

UNIVERSITATEA „ALEXANDRU IOAN CUZA” DIN IAȘI
ȘCOALA DOCTORALĂ DE ECONOMIE ȘI ADMINISTRAREA
AFACERILOR
DOMENIUL DE DOCTORAT: INFORMATICĂ ECONOMICĂ

Etica tehnologiilor inteligenței artificiale implementate în mediul de afaceri

REZUMAT

Conducător de doctorat:

Prof. univ. dr. Gabriela Meșniță

Doctorand:

Constantin Alexandru Ciobanu

MULȚUMIRI

Demersul științific și elaborarea acestei teze de doctorat nu ar fi fost posibilă fără ajutorul și suportul unor oameni minunați, cărora doresc să le mulțumesc și să le ofer pe această cale recunoștința mea.

*În primul rând, aș dori să aduc alese mulțumiri și deosebită considerație, conducătorului de doctorat, **prof. univ. dr. Gabriela Meșniță**, pentru încrederea și înțelegerea acordată, pentru înaltul grad de profesionalism arătat pe parcursul îndrumării științifice. Sprijinul, răbdarea și sfaturile ce mi-au fost oferite, m-au provocat și motivat pe parcursul cercetării doctorale și au contribuit la formarea mea profesională.*

*Totodată, aș dori să mulțumesc și să manifest deplină recunoștință comisiei de îndrumare, **prof. univ. dr. Florin Dumitriu, conf. univ. dr. Laura-Diana Radu, conf. univ. dr. Valerică Greavu-Șerban**. Această teză de doctorat nu ar fi fost posibilă fără îndrumările, comentariile și provocările științifice aduse pe parcursul celor trei ani de cercetare.*

*În egală măsură țin să mulțumesc **colegilor și colectivului Școlii Doctorale de Economie și Administrarea Afacerilor** pentru dezbaterile constructive pe care le-am avut pe diferite teme științifice, pentru sfaturile și suportul moral oferit.*

*Nu în ultimul rând, aduc infinite mulțumiri **familiei mele**, în special Ramonei și celor două fete, Thea și Catinca. Răbdarea lor nemărginită, dragostea necondiționată și sprijinul moral m-au motivat și m-au determinat să depășesc orice obstacol, reușind, astfel, finalizarea tezei de doctorat.*

CUPRINSUL TEZEI DE DOCTORAT

Introducere

Capitolul 1. Analiza din perspectiva conceptuală a eticii privind inteligența artificială în mediul de afaceri

- 1.1 Etica din perspectiva digitală
- 1.2 Evoluția inteligenței artificiale și a provocărilor etice asociate
 - 1.2.1 *Scurt istoric al provocărilor etice în evoluția IA*
 - 1.2.2 *Tipuri de inteligență artificială*
 - 1.2.3 *Domenii de aplicabilitate a inteligenței artificiale*
- 1.3 Impactul inteligenței artificiale în mediul de afaceri
- 1.4 Teme asociate eticii IA implementate în mediul de afaceri
 - 1.4.1 *IA și autonomia în luarea deciziilor*
 - 1.4.2 *Prejudecata și echitatea în tehnologiile IA*
 - 1.4.3 *Confidențialitatea și securitatea IA*
 - 1.4.4 *IA și viitorul pieței muncii*
 - 1.4.5 *Reglementarea inteligenței artificiale*
- 1.5 Concluzii în urma analizei literaturii de specialitate

Capitolul 2. Perspective ale adoptării tehnologiilor IA în afaceri

- 2.1 Caracteristici ale adoptării tehnologiilor inteligenței artificiale
 - 2.1.1 *Calitatea seturilor de date*
 - 2.1.2 *Definirea unor cazuri de utilizare pentru IA (Business Use Case)*
 - 2.1.3 *Compatibilitatea IA cu mediul de afaceri*
- 2.2 Limitări ale adoptării tehnologiilor inteligenței artificiale
 - 2.2.1 *Limitări tehnologice*
 - 2.2.2 *Limitări organizaționale*
 - 2.2.3 *Limitări socioculturale*
- 2.3 Modele de adoptare a tehnologiei în contextul IA

2.3.1 Modele de referință utilizabile pentru adoptarea IA în afaceri

2.3.2 Adoptarea centrată pe valoare a tehnologiei IA în mediul de afaceri

2.4 Concluzii privind perspectivele adoptării IA în afaceri

Capitolul 3. Contextualizarea frameworkului pentru etica tehnologiilor IA

3.1 Metodologia folosită pentru identificarea principiilor etice specifice IA

3.2 Principii etice specifice tehnologiilor inteligenței artificiale

3.2.1 Principiul transparenței

3.2.2 Principiul dreptății și al corectitudinii

3.2.3 Principiul non-vătămării

3.2.4 Principiul responsabilității

3.2.5 Principiul binefacerii

3.2.6 Principiul libertății și al autonomiei

3.2.7 Principiul sustenabilității

3.2.8 Principiul solidarității

3.2.9 Principiile etice IA promovate de Uniunea Europeană

3.3 Provocări în operaționalizarea principiilor etice în IA

3.4 Ethics by Design și etica algoritmilor

3.5 Modele de etică pentru IA

3.6 Frameworkuri de etică pentru inteligența artificială: limite și oportunități

3.7 Concluzii privind frameworkurile pentru etica tehnologiilor IA

Capitolul 4. Framework de etică aplicabil tehnologiilor inteligenței artificiale implementate în mediul de afacere

4.1 Importanța frameworkului propus

4.2 Metodologia dezvoltării FEIA

4.2.1 Analiza modelului VCIO

4.2.2 Prima versiune a frameworkului de etică IA

4.2.3 Versiunea a doua a frameworkului de etică IA

4.3 Descrierea componentelor FEIA

- 4.3.1 AI Embedded Ethics by Design*
- 4.3.2 AI Desired State Configuration*
- 4.3.3 Interoperabilitatea celor două componente*
- 4.4 Aspecte funcționale ale frameworkului de etică în IA
 - 4.4.1 Integrarea principiilor etice la nivelul FEIA v2*
 - 4.4.2 Seturi de date pentru un model de antrenament etic*
 - 4.4.3 Proces pentru validarea eticii modelului de antrenament*
 - 4.4.4 Rolul funcțional al APIMngmtSvcS în cadrul frameworkului*
 - 4.4.5 Monitorizarea și validarea elementelor de etică în IA*
- 4.5 Concluzii privind frameworkul de etică pentru IA

Concluzii finale

Limite și direcții de cercetare

Diseminarea activităților de cercetare

Referințe bibliografice

Introducere

Tehnologia a reprezentat dintotdeauna un factor determinant în ceea ce privește evoluția societăților și a mediilor organizaționale de afaceri. Printre interesele studiilor de cercetare asupra tehnologiei s-au numărat, de cele mai multe ori, aspecte ale impactului pe care tehnologia îl poate avea asupra comunităților și a mediului economic.

