ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/336680544

A Ética na Inteligência Artificial: Desafios Ethics of Artificial Intelligence: Challenges

Conference Paper · July 2019

CITATIONS
2

READS
854

3 authors, including:

Manuela Aparicio
Universidade NOVA de Lisboa
118 PUBLICATIONS **1,006** CITATIONS
SEE PROFILE

Carlos J. Costa
ISEG – Lisboa School of Economics and Management. University of Lisbon
217 PUBLICATIONS **1,115** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

WICTA 9th ed Special Interest Group on ICT for Auditing & Accounting, Chaves, Portugal (June, 2021) View project

Playing, Learning, Investing – a quiz game to improve the financial literacy of bank private investors. View project

All content following this page was uploaded by Manuela Aparicio on 20 October 2019.

The user has requested enhancement of the downloaded file.

A Ética na Inteligência Artificial: Desafios

Ethics of Artificial Intelligence: Challenges

Martinha Piteira
Instituto Politécnico de Setúbal, IPS, Setúbal
Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL
Portugal
martinha.piteira@estsetubal.ips.pt

Manuela Aparicio
Instituto Universitario de Lisboa (ISCTE-IUL) ISTAR-IUL
Information Management School (Nova IMS), Universidade Nova de Lisboa, Portugal
manuela.aparicio@acm.org

Carlos J. Costa
Advance/CSG, ISEG (Lisbon School of Economics & Management), Universidade de Lisboa
cjcosta@iseg.ulisboa.pt

Resumo - **A inteligência artificial (IA) tem nos últimos tempos assumido um papel relevante nos mais diversos setores da nossa sociedade. Estamos num ponto sem retorno, e o nosso futuro passará naturalmente pela incorporação da inteligência artificial, na nossa vida diária, seja profissional, seja pessoal. A ideia da existência de máquinas "pensantes" e que tomem decisões pelos Humanos levanta uma série de questões éticas que devem estar presentes aquando do desenvolvimento e incorporação da inteligência artificial nos mais diversos setores da sociedade. É fundamental estudar e investigar as melhores abordagens à sua integração. O presente estudo identifica, os principais princípios orientadores da ética, na área da informática e sistemas de informação, em contexto de uso de sistemas inteligentes e autónomos. Apresenta também os resultados de um estudo bibliométrico, no qual são identificados os principais grupos de assuntos estudados e publicados na ACM e IEEE pela comunidade científica.**

***Palavras Chave** - Ética; Inteligência Artificial; Framework, Estudo Bibliométrico.*

Abstract - ***Artificial intelligence (AI) has in recent times assumed a relevant role in the most diverse sectors of our society. We are at a no return point, and our future will incorporate artificial intelligence into our everyday life, professional or personal. The idea of "thinking" machines existence, making decisions by Humans raises several ethical questions. It is fundamental to study and investigate the best approaches to their integration. This article identifies the guiding principles of ethics in the context of using intelligent and autonomous systems. Here we present a bibliometric study, reporting the main ACM and IEEE studies on Ethics and AI. Our results indicate various clusters of Ethics and AI, that the scientific community has been focused on.***

Keywords - Ethics; Artificial Intelligence; Framework, Bibliometric Study.

I. INTRODUÇÃO

As tecnologias que emergiram nos últimos anos, têm nos transportado para o imaginário de um admirável "novo mundo", com a inteligência artificial (IA) a assumir um papel de destaque. A sua utilização nas organizações tem enormes impactos, principalmente na tomada de decisão. Através da sua capacidade de fornecer quantidades enormes de informação e conhecimento. A inteligência artificial influenciará a dinâmica da tomada de decisões em diversos contextos [1]. Por exemplo, no setor bancário, no sector automóvel, na educação, entre outros. A utilização de algoritmos de *machine learning*, nestes setores levantam diversos desafios éticos. Por exemplo, os algoritmos que recomendam com base em determinados critérios a aprovação ou não da hipoteca [1]–[4]. E, se a rejeição acontece baseada em discriminação racial? E, os algoritmos que são incorporados nas viaturas de condução autónoma e que tomam decisões perante determinadas situações. E, se for inevitável o choque frontal com um conjunto de pessoas? Qual, ou quais as que escolhe para esse choque? Baseia a sua decisão na idade? Na raça?[1]. A IA assume também um papel relevante na educação. Por exemplo, no reconhecimento da aquisição de competências [5]–[7]. Que os algoritmos de inteligência artificial representam um papel importante e crescente na nossa sociedade, é uma realidade. Os cenários acima descritos, são assustadores, por isso, a importância crescente de desenvolver algoritmos de inteligência artificial, que não sejam apenas poderosos e escaláveis, mas acima de tudo, que sejam transparentes para inspeção. Por outro lado, é fundamental que os algoritmos sejam robustos o suficiente de forma a evitar a manipulação.

