); Sassmannshausen et al. (2021);

	emocionais/afetivas. Os fatores cognitivos e emocionais são diretamente afetados pelas crenças, culturas e ambiente em que as pessoas estão inseridas.	Asan, Bayrak, e Choudhury (2020); Glikson e Woolley (2020); Sharan e Romano (2020); Edmonds et al. (2019); Nygard et al. (2017).
--	--	--

Quadro 1 – evolução dos conceitos na literatura

Dentre os 15 artigos revisados, encontrou-se um artigo em potencial: *Human trust in artificial intelligence: Review of empirical research* de Glikson e Woolley (2020) cujo abrange todos os fatores cognitivos e emocionais identificados na RI e mais alguns, em um framework desenvolvido pelas autoras.

Glikson e Woolley (2020) consideram que confiança é uma construção e pode ser influenciada por fatores cognitivos (avaliação racional e características situacionais) ou emocionais (baseado no afeto, conexões sociais), neste sentido, as autoras apresentam como tais fatores influenciam o grau de confiança em cada tipo de IA (Robô físico, Agentes virtuais e embutida) propõem em seu framework as dimensões de confiança cognitiva (tangibilidade, transparência, confiabilidade, características da tarefa e comportamentos imediatos) e emocional (tangibilidade comportamentos imediatos e Antropomorfismo).

- Tangibilidade: Se refere a capacidade de ser percebido ou tocado durante o desenvolvimento de confiança (Glikson & Woolley, 2020);
- Transparência: Está relacionada ao quanto a lógica interna da tecnologia é aparente para os usuários, reflete o nível em que as regras operacionais subjacentes estão sendo consideradas e apresentadas para desenvolver confiança na nova tecnologia (Glikson & Woolley, 2020);
- Confiabilidade: Capacidade de apresentar o mesmo comportamento esperado ao longo do tempo, também é fundamental para a confiabilidade da tecnologia (Glikson & Woolley, 2020);
- Características da tarefa: São tarefas relacionadas ao trabalho que a “tecnologia” ou IA está realizando. Como IA lida com julgamentos amplamente técnicos ou interpessoais, pode ser um importante antecedente o papel das características da tarefa no desenvolvimento da confiança cognitiva (Glikson & Woolley, 2020);
- Comportamentos imediatos: É o grau de percepção física e/ou psicológica relacionado à proximidade entre as pessoas. Referem-se a gestos socialmente orientados destinados a aumentar a proximidade interpessoal, como proatividade, escuta ativa e responsividade (Glikson & Woolley, 2020); e
- Antropomorfismo: Semelhança humana, refere-se à percepção da tecnologia ou de um objeto como tendo qualidades humanas, como sentimentos. Percepções podem ser impulsionadas por recursos de interface, como a forma humana do robô, por características comportamentais, como olhar, e por enquadramento intencional, como dar a um robô ou *bot* um nome humano (Glikson & Woolley, 2020).

Com isto, optou-se pela adoção destas dimensões para estruturação do segundo quadro. O quadro 2, abaixo apresentado como Figura 2, aponta em quais dimensões os fatores cognitivos e emocionais são encontrados de forma mais substancial.

DIMENSÕES	FATORES COGNITIVOS E EMOCIONAIS QUE INFLUENCIAM CONFIANÇA EM IA	Song, Xu b. e Zhao (2022)	Lewis e Marsh (2022)	Chong et al. (2022)	Akula et al. (2022)	Yokoi et al. (2021)	Zhang et al. (2021)	Lapinska et al. (2021)	Vereshak, Bailly e Caramiaux (2021)	Sassmannshausen et al. (2021)	Asan, Bayrak, e Choudhury (2020)	Glikson e Woollay (2020)	Sharan e Romano (2020)	Edmonds et al. (2019)	Curioni (2018)	Nygard et al. (2017)
Tangibilidade	Presença ou representação humana física/virtual	X						X				X	X			
Transparência	Visibilidade e entendimento de como IA está sendo implantada		X		X		X		X		X	X	X		X	X
Confiabilidade	Confiabilidade do modelo	X	X		X		X	X		X	X	X				
Características da tarefa	Nível de complexidade de tarefas realizadas por IA	X	X	X			X	X	X	X	X	X	X	X	X	X
Comportamentos imediatos	Capacidade de resposta, adaptabilidade e comportamentos pró-sociais		X			X		X	X		X	X	X			X
Antropomorfismo	Similaridade de IA com humano	X						X				X	X			X