Din perspectivă istorică, se poate observa evoluția procesului de transformare din domeniul tehnologiei. Momente semnificative din punct de vedere tehnologic, precum *Revoluția Industrială*, care poate fi asociată cu dezvoltarea primelor modele de afaceri, apoi progresul generat de apariția Internetului și schimbările semnificative din ultimul deceniu (tehnologii noi de inteligență artificială, *cloud computing*, *Blockchain* etc.), reprezintă momente care au generat provocări importante în economie, dar și în societate. Cea mai recentă provocare, dar în același timp inovație tehnologică, este aceea reprezentată de inteligența artificială (IA), care, prin diferitele forme de aplicabilitate, redefiniște sau chiar introduce concepte noi legate de productivitate în afaceri.

Fie că discutăm despre soluții de automatizare, de învățare automată, de procesare a limbajului sau despre augmentare a realității, inteligența artificială, în sens larg, schimbă deja modul în care companiile își desfășoară activitățile specifice afacerii. Scopul utilizării inteligenței artificiale poate varia în funcție de contextul de aplicabilitate. În general, utilizarea IA pleacă de la ideea de a replica sau chiar de a augmenta inteligența umană. Astfel, se pot automatiza anumite sarcini repetitive sau se pot obține predicții, sau chiar recomandări personalizate pentru situații specifice. Ca urmare a inovării tehnologice continue, regăsim astăzi tehnologiile IA utilizate în scopuri multiple. De cele mai multe ori, inteligența artificială este folosită ca un instrument tehnologic care, cu ajutorul intervenției umane, poate rezolva probleme complexe, permițând oamenilor să se concentreze mai mult pe creativitate și inovare. Valorificând capacitățile IA, companiile pot îmbunătăți eficiența, luarea deciziilor, experiențele utilizatorilor și pot, în definitiv, stimula inovația

în diferite domenii. În acest context, inteligența artificială arată un potențial practic nelimitat, de care poate beneficia întreaga societate, dar mai ales mediul de afaceri, care poate capitaliza o cotă de piață importantă prin eficientizarea și automatizarea anumitor procese de afaceri.

Totodată, în contextul avansului pe care IA îl are, apar tot mai multe dezbateri publice inițiate fie de mediul academic, fie de autorități guvernamentale sau de organizații ne-guvernamentale, sau chiar de societatea civilă, referitoare la aspectele de etică asociate acestor tehnologii.

Cercetarea propune un studiu mixt asupra aspectelor de etică asociate tehnologiilor IA, în contextul utilizării acestora de către mediul de afaceri. Dat fiind faptul că inteligența artificială are o dezvoltare fulminantă (mai ales în ultimul an, prin creșterea aplicațiilor de IA generativă – vezi ChatGPT), considerăm esențială înțelegerea beneficiilor aduse, dar și explorarea și prevenirea riscurilor asociate. Provocarea este dată tocmai de faptul că acest progres tehnologic al IA nu trebuie oprit sau temporizat, dar considerăm esențială dezvoltarea unui cadru propice, în care tehnologiile pot să evolueze în mod inovator și în concordanță cu respectarea cerințelor de etică, care pot varia în funcție de contextul utilizării IA.

Stadiul actual al cercetării privind etica aplicabilă IA în mediul de afaceri constată o dinamică în creștere în ceea ce privește numărul de articole, frameworkuri, regulamente care propun norme etice aplicabile IA. Se recunosc tot mai mult implicațiile etice ale tehnologiilor IA în diferite sectoare de afaceri și există nevoia de abordare a provocărilor etice asociate dezvoltării, implementării și utilizării acestora (Floridi & Cowls, 2019). Cu toate acestea, menționăm că, din punct de vedere istoric, tehnologia, în general (subscriem aici și inteligența artificială), deși la momentul apariției a generat provocări și contestări inclusiv de natură etică, a condus la progres științific și dezvoltare economică.

Cercetătorii explorează dimensiunile etice ale IA în mediul de afaceri, în încercarea de a înțelege riscurile și beneficiile potențiale, precum și impactul adoptării IA la nivelul societății. Pe de o parte se

investighează aplicabilitatea unor norme etice la nivelul IA, precum prejudecata, echitatea, responsabilitatea și transparența, dar și impactul IA asupra forței de muncă sau asupra luării deciziilor (Jobin & Vayena, 2019). În acest context, se constată un trend pozitiv în literatura de specialitate, care pune accent pe dezvoltarea unor frameworkuri de etică, linii directoare și practici recomandate pentru integrarea etică a IA în procesele de afaceri.

Cercetarea științifică referitoare la etica aplicată IA, presupune o abordare interdisciplinară, care să aducă împreună experți din diferite domenii precum informatică, etică, drept, științe sociale, managementul afacerilor. O astfel de abordare favorizează o înțelegere cuprinzătoare a provocărilor etice și permite dezvoltarea de soluții holistice care să adreseze etica în IA.

În egală măsură, reținem faptul că studiul eticii IA, ca domeniu de cercetare, este și va fi influențat de evoluția cadrului legal și de reglementare. Guvernele și organizațiile de reglementare recunosc importanța eticii IA și dezvoltă frameworkuri și linii directoare pentru a asigura o implementare responsabilă a IA în afaceri. Cu toate acestea, este nevoie de o explorare multidimensională pentru a înțelege care sunt sau care pot fi practicile responsabile și etice referitoare la implementarea IA în mediul de afaceri, precum și aspectele tehnice legate de implementarea efectivă a acestor practici de etică în IA.

Această cercetare promovează practici responsabile și etice aplicabile dezvoltării IA, și care pot fi operaționalizate în funcție de contextul (de exemplu social, economic, politic, cultural) de utilizare al IA.

Etica în inteligența artificială joacă un rol esențial în mediul de afaceri. Pe măsură ce tehnologia IA avansează și se integrează în diferite industrii, considerațiile etice devin esențiale pentru o implementare responsabilă, sustenabilă în timp și benefică a IA (Floridi și Cowsls, 2019). Totodată, este vital ca afacerile să dezvolte și să utilizeze tehnologii IA într-un mod care să respecte drepturile individuale, să promoveze

echitatea și să minimizeze prejudecățile și discriminarea (Jobin, Ienca și Vayena, 2019).

Astfel, plecând de la contextul descris, **întrebarea fundamentală a cercetării** este referitoare la *Cum putem valida etica aplicată tehnologiilor inteligenței artificiale implementate în mediul de afaceri?*

Considerăm că abordarea eticii aplicate în inteligența artificială poate contribui la reducerea riscurilor legale și de reglementare. Prin adoptarea unor practici etice solide, afacerile pot respecta cerințele legale și pot evita sancțiunile, menținând în același timp o relație pozitivă cu autoritățile de reglementare.