Os novos desafios éticos que se colocam atualmente na inteligência artificial, estão relacionados com o facto de os algoritmos de IA serem utilizados para tarefas com dimensões sociais – cognitivas anteriormente realizada por Humanos. Neste caso, os algoritmos herdam requisitos sociais [1]. Nesse sentido, perceber o impacto que a utilização da IA nas organizações e nas vidas das pessoas tem é fundamental. Mas, principalmente identificar os princípios éticos dessa aplicação e como os podemos monitorizar e agir. Nesse sentido, este este artigo identifica os principais desafios éticos relativos à utilização da inteligência artificial. A abordagem metodológica seguida neste artigo seguiu o método de estudo documental e bibliométrico para identificar as diversas áreas cobertas na literatura que incluem estudos relacionados com ética e inteligência artificial, nomeadamente nas duas principais bibliotecas digitais da área das ciências da computação a *Association for Computing and Machinery* (ACM) [8], [9] e a *Institute of Electronical and Electronics Engineers* (IEEE) [10], [11].

Este artigo está estruturado em quatro secções. Na secção seguinte (secção II) são apresentados os princípios de ética preconizados pelas duas principais associações mundiais das ciências da computação. Na secção III é apresentado um mapeamento dos principais trabalhos científicos que se realizaram. Finalizamos com as conclusões.

II. ENQUADRAMENTO TEÓRICO

A. *Conceitos Associados*

Associados à temática do presente artigo, a literatura aborda diversos conceitos. Nesta secção identificamos e descrevemos

2019 14th Iberian Conference on Information Systems and Technologies (CISTI)
19 – 22 June 2019, Coimbra, Portugal
ISBN: 978-989-98434-9-3

alguns desses conceitos. Dois dos conceitos que são referidos na literatura são: "responsabilidade social" e "ética". A literatura apresenta diversos pontos de vista sobre estes conceitos. A responsabilidade social e a ética são conceitos muitas das vezes utilizados de forma indistinta. Por exemplo, a responsabilidade social é apresentada como sendo a ética num contexto organizacional; a responsabilidade social foca-se no impacto que a atividade do negócio tem em toda a sociedade. Enquanto que o conceito de ética está relacionado com a conduta daqueles que fazem parte da organização. A literatura não apresenta relação entre a responsabilidade social e a ética, apresentando ainda que a responsabilidade social tem diversas dimensões, entre elas a ética [12]. Um terceiro conceito revisto na literatura é a moralidade. De acordo com Beauchamp e Bowie [3], a moralidade refere-se a princípios ou regras de conduta moral tal como estão definidas na sociedade. A moralidade existe antes da aceitação ou rejeição dos seus standards pelos indivíduos, nesse sentido a moralidade não pode ser meramente uma política ou um código. Nesse sentido, a ética tem a ver com o refletir sobre a natureza e justificações do certo e do errado [3].

B. The Code ACM

A ACM – *Association for Computing Machinery* apresentou recentemente um documento com o código de ética e conduta profissional, (*The ACM Code*) atualizado [13]. Esse documento identifica os princípios gerais de ética, orientadores de uma conduta profissional. Os princípios definidos pela ACM foram organizados em quatro grupos, nomeadamente: princípios gerais de ética, responsabilidades profissionais, princípios de profissionais de liderança e conformidade com o *The Code*.

Tabela 1 – Tabela resumo do Código de Ética e Conduta Profissional - Adaptado da ACM [13]

Dimensões de Atuação		Principios de ética	Ideias Chave
Principios Gerais de Ética	Um profissional dos sisteamas e tecnologias da informação deve:	Contribuir para a sociedade e para o seu bem-estar, tendo presente que todas as pessoas são partes interessadas nos sistemas de informação (SI /TI)	Deve ser tido em consideração que todas as pessoas são partes interessadas em SI/TI e que devemos contribuir para a sociedade como um todo A honestidade é uma componente essencial da confiança entre as partes
		Evitar danos.	
		Ser honesto e confiável	
		Ser justo e não tomar ações de discriminação	
		Respeitar o trabalho que requer a produção de ideias, invenções, trabalho criativo e artefactos	
		Respeitar a privacidade	
		Honrar a Confidencialidade	
Responsabilidade Profissional		Buscar elevados padrões de qualidade: produto e no trabalho profissional	Um profissional deve orientar a sua atuação para que crie um impacto positivo na sociedade A computação é um serviço para a sociedade Apoiar de forma consistente o bem comum
		Manter altos os standards de competência profissional, competência e praticar a ética	
		Conhecer e respeitar as regras existentes pertinentes para o trabalho profissional	
		Aceitar e providenciar uma revisão profissional do trabalho a realizar ou realizado	
		fornecer avaliações abrangentes e completas dos sistemas de computação e seus impactos, incluindo análises de possíveis riscos.	
		Realizar tarefas apenas em áreas da sua competência	
		Promover a conscientização pública e compreensão da computação, relativas à tecnologia e às suas consequências	
		Aceder a recursos de computação e comunicação somente quando autorizado ou quando for obrigado para o bem público	
		Desenhar e implementar sistemas que sejam robustos e usáveis em segurança	
Conformidade com o *The Code*		Defender, promover e respeitar os princípios do The Code.	Um profissional deverá conhecer, defender, divulgar, promover os principios de ética do The Code. Deve identificar violações e tratar da sua resolução junto das entidades respetivas.
		Trate as violações do código como inconsistentes com a participação da ACM.	
Liderança Profissional	Um profissional de computação, quando actua como líder, deve:	Assegurar que o bem público é a principal preocupação.	Um líder deve incentivar, articular, implementar e dar suporte a uma conduta ética de todos os profissionais dos sistemas de informação.
		Articular, incentivar a aceitação e avaliar o cumprimento das responsabilidades sociais pelos membros da organização ou grupo	
		Gerir as pessoas e recursos para promover / melhorar a qualidade de vida no trabalho	
		Articular, aplicar e apoiar políticas e processos que reflitam os princípios do The Code	
		Criar oportunidades para os membros da organização crescerem como profissionais.	
		O profissional deve estar ciente das alterações que forem realizadas nos sistemas ou mesmo a sua descontinuidade.	
		Reconhecer e cuidar de sistemas que se integrem na infraestrutura da sociedade	