Figura 2 – evolução dos conceitos na literatura

#### 4.2 Distribuição dos artigos por dimensão

Os artigos foram analisados e classificados em ordem decrescente considerando a influência dos fatores cognitivos e emocionais nas dimensões abordadas de forma mais substancial em cada trabalho. A partir desta análise foi possível verificar que os fatores que apresentaram maior popularidade estão nas dimensões de transparência presente em pelo menos (60%) dos artigos e Características da tarefa presente em (87%) dos artigos.

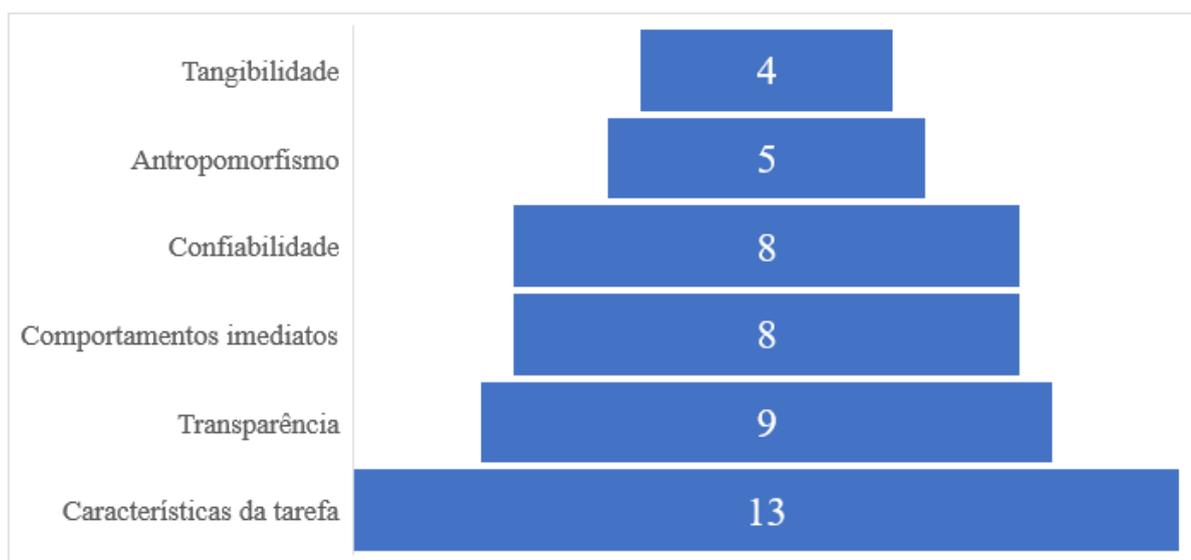


Figura 3 – Distribuição dos artigos por dimensão

## 5. DISCUSSÕES E ESTUDOS FUTUROS

Neste estudo, foram identificados alguns dos principais fatores cognitivos e emocionais que influenciam confiança humana em IA em distintas dimensões. Conhecer estes fatores é fundamental para realçar as oportunidades que existem para o desenvolvimento de confiança em IA, uma vez que esta tecnologia está se desenvolvendo muito rapidamente, possivelmente de forma exponencial, e como as pessoas vão se ajustar a isso é ainda um problema fascinante, comenta Kahneman (2021).

Promover confiança humana na utilização de IA é um fator chave para o sucesso da implantação desta tecnologia, principalmente porque os humanos temem serem substituídos pelas máquinas e este temor gera resistência criando barreiras no avanço desta tecnologia, porém, suas fortalezas combinadas elevam o potencial de aprendizagem melhorando suas capacidades e não o contrário (Curioni, 2018).