Întrebările secundare care vin au ghidat cercetarea sunt:

Care sunt perspectivele etice în contextul implementării tehnologiilor de inteligență artificială?

Ce modele pot fi folosite în vederea adoptării IA în mediul de afaceri?

Care sunt mecanismele de monitorizare a eticii tehnologiilor inteligenței artificiale implementate în companii?

Cum pot fi transpuse principiile etice la nivel algoritmic, ca parte din modelul de antrenament al IA?

Cum se pot defini normele etice astfel încât acestea să poată fi înțelese corect de către un dezvoltator sau practician IA?

Prin răspunsurile pe care le vom oferi întrebărilor menționate urmărim îndeplinirea obiectivelor specifice ale cercetării și, în final, atingerea scopului detaliat în secțiunea următoare.

Scopul cercetării doctorale este acela de a crea un **cadru de lucru (framework)** conceptual, adaptabil, care să poată fi utilizat pentru a valida etica specifică tehnologiilor IA implementate în mediile de afaceri. Acesta va fi numit în continuare **Framework de Etică pentru IA** (abreviat pe parcursul lucrării **FEIA**) și reprezintă o modalitate practică prin care o companie va putea gestiona sau asigura o guvernanta etică a tehnologiilor IA în diferitele etape ale adoptării acestora. Totodată, cercetarea urmărește ca prin acest framework să ofere soluții punctuale în

ceea ce privește mecanismele care pot fi utilizate în operaționalizarea principiilor etice și detectarea abaterilor etice la nivelul algoritmilor IA în diferite etape (dezvoltare, implementare, post-implementare tehnologie IA).

Utilitatea implicită a acestui demers științific constă în prezentarea unor metode și instrumente de lucru (ca parte integrantă din framework) care pot ajuta în operaționalizarea principiilor etice la nivelul tehnologiilor IA. Se va sublinia astfel nevoia unei schimbări de paradigmă prin trecerea de la *abordarea normativă* (enumerarea unor norme etice pe care o tehnologie IA trebuie să le respecte) la o *abordare pragmatică* (cum putem implementa și defini principii etice la nivel algoritmic).

Nevoia unui framework care să valideze etica unei tehnologii IA, vizează crearea unui context favorabil acolo unde tehnologiile emergente se dezvoltă în concordanță cu norme etice specifice și cu nevoile individuale de intimitate, securitate și conformitate. Un astfel de framework nu reprezintă un blocaj în calea progresului tehnologic, ci, dimpotrivă, poate ajuta în adoptarea de tehnologii responsabile și sustenabile.

În figura 1 este prezentată o abordare schematică a proiecției de cercetare care cuprinde întrebarea fundamentală, scopul, obiectivele și întrebările secundare, precum și abordarea metodologică.

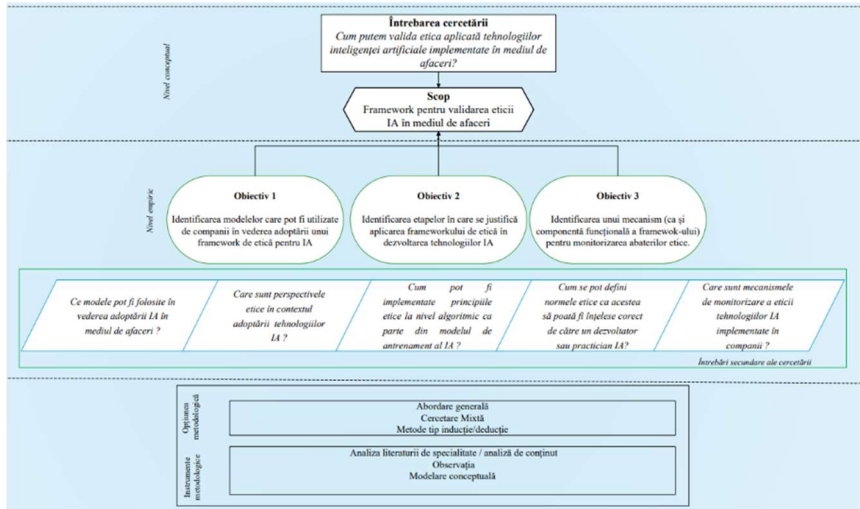


Figura 1. Structura și abordarea metodologică a cercetării

Astfel, plecând de la întrebarea fundamentală a cercetării, scopul stabilit este acela al dezvoltării unui framework conceptual (FEIA) care să valideze elementele de etică asociate tehnologiilor IA permițând companiilor implementarea controlată a acestora. Acest framework este conceptual, având un caracter generalizator din punctul de vedere al aplicabilității. Totodată, acesta este adaptabil și personalizabil conform cerințelor/nevoilor etice specifice unei afaceri. Considerăm esențial ca un framework pentru etica IA să aibă componente care permit:

- a) *Definirea* normelor etice din faza de design a IA, ceea ce numim în cercetarea noastră: Etica încorporată în designul IA (din engleză *AI Embedded Ethics by Design – AI EED*).
- b) *Posibilitatea de configurare și adaptare continuă a IA*. Ne referim prin aceasta la faptul ca o companie să poată defini o stare dorită sau ideală pentru o tehnologie IA și apoi funcție de aceasta se poate proiecta sau antrena respectiva tehnologie pentru a îndeplini această

stare (din engleză *AI Desired State Configuration – AI DSC*). Ideea a fost experimentată de Microsoft printr-un proiect similar aplicat soluțiilor cloud Microsoft 365 (Microsoft365 DSC, 2020). Această componentă trebuie de altfel să permită mecanisme practice prin care eventualele abateri etice, care nu sunt definite în starea de configurare inițial antrenată, să poată fi automat depistate și rezolvate oportun.

Obiectivele secundare pe care le vom aborda în atingerea scopului au fost:

1. **Identificarea etapelor în care se justifică aplicarea unui cadru etic în dezvoltarea tehnologiilor IA.** Acest obiectiv își propune să exploreze și să definească etapele din procesul de dezvoltare al IA, în care considerațiile etice devin cruciale. Implică analizarea diferitelor etape, cum ar fi colectarea datelor, proiectarea algoritmilor, antrenarea modelelor, implementarea și monitorizarea continuă, pentru a determina când și cum trebuie aplicate frameworkurile etice.
2. **Identificarea modelelor care pot fi utilizate de companii în vederea adoptării frameworkului de etică pentru IA.** Acest obiectiv urmărește să exploreze factorii care determină decizia unei afaceri de a adopta un cadru etic. Totodată, prin acest obiectiv se urmărește explorarea modelelor de adoptare a tehnologiei care pot fi utilizate în adoptarea IA de către companii. Implică analiza factorilor precum cerințele legale și reglementare, cerințele pieței, gestionarea reputației și valorile organizaționale, pentru a înțelege procesul decizional și pentru a promova o mai largă adoptare a frameworkurilor de etică în dezvoltarea IA.
3. **Identificarea unui mecanism (ca o componentă funcțională a unui framework) de monitorizare a abaterilor etice.** Acest obiectiv se concentrează asupra importanței implementării unui sistem de monitorizare care să poată detecta și aborda problemele etice pe parcursul ciclului de viață al tehnologiilor de IA. Implică examinarea mecanismelor existente de monitorizare și propunerea de strategii eficiente pentru detectarea și reducerea abaterilor etice în tehnologiile IA.