De acordo com o preconizado ela ACM [13], os princípios gerais de ética é a primeira área de atuação, composta por sete princípios orientadores no desenvolvimento dos sistemas de informação, cf. Tabela 1. Sendo a ética e os princípios associados como orientadores de uma conduta organizacional que os indivíduos devem seguir, um dos princípios de partida nesta área de atuação é consciência que os profissionais da área de Sistemas de Informação e Tecnologias da Informação devem ter contribuindo para sociedade e para o seu bem-estar, tendo presente que todos são partes intervenientes e interessadas nos SI. Nesse sentido, os profissionais deve adotar uma conduta que tenham em consideração o desenvolvimento dos sistemas com a preocupação de melhorar a qualidade de vida das pessoas, consequentemente devem usar as suas competências para o benefício da sociedade. A sua atuação deverá também evitar danos, i.e. evitar impactos negativos, especialmente aqueles que sejam significantes e desajustados, como ameaças físicas e mentais, e a destruição não justificadas de informação, entre outros. A base de uma relação organizacional deve ser construída assente em pilares de honestidade e confiança. Nesse sentido, um profissional deve ser transparente e fornecer

ISBN: 978-989-98434-9-3

todas as informações relativas às capacidades, limitações e potenciais problemas de um sistema, assim como ser honesto sobre as suas qualificações e limitações para a concretização de uma determinada tarefa. Deverá também ser justo e não tomar ações de discriminação, defender e aplicar valores de igualdade, tolerância, respeito pelos outros. Nesse sentido, deve respeitar o trabalho que requer a produção de ideias, invenções, trabalhos criativos e a produção de artefactos computacionais que permitam acrescer valor para a sociedade. Deve assim contribuir para que o valor desses profissionais sejam reconhecidos e atribuídos os créditos a quem de direito e respeitar os direitos autorais, patentes, licenças e outros métodos de proteção dos trabalhos autorais. Respeitar a privacidade e honrar a confidencialidade: o princípio da privacidade aplica-se aos profissionais da área de informática de um modo muito particular. Estes profissionais lidam com a tecnologia que permite recolher, monitorar e trocar informação pessoal muito rapidamente, sem o conhecimento da(s) pessoa(s) afetadas. Por outro lado, a estes profissionais é entregue ou dado o acesso a informações confidenciais, como segredos de negócio, dados de clientes, estratégias empresariais não divulgadas, informações financeiras, dados de pesquisas, entre outros. Os profissionais devem proteger a confidencialidade dessas informações, exceto quando existem evidências da violação da lei, violações dos regulamentos organizacionais, ou violações do *The ACM Code*.