Toda mudança requer preparação e passa por um período de adaptação, e neste sentido Glikson e Woolley (2020, P.633) destacam que “pesquisas existentes abordando a trajetória do ser humano e confiança em IA sugerem que a confiança inicial começa em um nível baixo e se desenvolve ao longo do tempo”. Com isto, conclui-se que, conhecer os fatores cognitivos e emocionais que influenciam confiança humana em IA oferece grande potencial para o desenvolvimento de confiança em IA nas organizações, favorecendo a aplicação e uso.

A partir da revisão integrativa, observou-se a necessidade de realização de estudos mais aprofundados sobre a influência dos fatores cognitivos e emocionais em relação a confiança humana em IA em função do baixo número de trabalhos encontrados na literatura. Recomenda-se que pesquisas futuras analisem também os fatores cognitivos e emocionais sob a perspectiva de longo prazo ao examinar o efeito da capacidade de resposta da IA a diferentes comportamentos de desuso e abuso no desenvolvimento da confiança e na relação de trabalho humano-IA (Glikson & Woolley, 2020).

Por fim, sugerem-se a condução de pesquisas envolvendo casos práticos nas organizações a fim de comparar os achados na literatura com o ambiente da vida real e como organizações que já usam IA em seus sistemas de gestão ou tomada de decisão lidam com confiança humano-IA.

## CONCLUSÃO

A revisão integrativa contribuiu para responder à pergunta de pesquisa “Como está a evolução dos conceitos de fatores cognitivos e emocionais, sobre confiança e inteligência artificial na literatura?”.

A partir dos resultados da RI foi possível identificar os fatores cognitivos e emocionais que influenciam confiança humana em IA na forma que vem sendo apresentados pela academia e sugerir que estes possam ser testados em estudos empíricos futuros nas organizações.

Acredita-se que as contribuições deste artigo permitiram identificar quais os fatores cognitivos e emocionais que influenciam confiança humana em IA.

Por fim, este estudo possui limitações devido ao número reduzido de artigos disponíveis sobre o tema estudado.

## REFERENCIAS

- Asan, O., Bayrak, a. E., Choudhury, A. (2020) Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians, *Journal of Medical Internet Research* (22:6), pp. 1–7.
- Botelho, L. L. R., de Almeida Cunha, C. C., Macedo, M. (2011). O método da revisão integrativa nos estudos organizacionais, *Gestão e sociedade* (5: 11), pp. 121–136.
- Bridge, F. (2020). Artificial Intelligence Reshaping the Automotive Industry. (Available at <https://www.futurebridge.com/industry/perspectives-mobility/artificial-intelligence-reshaping-the-automotive-industry/>).
- Christenberry, T. (2017). Standalone Interviews Do Not Equal Qualitative Research, *Nurse Author & Editor* (27:4), pp. 1–9.
- Cooper, D. R., Schindler, P. S. (2016). *Métodos de pesquisa em Administração* (12th ed). Porto Alegre, AMGH.
- Creswell, J. W. (2014). *Projeto de pesquisa: métodos qualitativo, quantitativo e misto* (3rd ed). Porto Alegre: Bookman.
- Curioni, A. (2018). Artificial intelligence: Why we must get it right, *Informatik-Spektrum* (41:1), pp. 7–14.
- Edmonds, M., Gao, F., Liu, H. (2019). A tale of two explanations: Enhancing human trust by explaining robot behavior, *Science Robotics*, (4:37), pp. 1–14.
- Eisenhardt, k. M. (1989). Building theories from case study research, *Academy of management review*, (14:4) pp. 532–550.
- Eisenhardt, k. M., Graebner M. E. (2014). Theory building from cases: opportunities and challenges diverse, *Academy of Management Journal* (50:1), pp. 25–32.
- GIL, A. C. (2019). *Métodos e Técnicas de Pesquisa Social* (7th ed). São Paulo: Atlas.
- Glikson, E.; Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, (14:2), pp. 627–660.
- Godoi, C.K; Bandeira-de-Mello, R.B.; Silva, A. B. (2010). *Pesquisa qualitativa em estudos organizacionais: Paradigmas, estratégias e métodos* (2nd ed). São Paulo: Saraiva.
- Kahneman, D. (2021, maio 10) Clearly AI is going to win. How people are going to adjust is a fascinating problem. (Entrevista com Daniel Kahneman). *The Guardian*. (Available at [https://www.theguardian.com/books/2021/may/16/daniel-kahneman-clearly-ai-is-going-to-win-how-people-are-going-to-adjust-is-a-fascinating-problem-thinking-fast-and-slow#:~:text=How%20people%20are%20going%20to%20adjust%20is%20a%20fascinating%20problem',-Tim%20Adams](https://www.theguardian.com/books/2021/may/16/daniel-kahneman-clearly-ai-is-going-to-win-how-people-are-going-to-adjust-is-a-fascinating-problem-thinking-fast-and-slow#:~:text=How%20people%20are%20going%20to%20adjust%20is%20a%20fascinating%20problem',-Tim%20Adams))).
- Kothari, C. R. (2004). *Research methodology: Methods and techniques*.