Prin abordarea acestor obiective, lucrarea contribuie la înțelegerea aspectelor etice aplicabile în dezvoltarea IA, oferind îndrumări pentru implementarea practică a normelor etice la nivelul IA în mediul de afaceri. Totodată, lucrarea oferă o radiografiere comprehensivă asupra impactului tehnologiilor menționate în mediul de afaceri, din perspectiva eticii digitale. Acest demers științific propune o trecere din *zona normativă* (nevoia de etică digitală) în *zona practică* (operaționalizarea eticii în IA), identificând modele de adoptare tehnologică și mecanismele care ar putea facilita o astfel de translație. Astfel, cercetarea caută să ofere direcții în vederea operaționalizării principiilor eticii în nevoile tehnologice specifice IA ale mediului de afaceri.

Considerând complexitatea pluridisciplinară a temei propuse și implicațiile acesteia de natură socioeconomică, tehnologică, dar și scopul care vizează crearea unui framework conceptual, demersul metodologic abordat este unul *mixt*. Se vor urmări, astfel, alegeri metodologice infralogice (caracteristice mai degrabă cercetării calitative), dar, subsecvențial, și a celor logice pentru a putea dezvolta în mod rațional soluții care să răspundă întrebării fundamentale de cercetare.

Așadar, se vor regăsi următoarele opțiuni metodologice și raționamentul utilizat în luarea deciziei.

Pentru primul obiectiv se va lua în considerare o abordare care se bazează pe *opțiunea metodologică (generalistă), de tip abductiv (cauză-efect) și inductiv*, astfel încât se analizează impactul tehnologiilor de inteligență artificială în mediile de afaceri digitalizate. Abordarea va fi generalistă și diacronă, în sensul în care se va urmări și evoluția în timp a conceptului de inteligență artificială.

Abducția va fi folosită ca o primă metodă, pentru că se vor urmări identificarea unor cauze (*ce pot genera implementarea inteligenței artificiale*) pentru a explora consecințele și, totodată, impactul pe care inteligența artificială îl are asupra companiilor. În această etapă se vor formula ipotezele de lucru centrate pe inteligență artificială în raport cu etica conceptuală.

Abordarea de tip cauză-efect este folosită complementar pentru a putea surprinde o primă posibilă concluzie, după ce se vor analiza date (cum ar fi indicatori de eficiență tehnologică sau de productivitate) referitoare la impactul inteligenței artificiale post-implementare în anumite modele de afaceri.

Pentru cel de-al doilea obiectiv se va urmări opțiunea metodologică care să facă referire la obiectul eticii IA implementate în afaceri. Se va folosi metoda deducției, intenția fiind de a explora frameworkuri de etică aplicabile IA deja existente. Cercetarea va explora modelele și etapele în care se pot defini principii etice la nivelul structurilor algoritmice care stau la baza dezvoltării unei tehnologii de inteligență artificială. Scopul va fi cel al studierii eticii ca parte integrantă a unui framework de etică aplicabil IA.

Deducția va fi folosită aici tocmai pentru a defini o tipologie a eticii digitale și a structura o suită de modele validabile/testabile, care să poată fi folosite în demersul tehnologic de adoptare a unei tehnologii de inteligență artificială, care să poată răspunde unor cerințe etice.

Pentru cel de-al treilea obiectiv se va urmări identificarea unor mecanisme tehnice post-implementare a IA, care să poată depista eventualele abateri etice în raport cu configurația inițială a modelelor de antrenament a IA. Vor fi utilizate metode deductive, precum și tehnici metodologice, precum observația și analiza calitativă, pentru a propune un astfel de mecanism ca parte din frameworkul creat.

Un aspect important de menționat aici este că metodologia care a stat la baza creării, prin modelare conceptuală, a frameworkului propus va fi detaliată în capitolul patru, acolo unde se descriu funcțional elementele și componentele FEIA.

Cercetarea, prin caracterul său pluridisciplinar, propune explorarea conceptelor de etică (în sens larg) care pot fi asociate cu tehnologia, aceasta reprezentând astăzi o prioritate atunci când se urmăresc premisele dezvoltării economice și inovării. De altfel, *motivația* alegerii acestui subiect de cercetare este dublu justificată. Considerăm, pe de o parte, necesară analiza modului de implementare a IA în companii,

subliniind beneficiile și valoarea adăugată reală, dar și potențialele riscuri de natură etică asociate. De cealaltă parte, propunem o soluție concretă, la nivel conceptual, care să poată fi utilizată în managementul eticii în contextul adoptării IA în mediul de afaceri.

Considerăm oportun ca, în contextul inteligenței artificiale, companiile să beneficieze de un *framework* adaptabil și practic, prin care să se poată valida și operaționaliza principiile etice care pot fi asociate IA într-un context dat (fie el economic, social, cultural).

În lucrarea de doctorat, tematica cercetării este elaborată pe parcursul a patru capitole, urmărind obiectivele menționate. Astfel, cercetarea pleacă de la analiza literaturii de specialitate referitoare la etică, în general, și etică aplicată în contextul inteligenței artificiale, pentru ca ulterior să reliefăm principiile etice relevante pentru IA, precum și principalele teme etice asociate implementării acesteia în mediul de afaceri. Pentru a se înțelege modalitățile și caracteristicile specifice implementării IA în mediul de afaceri se vor analiza modelele de adoptare a tehnologiei pe care companiile le pot utiliza în vederea adoptării IA. Nu în ultimul rând, se explorează principalele frameworkuri de etică și principalele limitări pe care acestea le au în privința operaționalizării la nivelul tehnologiilor IA.

Capitolul 1. Analiza din perspectiva conceptuală a eticii privind inteligența artificială în mediul de afaceri

Primul capitol este dedicat analizei principalelor concepte (*etică* și *intelență artificială*) prezentate în cercetarea doctorală, precum și stabilirea legăturilor dintre acestea, din perspectiva implementării lor în mediul de afaceri. Astfel, tematica principală a cercetării se concentrează pe analiza conceptului de *etică* și a celui de *intelență artificială*. Specific, în acest capitol se explorează aspecte ce țin de aceste două concepte, atât din punct de vedere definitoriu, cât și evolutiv. Totodată, încadrăm în această etapă a cercetării, principalele aspecte referitoare la impactul pe care IA îl are în mediul de afaceri, precum și cele mai

relevante teme etice asociate IA, așa cum sunt ele prezentate de literatura de specialitate.