Um profissional deverá ser orientado e guiado por um conjunto de princípios que definam as suas responsabilidade profissionais, e nesse sentido o *The Code* [13] identifica esta como sendo uma segunda área de atuação e que é composta pelos seguintes princípios: a) Buscar elevados padrões de qualidade: produto e no trabalho profissional e manter altos standards de competência profissional, competência e praticar a ética, estes dois princípios centram-se na qualidade do trabalho que deve ser desenvolvido, sendo que essa qualidade depende de quem individualmente ou em equipa assuma a responsabilidade por garantir e manter um elevado nível de competência. A competência profissional começa com os conhecimentos técnicos, e consciente do contexto social onde o seu trabalho vai ser implementado. Para além de *soft skills* necessárias o profissional deverá ser capaz de reconhecer e identificar os desafios éticos associados ao seu trabalho. b) Conhecer e respeitar as regras existentes pertinentes para o trabalho profissional, aceitar e providenciar uma revisão profissional do trabalho a realizar ou realizado e realizar apenas tarefas da sua competência: deverão ser tidas em contas as regras, procedimentos locais, regionais e até leis internacionais. Um profissional deverá tê-las em consideração a menos exista um motivo ético convincente para não serem tidas em consideração. Também a qualidade do trabalho depende de uma revisão em todas as suas fases. Deverá haver uma avaliação pelos pares e *stakeolders*, assim como manter uma análise critica do seu trabalho e dos seus colegas. Naturalmente, para que um trabalho seja bem-sucedido o profissional apenas deverá realizar tarefas que são da sua competência. c) Promover a conscientização pública e compreensão dos SI/TI, relativas à tecnologia e às suas consequências: cabe ao profissional da área de informática, dado estar numa posição de confiança e conhecimento fornecer objetivamente, avaliações e informações aos vários intervenientes no processo. Deverá ter um cuidado maior quando identifica e mitiga potenciais riscos associados ao *machine learning*.

A terceira área de atuação identificada é a Liderança Profissional. Esta área é composta por sete princípios, cf. Tabela 1. Esta área reforça o papel que um profissional deve assumir, especialmente quando atua como líder. Um líder deverá ter sempre como ponto de partida o bem comum da sociedade e nesse sentido deverá tê-lo presente quando define tarefas e executa todos as fases de desenvolvimento de projeto. Esse deverá ser o seu principal objetivo independentemente da metodologia utilizada. Deverá também ser um articulador, um elemento ativo e incentivar as equipas que lideram no sentido de aceitarem as responsabilidades sociais e desencorajar tendências que as ponham em causa. Nesse sentido deve encorajar atitudes orientadas à transparência, qualidade e bem-estar da sociedade. Consequentemente deve evitar danos às pessoas e aumentar a sensibilização dos efeitos da tecnologia na sua vida. Um líder é o responsável pela equipa que lidera, nesse sentido deve zelar por serem dados os devidos recursos e garantir assim a qualidade do ambiente do trabalho. Proporcionar oportunidades para os membros da organização / colaboradores aprenderem e crescerem profissionalmente. Estas oportunidades devem incluir experiências que demonstrem as consequências e limitações de determinados sistemas, e devem ser alertados para os perigos de abordagens simplificadas, da improbabilidade de poderem antecipar todas as condições de operação, da inevitabilidade de erros de software, da interação dos sistemas em diversos contextos, e outros relacionados com a complexidade da sua profissão, e ainda assim estarem confiantes em assumirem as responsabilidades do trabalho que realizam. Os sistemas devem ser utilizados para o bem comum da sociedade e assim devem ser tidos especiais cuidados com os sistemas que são integrados na infraestrutura da sociedade, como também serem tidos em consideração os efeitos de modificar ou retirar sistemas que tenham um impacto negativo na produtividade e na vida das pessoas. Deve um líder avaliar e identificar alternativas que mitiguem esses efeitos. Por último, cabe ao líder articular, aplicar e suportar-se em políticas e processos que reflitam a aplicação do *The Code*. A última área de atuação identificada é a Conformidade com o Código de Ética. Os sistemas de informação no futuro dependem essencialmente da componente técnica e da ética, na sua conceção, desenvolvimento e utilização. Relativamente aos profissionais que desenham e desenvolvem esses sistemas devem fazê-lo tendo presente os princípios do *The Code* e contribuir para o melhorar. Devem ser tomadas ações no sentido de resolver questões de ética quando identificadas e que violem os princípios de ética.

Tendo por base o código de ética e a conduta profissional definida pela ACM [13], torna-se necessário ter um instrumento que permita aferir em que medida este código está a ser aplicado e de que modo a conduta profissional está a ser cumprida. Para tal, foi proposta um processo composto por quatro passos, através do qual é possível aferir essas conformidades. O processo designa-se na língua inglesa pelo acrónimo CARE: *Consider*, *Analyze*, *Review*, *Evaluate*. Através da aplicação deste processo, é possível avaliar a conformidade, identificar e estabelecer ações a serem tidas em consideração e que sejam consistentes com o código de conduta profissional definido pelo *The Code* da ACM. Na Tabela 2, descrevemos resumidamente o processo CARE. O documento *The Code* não defini a metodologia de aplicação do processo de validação CARE, contudo acreditamos que este processo poderá ser utilizado iterativamente, permitindo envolver todas as partes interessadas, facilitando desse modo a identificação e correção de não conformidades, até ser encontrado um compromisso ético, social e legal entre as partes envolvidas e as ações a serem tomadas.