- Lapińska, J., Escher, I., Górka, J., Sudolska, A., Brzustewicz, P. (2021). Employees' trust in artificial intelligence in companies: The case of energy and chemical industries in Poland. *Energies*, (14:7).
- March, J. G., Simon, H. A. (1993). *Organizations*. John Wiley & Sons.
- Martins, G. DE A., Theóphilo, C. R. (2007). *Metodologia da Investigação Científica para Ciências Sociais Aplicadas* (3rd ed). São Paulo: Atlas.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, (151:4), pp. 264–269, 2009.
- Nahavandi, S. (2019). Industry 5.0 a human-centric solution. *Sustainability* (11:16).
- Nutt, P. C., Wilson, D. C. (2010). *Handbook of decision making*. John Wiley & Sons.
- Nygard, K., Chowdhury, M., Bugalwi, A., Kotala, P. (2017). *People and Intelligent Machines in Decision Making*. *Int. J. Comput.*
- Pereira, J. M. (2010). *Manual de metodologia da pesquisa científica* (4th ed). São Paulo: Atlas.
- Raisch, S., Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, (46:1), pp. 192–210.
- Sassmannshausen, T., Burggräf, P., Wagner, J. (2021). Trust in artificial intelligence within production management—an exploration of antecedents. *Ergonomics*, (0:0), pp. 1–18.
- Seuring, S. A. (2008). Assessing the rigor of case study research in supply chain management. *Supply Chain Management: An International Journal*, (13:2), pp. 128–137.
- Sharan, N. N., Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon*, (6:8).
- Simon, H. A. (1997). *Administrative behavior* (4th ed). New York: The free Press.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, pp. 333–339.
- Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, (4:3), pp. 356–367.
- Triviños, A. N. S. (1987). *A pesquisa qualitativa em educação*. São Paulo: Atlas.
- Vergara, S. C. (2016). *Projetos e relatórios de pesquisa em administração* (16th ed). São Paulo: Atlas.
- Voss, C (2010). Case research in operations management. *Researching operations management*. pp.176–209.
- Xu, X., Lu, Y., Vogel-Heuser, B.; Wang, L. (2021). Industry 4.0 and Industry 5.0—Inception, conception, and perception. *Journal of Manufacturing Systems*, (61), pp. 530–535. (Available at <https://www.sciencedirect.com/science/article/pii/S0278612521002119>).
- Yin, R. K. (2015). *Estudo de Caso: Planejamento e métodos* (5th ed). Porto Alegre: Bookman.

Yokio, R., Eguchi, Y., Fujita, T., Nakayachi, K. (2021). Artificial Intelligence Is Trusted Less than a Doctor in Medical Treatment Decisions: Influence of Perceived Care and Value Similarity. *International Journal of Human-Computer Interaction*, (37:10), pp. 981–990.

Zhang, Z., Citardi, D., Wang, D. (2021). Patients' perceptions of using artificial intelligence (AI)-based technology to comprehend radiology imaging data. *Health Informatics Journal*, (27:2).