În urma analizei literaturii științifice, constatăm în acest capitol, că adoptarea și utilizarea IA în afaceri nu presupune doar un comportament responsabil al unor anumiți actori implicați (ca de exemplu dezvoltatori de cod sursă sau analiști de date). Pentru a obține *valoare adăugată din IA* este nevoie de strategie și de un cadru de lucru care să permită validarea etică continuă într-un context dat. Eventualele problematice etice asociate cu inteligența artificială în afaceri pot reprezenta diferite riscuri pentru acele afaceri, cum ar fi eșecul lansării unor produse, problemele legale sau prejudiciul adus reputației companiei (Talgala, 2022). În lipsa unor reglementări etice referitoare la IA, există posibilitatea ca aceste tehnologii să fie utilizate în mod greșit, generând o serie de nedreptăți (Passanen, 2022), cum ar fi: manipularea, încălcarea dreptului la intimitate și perpetuarea prejudecăților și discriminării. În plus, lipsa transparenței din anumite tehnologii IA poate genera neîncredere și suspiciune în rândul utilizatorilor, care pot pune sub semnul întrebării fiabilitatea rezultatelor și recomandărilor dacă nu înțeleg modul în care funcționează tehnologiile IA. (Fahmideh, et al., 2021).

Se constată astfel o gamă largă de implicații etice ce pot fi asociate tehnologiilor IA, care vizează nu doar modul de dezvoltare software a IA, dar și modul de implementare și utilizare a acesteia.

Astfel, s-a constatat că în general, definirea termenului *etică* presupune explicarea unor concepte conexe, interdependente, care pot fi asociate contextual cu ceea ce se explică a fi etic sau moral. În cercetarea noastră, folosim definiția *eticii* (în sens larg) conform căreia este o ramură a filosofiei care se ocupă cu studiul a ceea ce este bun sau rău, drept sau greșit din punct de vedere moral, termenul putând fi asociat cu orice teorie a valorilor și principiilor morale (Singer, 2021; Velasquez, et al., 2010). Totodată, prin etică putem înțelege standardele de dreptate și corectitudine care pot prescrie ceea ce oamenii ar trebui să facă, de obicei, în termeni de drepturi, obligații și beneficii pentru societate (Canada Gov,

2015). În egală măsură, *etica* se poate referi și la studiul și dezvoltarea unor seturi de principii sau standarde etice într-un context dat, sau chiar la codurile de conduită care stabilesc ce comportament este corect și ce este greșit pentru indivizi și grupuri (Centrul de etică, 2020).

Inteligența artificială poate fi definită prin raportarea la programe, algoritmi, sisteme și mașini care pot demonstra inteligență (Shankar, 2018). Analizând literatura de specialitate se observă că o definiție universal aplicabilă pentru inteligența artificială nu a fost agreată între cercetători, din cauza aspectului interdisciplinar al termenului. Cu toate acestea, în general, inteligența artificială poate fi văzută ca fiind: „capacitatea unei mașini de a imita funcții umane, cum ar fi raționamentul, învățarea, planificarea și creativitatea” (Parlamentul European, 2023).

Termenul de inteligență artificială este adesea utilizat pentru a descrie procese de creare a unor sisteme tehnologice echipate cu procese cognitive asociate în mod normal cu oamenii, cum ar fi capacitatea de a raționa, de a deduce sau de a învăța din experiențele anterioare (Copeland, 2020). Pe de altă parte, anumite tehnologii IA au ajuns la nivelul de competență al experților umani în executarea anumitor sarcini specifice, rezultând astfel aplicații diverse precum diagnosticul medical, motoare de căutare inteligente și recunoașterea vocală, a imaginilor sau a textelor. Inteligența artificială sau, în sens general, teoria dezvoltării sistemelor de tip „*computing*”, care pot efectua activități specific inteligenței umane, sunt considerate ca revoluții în curs care transformă societățile și știința (Harari, 2017).

O altă modalitate de a descrie inteligența artificială este aceea în care acest concept nu depinde de tehnologiile caracteristice, ci mai degrabă de scopurile asociate utilizării și implementării IA. În acest sens, de cele mai multe ori, companiile vor folosi tehnologii de IA pentru eficientizarea proceselor de afaceri sau a strategiilor de marketing prin accesul facil la date în vederea unei interacțiuni mai eficiente cu clienții sau cu propriii angajați (Davenport și Ronanki, 2018). De altfel, cercetarea noastră va reveni asupra acestui aspect referitor la

eficientizarea proceselor de afaceri, ca scop fundamental în utilizarea tehnologiilor IA.

Inteligența artificială reprezintă un domeniu vast și în continuă evoluție, care se concentrează pe crearea și implementarea de algoritmi și modele capabile să simuleze anumite aspecte ale inteligenței umane. Există mai multe tipuri de IA, fiecare cu caracteristici și aplicabilități specifice. Conform Comisiei Europene (2020), inteligența artificială este clasificată în două tipuri:

- *software*: asistenți virtuali, programe informatice de analiză a imaginilor, motoare de căutare, sisteme de recunoaștere vocală și facială;
- *IA încorporată*: roboți, automobile autonome, drone, internetul obiectelor.

Sigur că această tipologie este una generalistă și simplificată, dar acoperă majoritatea formelor de IA utilizate de fiecare dintre noi în viața zilnică, cât și de companii, în scop de productivitate în afaceri.

Cercetarea noastră utilizează termenul de *tehnologii ale inteligenței artificiale* făcând referire la multitudinea de sisteme și soluții care prezintă capabilități IA în diferite forme. În general, distingem două mari categorii asociate inteligenței artificiale: *Învățare Automată (ML, din engleză Machine Learning)* și *Învățare Profundă (DL, din engleză Deep Learning)*.

Cele două categorii sunt tratate de literatura de specialitate ca domenii distincte, dar interconectate din punctul de vedere al aplicabilității. În timp ce există anumite similitudini între ele, fiecare categorie prezintă caracteristici și metodologii de dezvoltare/implementare unice.

Legătura dintre *Inteligența artificială (IA)*, *Învățarea Automată (ML)* și *Învățarea Profundă (DL)* sunt adesea interpretate greșit, generând confuzie în utilizarea de zi cu zi, în special în anumite domenii. O înțelegere mai clară poate fi obținută prin vizualizarea acestora ca pe niște cercuri concentrice (Dubey, 2018; Rouse, 2018a) (figura 2). Cercul mai mare, *IA*, reprezintă conceptul fundamental – capacitatea computerelor de

a imita comportamentul uman și de a replica abilitățile de luare a deciziilor umane (Copeland, 2016; Finlay, 2017). În interiorul IA, ca un subset, se află *Învățarea Automată*, care implică algoritmi care învață din date și fac predicții pe baza acestora (Dubey, 2018). *Învățarea Profundă* este plasată în interior, ca un subset specializat al *Învățării Automate*, concentrându-se pe rețele neuronale multistrat. (Rouse, 2018b; Hao, 2019). Această conceptualizare ierarhică oferă o înțelegere nuanțată a naturii distincte, dar interconectate a *IA*, *ML* și *DL*.