ISBN: 978-989-98434-9-3

***Tabela 2** – Resumo do processo CARE*

CARE	Descrição
Considerar *(**C**onsider)*	Nestes passos devem ser idenficados quem são os atores e os *stakeolders*. Quais os efeitos observáveis das ações ou decisões para esses *stakeholders*. E, o que pode ser considerado ou melhorado para uma maior compreensão do contexto situacional.
Analisar *(**A**nalyse)*	Identificar quais os direitos dos *stakeholders*. Direitos legais, sociais que possam influenciar e em que extensão. Quais os aspetos técnicos mais relevantes para as decisões dos atores. Quais os princípios do *The Code* que sejam mais relevantes no contexto situacional que se está a estudar. Quais os valores legais que devem ser tidos em conta: pessoais, institucionais, legais entre outros.
Rever *(**R**eview)*	Quais as responsabilidades, autoridade, praticas, ou políticas que podem moldar as decisões dos atores. Que potenciais ações podem alterar os resultados inicialmente estabelecidos.
Avaliar *(**E**valuate)*	Em que medida as decisões tomadas num contexto em particular, podem ser utilizadas em contextos similares. Quais as ações ou a falta delas que suportam ou violam o *The Code*. As ações tomadas num caso em particular, consideram os direitos e os impactos que têm em todos os *stakeholders*.

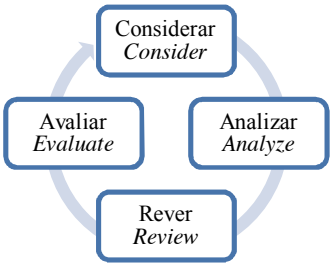

***Figura 1** - Processo CARE*

C. IEEE Ethical Design

O código de ética da IEEE [14] é composto por dez princípios gerais. Sendo o primeiro princípio de ética direcionado para primar a segurança, saúde e bem-estar dos indivíduos, este princípio deve estar alinhado com o desenvolvimento de práticas de design ético e sustentabilidade. O segundo princípio define que devem ser evitados os conflitos de interesse sempre que possível. O terceiro princípio apela a honestidade e realismo do trabalho a desenvolver pelos profissionais. O quarto princípio invoca a rejeição de suborno. O quinto princípio refere a importância do aumento de literacia digital e tecnológica da sociedade em geral. O sexto princípio é referente a necessidade de atualização e desenvolvimento de conhecimento e competências nas áreas tecnológicas. O sétimo princípio refere a necessidade de procurar aceitar e propor críticas honestas dos outros, bem como a de dar crédito aos contributos dos outros. O oitavo princípio invoca o tratamento cordial e a não discriminação baseada na raça, religião, género, deficiência, idade, nacionalidade, orientação sexual, identidade de género, ou expressão de género. O nono princípio declara a necessidade de evitar prejudicar os outros, a sua propriedade ou reputação. O décimo princípio estabelece a interajuda entre colegas no cumprimento do código de ética da IEEE [14].

A IEEE lançou a discussão publica um documento [15] com um conjunto de linhas diretoras sobre o alinhamento práticas de design ético e de sustentabilidade dos sistemas inteligentes. Neste documento definem-se princípios nos quais as tecnologias e os sistemas inteligentes devem priorizar o bem-estar dos seres humanos. Assim o documento [15] preconiza que a inteligência artificial ao preservar valores éticos e morais, permitem não só solucionar objetivos funcionais dos sistemas, mas também aumentar o nível de confiança entre os humanos. O documento tem três propósitos a saber. O primeiro deles é o impulsionar a discussão publica acerca do uso de sistemas inteligentes e autónomos de modo alinhado com os valores e com os princípios éticos dos seres humanos no seu contexto cultural. O segundo propósito deste documento é o de inspirar a criação de normas internacionais para certificar sistemas (IEEE P7000 ™). O terceiro propósito consiste na facilitação da criação de políticas nacionais e globais alinhadas com estes princípios.

O documento [15] define na sua base a articulação de três princípios de ética base para os sistemas inteligentes/autónomos (A/IS), independente de serem robots físicos ou sistemas de software. Os três princípios base para A/IS são os seguintes:

1. *"Incorporar os mais altos ideais de valorização e benefício humano como um superconjunto dos direitos humanos.*
2. *Priorizar os benefícios para a humanidade e o meio ambiente do uso de A/IS. Note-se que, estes não devem estar em substituição um do outro, um depende do outro. Priorizar o ser humano e o bem-estar não significa degradar o meio ambiente.*
3. *Mitigar os riscos e os impactos negativos, incluindo o uso indevido, à medida que o A/IS evoluem como sistemas sociotécnicos. Em particular, assegurando que os A/IS sejam responsáveis e transparentes.*" [15, p. 20].

Na versão inicial do documento [15] foram constituídas diversas comissões, estas comissões definiram um conjunto de 5 princípios gerais orientadores na conceção, desenvolvimento, implementação e monitorização destes sistemas A/IS. O documento define que em primeiro lugar os direitos humanos devem estar salvaguardados dos sistemas A/IS. Em segundo lugar os sistemas autónomos e inteligentes devem dar prioridade ao bem-estar, com sendo um resultado de todos os sistemas desenhados, segundo as regras geralmente aceites e métricas de referência. Em terceiro lugar a responsabilidade e corresponsabilidade (*accountability & responsability*) devem estar bem definidas, quer na parte de legislação, quem desenha e desenvolve os sistemas, quer na definição de ecossistemas de *stakeholders* múltiplos, quer em sistemas de registo que permitam identificar a evolução do desenvolvimento dos sistemas (i.e., objetivos de uso, dados de treino, algoritmos, gráficos de processos, diferentes níveis de funcionalidades, interfaces com o utilizador, objetivos de otimização onde se pondera ganhos vs. perdas). Em quarto lugar o documento define que deve existir transparência quer para os utilizadores perceberem o que os programas estão a fazer e porquê. A transparência é também importante para a certificação dos sistemas A/IS, bem como para verificação dos sistemas inteligentes em caso de acidente e em caso de deliberação e atribuição de responsabilidades legais. A transparência é também um aspeto importante para fortalecer o nível da confiança e aceitação da sociedade em geral. Em quinto lugar o documento refere a importância da criação de consciência dos usos dos sistemas autónomos e inteligentes, nomeadamente a consciência de que estes sistemas podem ser mal utilizados com fins para os quais não foram desenvolvidos, sublinhado a necessidade de serem ministradas ações de sensibilização sobra a ética e o uso de A/IS.

ISBN: 978-989-98434-9-3

As comissões que foram constituídas para debater o alinhamento da ética com o desenvolvimento de sistemas A/IS debateram e produziram relatórios de atuação em diversa áreas, nomeadamente na introdução de valores societais e culturais nos sistemas inteligentes, nas linhas de investigação e desenvolvimento, na identificação dos benefícios do uso de A/IS, no acesso e controlo dos dados pessoais, na área de construção de sistemas autónomos de armamento, nas áreas da economia, causas humanitárias e direito. Foram ainda constituídas comissões com o objetivo de definir linhas orientadoras para as áreas de computação afetiva, política, ética clássica, realidade mista e aumentada e bem-estar [15].

III. MAPEAMENTO DOS PRINCIPAIS ESTUDOS REALIZADOS

No presente estudo foi conduzido um trabalho empírico, que consistiu num estudo bibliométrico das principais referencias das duas bibliotecas digitais, ACM [8] e IEEE [11]. Foram realizadas pesquisas com a combinação do termo "Ética" e do termo composto "Inteligência Artificial" em ambos os motores de busca. De seguida foi realizado um tratamento de dados onde se verificou através da coocorrência das palavras-chave dos diversos artigos quais eram essas ocorrências e de que modo se relacionavam. Foram ainda identificados clusters de assuntos diferenciados entre os artigos publicados pela ACM e IEEE.

A. *Estudos realizados e indexados na ACM e IEEE em matéria de Etica e Inteligencia Artificial*

As associações internacionais de referência, na comunidade científica mundial, nas ciências da computação e engenharia, ACM [9] e IEEE [11], produziram e continuam a aprofundar investigação sobre aspetos da ética na computação. Neste sentido, entende-se que é pertinente identificar os aspetos que os diversos autores têm desenvolvido nos seus trabalhos científicos. Foi conduzido um estudo empírico com o objetivo de identificar os aspetos que até a data mereceram a atenção da comunidade de investigadores. Para tal fomos junto das duas bibliotecas digitais [8], [10] recolher dados sobre todos os artigos até Fevereiro de 2019 que tivessem indexados, quer nas palavras-chave, quer no titulo, em coocorrência os termos "Inteligência Artificial" e "Ética". Em primeiro lugar foram realizadas duas extrações

1) Na biblioteca digital da ACM *[8]* foi realizada uma pesquisa correspondente aos seguintes termos de busca (+"Artificial intelligence" +Ethics) o que corresponde a "query": { (+"Artificial intelligence" +Ethics) } com filtro (expandir) "filter": {owners.owner=GUIDE}. Com retorno desta pesquisa forma obtidos 807 resultados, artigos científicos, que estão indexados até Fevereiro de 2019 inclusive, sob os indexadores mencionados. De seguida utilizando uma ferramenta para estudos bibliométricos *[16]* para a identificação de clusters, foi conduzida uma análise de co-ocurrências de palavras-chave, com o método de fullcounting que tenha ocorrido pelo menos 5 vezes e as palavras que não tinham ligações forma ignoradas, com um número mínimo de 2 clusters. Os resultados obtidos permitiram identificar 4 clusters (Tabela 3), grupo 1- Big data, machine learning (a vermelho na Figura 2); grupo 2-Agentes (verde); grupo 3- robótica (azul) e o grupo 4-inteligencia artificial (amarelo).