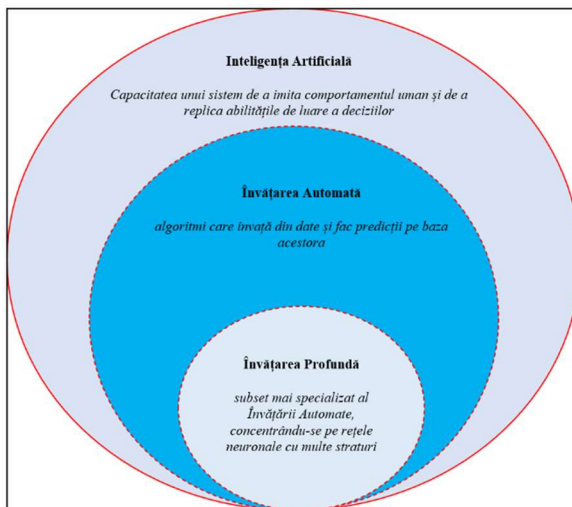


Figura 2. Principalele componente care alcătuiesc tehnologiile inteligenței artificiale

Din figura 2 se poate constata că *inteligenta artificială*, în sens larg, este domeniul cel mai amplu, care cuprinde dezvoltarea sistemelor inteligente capabile să efectueze sarcini care necesită, în mod obișnuit, inteligență umană (Russell & Norvig, 2016). Aceasta implică crearea de algoritmi și sisteme capabile să perceapă, să raționeze, să învețe și să ia decizii autonom. IA include diverse subdomenii, printre care se numără ML și DL.

Fiecare dintre aceste ramuri, poate avea mai multe domenii de aplicabilitate practică, acestea reprezentând un mod unic în care

tehnologia IA poate fi utilizată pentru a automatiza sarcini, interpreta date complexe sau interacționa cât mai natural cu oamenii. Totodată, companiile, în funcție de specificul proceselor de afaceri, pot implementa tehnologii IA care adresează unul sau mai multe domenii de aplicabilitate, creând ecosisteme de tehnologii IA. Astfel, când aceste tipuri de tehnologii IA sunt proiectate, pentru diferite domenii de aplicabilitate, va fi nevoie de o abordare etică unică.

Cercetarea noastră surprinde, însă, nevoia unui framework de etică universal, dar adaptabil pentru fiecare situație unică a implementării IA.

Inteligența artificială este considerată un element central al transformării digitale și a devenit o prioritate inclusiv pentru Uniunea Europeană, preconizându-se că viitoarele aplicații vor aduce schimbări enorme, însă IA este deja prezentă în viața noastră de zi cu zi. (Rezoluția 2088-INI a Parlamentului European, 2019). Transformarea digitală a devenit esențială pentru companii și este abordată ca un proces prin care acestea pot rămâne competitive tehnologic (Astapciks, 2023) (a se reține faptul că tehnologia este un domeniu în continuă evoluție). Motivele pentru care companiile decid să se transforme digital pot fi multiple, dar de cele mai multe ori acestea se rezumă la avantajul competitiv în piață, eficientizarea productivității, securitate, reducere de costuri, răspuns la nevoile clienților finali sau ale partenerilor.

IA are un potențial imens atât din punct vedere tehnologic, cât și din perspectiva capitalizării economice. Însă acest potențial este reprezentat și de provocări etice din ce în ce mai relevante. În această secțiune propunem o explorare a literaturii pentru a surprinde principalele teme etice care rezultă din implementarea și integrarea pe scară largă a IA în afaceri. Considerațiile etice cheie se învârt în jurul responsabilității sistemelor de IA tot mai autonome, în special în sectoare sensibile, unde luarea deciziilor poate avea consecințe ce pot schimba viața indivizilor sau societatea. Pe de o parte principalele teme etice vizează provocările legate de prejudecățile inerente ce pot exista în algoritmi IA și care pot amplifica inegalitățile societale. Pe de altă parte, unicitatea aplicațiilor de

IA necesită o examinare riguroasă a implicațiilor privind confidențialitatea și securitatea datorită colectării și utilizării substanțiale de date (Floridi & Taddeo, 2016). Cercetarea va explora, de asemenea, aspectele etice asociate cu influența transformatoare a IA asupra piețelor muncii. Totodată, în această secțiune subliniem necesitatea pentru o reglementare eficientă și adaptivă a IA, căutând un echilibru între încurajarea inovației și atenuarea riscurilor (Bostrom & Muller, 2016). În plus, se analizează implicațiile societale ale interacțiunilor om-IA, explorând efectele emoționale, psihologice și sociologice. Constatările subliniază esențialitatea discursului și cercetării continue în etica IA pentru a asigura că avansarea tehnologică se aliniază cu normele și valorile societale (Etzioni și Etzioni, 2017).

Ca urmare a acestei analize holistice a literaturii de specialitate, efectuate în primul capitol, am identificat cinci teme principale, recurente, asociate eticii IA (tabelul 1). În secțiunile următoare, fiecare temă va fi prezentată subliniind elementele pozitive, dar și provocările de care ar trebui să țină cont companiile care implementează tehnologii IA în procesele de afaceri.

Tabel 1. Teme etice asociate inteligenței artificiale conform analizei literaturii de specialitate

Dimensiuni etice ale IA	Autori relevanți
IA și autonomia în luarea deciziilor	Doshi-Velez, F. & Kim, B. (2017); Miller (2019); (Angeschmid et al., 2022)
Prejudicierea și echitatea în IA	Barocas, S. & Selbst, A. D. (2016); Dwork, C. et al. (2012)
Confidențialitatea și securitatea IA	Sweeney, L. (2013); Floridi, L. (2014)
IA și viitorul pieței muncii	Russell, S. et al. (2015)
Reglementarea IA	Kroll, J. A., Huey, J., Barocas, S. (2017); Wagner, B. (2018)

Acest prim capitol a explorat literatura de specialitate pentru a identifica principalele dimensiuni etice ale inteligenței artificiale și subiectele adiacente legate de tema de cercetare. Contextul globalizării informației, problemele de mediu sau economiile preponderent digitalizate, unde consumatorii sunt influențați de tehnologie în alegerile lor, reprezintă provocări care necesită reacții holistice. Transformarea digitală prin tehnologii de inteligență artificială presupune că tehnologiile pot răspunde acestor provocări prin inovație. Ideea este susținută de unele studii și statistici (Kesidou & Sovacool, 2019; Soni et al., 2020; Minevich, 2022; Zapanta, 2023; Perifanis & Kitsios, 2023).