Tabela 3 *– Resultados dos Clusters dos trabalhos da ACM*

Cluster 1 (8 itens)	Big Data Decision Making Explainable AI Fairness Machine Learning Natural Language processing Privacy Transparency	**Cluster 3 (6 itens)**	Design Ethics Robotics Robots Value Alignment Values
Cluster 2 (8itens)	Artificial agents Computational ethics Computer ethics Human-robot interaction Machine Ethics Multi-agent systems Robot ethics Trust	**Cluster 4 (5 itens)**	Ambient intelligence Artificial intelligence Autonomy Human-computer interaction Turing test

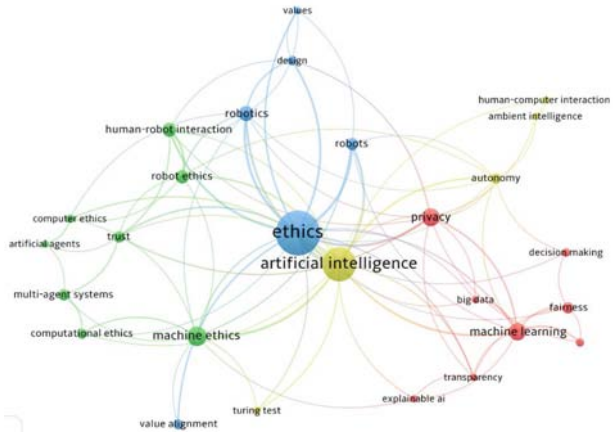

Figura 2*-Resultados da visualização dos clusters da ACM*

2) Na biblioteca digital da IEEE *[10]* foi realizada a seguinte pesquisa ((ethics) AND artificial intelligence), foram obtidos 258 artigos de revista e proccedings. Para esta análise foram utilizados os resumos. Com base nestes dados foram criados clusters, o processo do tratamento de dados foi o seguinte. Foi realizada uma contagem total em que o número de ocorrências fosse pelo menos 10 vezes, tendo sido identificados 112 termos relevantes e desses foram ainda selecionados 60% dos termos mais relevantes para a formação dos clusters. Dos resultados forma obtidos 3 clusters, repostados na Tabela 4. Da análise de clusters foram identificados três grupos. O primeiro grupo tem a ver com ética relacionada com a IA e suas aplicações (a vermelho na Figura 3). O segundo grupo tem a ver com ética ao nível da produção científica (verde). O terceiro sugere ética relacionada com machine learning (área especifica da IA) (azul). A figura 3 apresenta o modo como estão agrupados os 3 clusters.

IV. DISCUSSÃO

Vários estudos bibliométricos realizados [17]–[20] em diversas áreas desde a saúde ao marketing, utilizando o mesmo método, foram conducentes a resultados semelhantes ao presente estudo. Nos estudos referidos foram identificados clusters de estudos interligados pela coocorrência.

ISBN: 978-989-98434-9-3

***Tabela 4** – Resultados dos Clusters dos trabalhos da IEEE*

Cluster 1 (42 itens)	Action Agent Algorithm Approach Argument Artificial intelligence Automation Autonomous system Condition Context Data Development Ethic Ethics Experiment Field Framework Future Goal Human Interaction Issue Kind Machine Machine ethic Model Number Ontology Participant Question Right Risk Robot Role Task Technique Technology Time Trust Word Year	**Cluster 2 (14 itens)**	Author Content Education Expert commitee IEEE Option Paper Past reference Publication principle Reasonable effort Review Student University Violation
		Cluster 3 (9 itens)	Area Car Computer Engineering Machine Learning Philosophy Self Study Topic

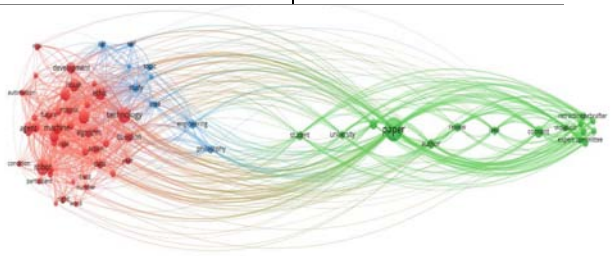

***Figura 3**-Resultados da visualização dos clusters da IEEE*

V. CONCLUSÕES

O principal objetivo deste estudo é o de identificar os principais desafios éticos relativos à utilização da inteligência artificial que se colocam as organizações e a sociedade em geral. Para tal através do estudo documental, fomos identificar as principais linhas orientadoras preconizadas pelas duas associações internacionais mais conceituadas. Conduzimos um estudo bibliométrico para mapear os principais assuntos estudados e publicados em artigos científicos onde ocorressem ética e inteligência artificial. Foram identificados 7 grupos distintos de assuntos estudados pela comunidade científica. Depreendeu-se pelos resultados do estudo que no momento nem todos os assuntos estão a ser cobertos pela comunidade científica. A enfase está mais focada na componente de produção e difusão de ciência, não se depreendendo dos resultados que haja um enfoque nas consequências da utilização de inteligência artificial. São necessários mais estudos com diversos âmbitos na utilização em contexto organizacional, bem como estudos que reportem o nível de consciencialização do uso de sistemas autónomos e inteligentes por parte da sociedade em geral. Tornando-se assim pertinente a realização de estudos, quer a nível individual, quer a nível organizacional, acerca dos impactos dos sistemas de inteligência artificial.