Cu toate acestea, în baza analizei literaturii se conturează tot mai mult nevoia abordării elementelor de etică în IA. Se poate observa că abordarea eticii în IA este mai degrabă conceptuală, abstractizantă, generalizatoare cu focus pe constatarea unor nevoi de etică și oferirea de soluții mai puțin pragmatice.

Una dintre oportunitățile de cercetare în acest sens este tocmai aceea că în studiul eticii IA este nevoie și de explorarea aspectelor tehnice care stau la baza acestor soluții și de nișarea pe soluții concrete, care să poată fi înglobate în aceste tehnologii.

Utilitatea cercetării eticii digitale a IA constă în identificarea a ceea ce ar putea fi limitele de natură morală ale transformărilor digitale: sunt limite de natură tehnologică sau mai degrabă de natură etică. Necesitatea unui cadru care să valideze etica unei tehnologii IA urmărește stabilirea unui mediu favorabil unde tehnologiile emergente evoluează în conformitate cu normele etice și nevoile individuale de intimitate, securitate sau conformitate. Un astfel de cadru nu inhibă progresul tehnologic, ci poate facilita adoptarea tehnologiilor responsabile și sustenabile.

Comunitatea științifică are nevoie de cercetare pragmatică în domeniul eticii IA care să propună soluții punctuale, concrete și nu doar ghiduri și explicații despre amenințările IA. Riscul real al inteligenței artificiale nu este asociat amenințării, ci competențelor. Un sistem IA superinteligent va fi extrem de bun în a atinge obiective specifice de lucru,

dar dacă acele obiective nu sunt aliniate cu obiectivele oamenilor, atunci va exista o problemă (Stephen Hawking, 2015).

Capitolul 2. Perspective ale adoptării tehnologiilor IA în afaceri

Capitolul doi al lucrării propune abordarea perspectivelor adoptării tehnologiilor IA la nivelul companiilor. Se analizează astfel, caracteristicile companiilor care intenționează adopția de IA și utilizarea acestora și eventualele limitări care pot asociate aceste intenții. Totodată se explorează cele mai relevante modele de adoptarea tehnologică, cercetarea propunând modelul bazat pe valoare adăugată (*Value Based Adoption Model – VAM*) care poate fi utilizat de companii atunci când se dorește implementarea IA, în procesele de afaceri.

Într-o economie globală și interdependentă, companiile au implementat întotdeauna soluții tehnologice, strategice pentru eficientizarea afacerilor lor. În acest capitol ne propunem, într-o primă fază, să identificăm caracteristicile pragmatice (referitoare la profitabilitate, adăugare de valoare sau compatibilitate) pe care companiile trebuie să le aibă în vedere atunci când vor să adopte tehnologii IA. Totodată, detaliem limitările pe care firmele le pot întâmpina în implementarea și adoptarea tehnologiilor IA, pentru ca ulterior să propunem un model de evaluare tehnologică, care poate fi utilizat de mediul de afaceri în adoptarea tehnologiilor IA. Demersul științific din această parte a cercetării lansează o direcție de analiză conceptuală în vederea identificării unei legături între:

- perspectivele adoptării tehnologiilor IA de către companii și modelele care pot fi utilizate în acest sens și
- necesitatea unui framework de etică care poate fi utilizat în timpul adoptării IA.

Acest element comun este cel al *valorii adăugate totale*, valoare pe care tehnologiile IA o poate aduce în companii sub diferite forme, dar doar dacă există premisele unui cadru intențional prin care să se asigure

îndeplinirea anumitor condiții specifice, pe care le prezentăm în secțiunile următoare.

În acest capitol cercetarea subliniază necesitatea ca firmele să aibă o înțelegere holistică și multidisciplinară. Însă acest lucru solicită implicarea diferiților actori care să contribuie la dezvoltarea, implementarea și gestionarea IA, respectiv: dezvoltatori de produs, programatori, ingineri, data scientists, echipe de conformitate și *etică*, departamente de resurse umane, operațiuni, vânzări, marketing, precum și echipe juridice. Ideea de colaborare (între oameni și tehnologie) în vederea implementării IA, cu scopul de a adăuga valoare, este susținută și de unele studii (La Torre et al., 2018; Wilson & Daugherty, 2018; Levine, 2022). Acestea sugerează că soluțiile IA nu funcționează întotdeauna cel mai bine singure. Tehnologiile IA pot fi excelente în facilitarea rapidă și complexă a procesului decizional sau în automatizarea unor procese, dar companiile pot obține cele mai mari îmbunătățiri ale performanței atunci când oamenii și mașinile/aplicațiile lucrează *împreună*.

Plecând de la cercetările menționate, apare întrebarea: *De ce o parte dintre companiile care utilizează soluții IA reușesc să genereze valoarea adăugată și altele nu?* De altfel, pe baza acestor constatări a fost stabilit și primult obiectiv al cercetării, acela al identificării unor modele care pot fi utilizate de afaceri în adoptarea tehnologiilor IA în general și, a adoptării unor frameworkuri de etică pentru IA în mod particular. *IA poate aduce eficientizare companiilor doar dacă există o abordare multidisciplinară focusată pe valoare adusă utilizatorilor în urma adoptării tehnologiei IA* (cu referire la utilizatori finali/clienti sau utilizatori interni ai unei firme). În acest sens, ultima secțiune a acestui capitol propune un model de adoptare a tehnologiilor IA care să fie centrat pe valoarea produsă de acea tehnologie. Această abordare este, de fapt, similară cu aceea pe care o avem în crearea frameworkului de etică: un framework care să gestioneze aspectele de etică ale IA, pentru a conserva plusvaloarea pe care acea tehnologie o aduce în afacere.

Lucrarea de doctorat, evidențiază în această etapă a cercetării, următoarele caracteristici pe care companiile trebuie să le aibă în vederea adoptării tehnologiilor IA la nivelul proceselor de afaceri:

1. Calitatea seturilor de date

Legătura dintre seturile de date și tehnologiile etice IA este fundamentală, deoarece seturile de date utilizate pentru antrenamentul IA influențează fundamental comportamentul și procesul de luare a deciziilor ale acestor tehnologii. Totodată, datele în sine ar trebui să fie explorate din punct de vedere etic (Blackman, 2020), adică să fie analizate din prisma principiilor etice aplicabile IA (prezentate în capitoul trei), în vederea selectării lor pentru a dezvolta și implementa tehnologii IA considerate etice.

2. Definirea unor cazuri de utilizare pentru IA (Business Use Case)

După cum se arată în unul dintre rapoartele celor de la O'Reilly (2020), cultura organizațională a companiilor care nu recunosc nevoia IA, precum și dificultățile de identificare a cazurilor de utilizare în afaceri, se numără printre principalele bariere în calea implementării IA. Un alt aspect important în definirea cazurilor pentru optimizarea proceselor de afaceri prin IA este cel al sprijinului pe care managementul strategic al companiilor îl oferă. Obținerea de avantaje prin adoptarea IA implică nu numai implementarea acesteia ca soluție la nivel de companie, dar și dedicarea, implicarea managementului în susținerea și comunicarea detaliilor soluției.