AGRADECIMENTOS

Este trabalho é parcialmente financiado por Fundos Nacionais através da FCT - Fundação para a Ciência e a Tecnologia no âmbito dos projetos de financiamento com as Referências UID/SOC/04521/2019; e UID/MULTI/4466/2019.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] N. Bostrom e E. Yudkowsky, «The ethics of artificial intelligence», em *The Cambridge Handbook of Artificial Intelligence*, K. Frankish e W. M. E. Ramsey, Eds. Cambridge University Press, 2014, pp. 316–334.

[2] E. Shekarian e A. Fallahpour, «Predicting house price via gene expression programming», *Int. J. Hous. Mark. Anal.*, vol. 6, n. 3, pp. 250–268, Jul. 2013.

[3] T. L. Beauchamp e N. E. Bowie, *Ethical theory and business*. Prentice Hall, 2001.

[4] R. B. Abidoye e A. P. C. Chan, «Valuers' receptiveness to the application of artificial intelligence in property valuation», *Pac. Rim Prop. Res. J.*, vol. 23, n. 2, pp. 175–193, Mai. 2017.

[5] M. Piteira, C. J. Costa, e M. Aparicio, «CANOE e Fluxo: Determinantes na adoção de curso online gamificado», *Revista Ibérica de Sistemas e Tecnologias da Informação (RISTI)*, pp. 34–53, 2017.

[6] M. Piteira e C. J. Costa, «Gamification: Conceptual framework to online courses of learning computer programming», em *12th Iberian Conference on Information Systems and Technologies (CISTI)*, 2017, pp. 1–7.

[7] M. Piteira, C. J. Costa, e M. Aparicio, «Computer Programming Learning: How to Apply Gamification on Online Courses?», *J. Inf. Syst. Eng. Manag.*, vol. 3, n. 2, p. 11, Abr. 2018.

[8] Association for Computing Machinery, «ACM Digital Library», 2014. [Em linha]. Disponível em: http://dl.acm.org/. [Acedido: 18-Fev-2014].

[9] «Association for Computing Machinery», 2019. [Em linha]. Disponível em: https://www.acm.org/. [Acedido: 06-Mar-2019].

[10] IEEE, «IEEE Xplore Digital Library», *Institute of Electrical and Electronics Engineers*, 2014. [Em linha]. Disponível em: http://ieeexplore.ieee.org/Xplore/home.jsp.

[11] «IEEE - The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.» [Em linha]. Disponível em: https://www.ieee.org/. [Acedido: 06-Mar-2019].

[12] J. Fischer, «Social Responsibility and Ethics: Clarifying the Concepts», *J. Bus. Ethics*, vol. 52, n. 4, pp. 381–390, Jul. 2004.

[13] Association for Computing Machinery, «ACM Code of Ethics and Professional Conduct». .

[14] «IEEE Code of Ethics». [Em linha]. Disponível em: https://www.ieee.org/about/corporate/governance/p7-8.html. [Acedido: 22-Fev-2019].

[15] «Ethically Aligned Design, Version 2 (EADv2) | IEEE Standards Association», *EADv2 - Ethics in Action*. [Em linha]. Disponível em: https://ethicsinaction.ieee.org/. [Acedido: 05-Mar-2019].

[16] «VOSviewer - Visualizing scientific landscapes», *VOSviewer*.

[17] A. Perianes-Rodriguez, L. Waltman, e N. J. van Eck, «Constructing bibliometric networks: A comparison between full and fractional counting», *J. Informetr.*, vol. 10, n. 4, pp. 1178–1195, Nov. 2016.

[18] F. J. Martínez-López, J. M. Merigó, L. Valenzuela-Fernández, e C. Nicolás, «Fifty years of the European Journal of Marketing: a bibliometric analysis», *Eur. J. Mark.*, vol. 52, n. 1/2, pp. 439–468, Jan. 2018.

[19] D. Holman, R. Lynch, e A. Reeves, «How do health behaviour interventions take account of social context? A literature trend and co-citation analysis», *Health (N. Y.)*, vol. 22, n. 4, pp. 389–410, Jul. 2018.

[20] T. Doleck e S. Lajoie, «Social networking and academic performance: A review», *Educ. Inf. Technol.*, vol. 23, n. 1, pp. 435–465, 2018.

ISBN: 978-989-98434-9-3

View publication stats