3. Compatibilitatea IA cu mediul de afaceri

Dezvoltarea tehnologiilor IA nu necesită doar deținerea cunoștințelor tehnice potrivite pentru a construi algoritmi, dar și specialiști în domeniu care cunosc activitățile, fluxurile de lucru, raționamentul clienților, procesele afacerii și care au capacitatea de a analiza dacă tehnologiile IA pot aduce valoarea estimată (Tarafdar et al., 2019). Inovarea accelerată poate implica un nivel ridicat de incertitudine

organizațională, cu schimbări care să necesite abilități tehnice și resurse dedicate, ceea ce poate duce la creșterea perioadei de implementare a proiectului, dar și la o evoluție imprevizibilă (Hengstler et al., 2016).

În acest capitol s-au abordat perspectivele adoptării IA în mediul de afaceri. S-a constatat că principala motivație a companiilor pentru a implementa IA este dată, în prezent, de promisiunile acestor tehnologii referitoare la profitabilitate, avantaj competitiv, eficientizare sau maximizare profit (Venkatraman, 2017; Weill & Woerner, 2015). Cu toate acestea, am arătat că, deși aceste promisiuni pot fi îndeplinite de tehnologiile IA, este nevoie de o strategie care să vizeze concret obiectivele urmărite prin implementarea IA. Totodată, printr-o astfel de strategie companiile pot analiza specificul seturilor de date pentru IA raportat la obiectivele așteptate. Simultan, companiile trebuie să aibă în vedere posibilele limitări care pot apărea în procesul de adoptare al IA și soluții punctuale pregătite. De altfel, lucrarea noastră propune stabilirea unui model de adoptare tehnologică care să poată fi utilizat pentru IA. În aceste sens, considerăm modelul VAM, bazat pe valoare totală adăugată, un model pretabil pentru adoptarea IA în mediul de afaceri și pentru a măsura intenția utilizatorilor sau a consumatorilor acestei soluții.

Apreciem ca fiind esențiale următoarele elemente ce trebuie luate în calcul de companii atunci când se intenționează adoptarea tehnologiilor IA:

- *identificarea specificităților tehnologiei IA ce urmează a fi implementată și coroborarea acestora cu obiectivele de afaceri,*
- *analiza potențialelor limitări care ar putea împiedica adoptarea cu succes a tehnologiei,*
- *stabilirea modului în care se măsoară a priori succesul implementării IA, cu focus pe valoarea adusă companiei în sens larg,*
- *crearea unei strategii care să vizeze înglobarea aspectelor de natură etică în toate etapele implementării IA.*

În cele din urmă, tehnologia a demonstrat de-a lungul istoriei că are capacitatea de a genera inovație și lucruri benefice pentru societate. Depinde doar de oameni să găsească cadrul necesar care poate fi folosit

pentru a maximiza valoarea pe care tehnologia o poate aduce. Această abordare este valabilă și în cazul tehnologiilor IA și frameworkul propus în teză reliefează procese și mecanisme care au rolul de a cataliza adoptarea etică a IA în vederea maximizării valorii adăugate.

Capitolul 3. Contextualizarea frameworkului pentru etica tehnologiilor IA

În **capitolul trei** a cercetării analizăm principiile și frameworkurile de etică aplicabile IA, reliefând principalele limitări ale operaționalizării acestora. Totodată, prezentăm posibilele modele de etică care pot fi luate în considerare în dezvoltarea unui framework de etică pentru IA. Plecând de la aceste analize, urmărim definirea și modelarea conceptuală a frameworkului propriu, care, printre alte caracteristici, propune elemente de operaționalizare a principiilor etice și tranziția acestora de la o descriere teoretică către o implementare practică într-o tehnologie IA.

Întrucât literatura de specialitate abundă în definirea și recomandarea de principii etice aplicabile IA, a fost necesară o rafinare a lor pentru a extrage cele mai relevante principii etice, în contextul scopului cercetării.

În vederea identificării principiilor etice aplicabile IA, în cercetarea noastră am adaptat metodologia PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses, 2020*), pentru a realiza o căutare structurată (vezi subcapitolul 3.1).

S-au utilizat trei etape în vederea identificării principiilor etice aplicabile IA, după cum urmează:

a) *Identificarea inițială:*

- s-au efectuat căutări succesive în bazele de date (*Scopus, Web of Science*), folosind termenii de căutare în limba engleză „*AI ethics*”, „*AI ethical principles*”, „*AI ethical framework*” și „*AI ethical guidelines*”;

- compararea documentelor identificate cu bazele de date online *The Global AI Inventory* *Algorithmwatch* (<https://algorithmwatch.org/en/>) și *AI Ethicist (AI Principles | AI Ethicist)*;
- b) *Faza de screening*:
- rezultatele obținute în faza inițială au fost filtrate pentru a selecta pentru analiză doar acele principii etice relevante pentru cercetare. Motivele folosite pentru excludere au fost codificate în figura 3.1 după cum urmează:
 - M1* – principii sau articole duplicat;
 - M2* – articole redactate în altă limbă decât engleză;
 - M3* – articole nerelevante, care nu conțin nici principii etice și nici frameworkuri de etică aplicabile IA.
 - pentru a valida datele obținute în etapele de căutare am analizat și alte cercetări de analiză sistematică a literaturii de specialitate (Jobin et al., 2019; Ryan și Stahl, 2020), având ca scop realizarea de ghiduri complexe care sintetizează cele mai relevante principii etice.
- c) *Studiile incluse în analiză*: 150 de studii legate de frameworkuri de etică pentru IA au fost extrase și apoi analizate pentru a identifica temele și principiile etice comune din cadrul lor.
- Întrucât interesul principal al acestei etape de cercetare este de a determina principiile existente pentru dezvoltatorii și utilizatorii de IA, am filtrat o parte dintre cele care ofereau aceleași recomandări, dar aveau nume diferite. Ca urmare, constatările noastre, în această etapă, sunt mai degrabă descriptive, urmând a avea o abordare pragmatică, *utilitaristă* în etapa ulterioară, a realizării frameworkului, care să permită aplicarea practică a principiilor identificate. Potrivit acestei abordări, companiile nu sunt agenți economici simpli, ci sunt responsabile de gestionarea relațiilor cu publicul și de promovarea propriei legitimități instituționale.

În același sens, considerăm că etica IA nu poate fi înfăptuită doar la nivel declarativ, ci este nevoie de construcția unor procese prin care etica IA să fie înglobată la nivelul tehnologiei, dovedind astfel o operaționalizare utilitaristă, în sensul mai sus menționat.

Evoluția rapidă a tehnologiilor de inteligență artificială a generat o atenție sporită asupra potențialelor implicații etice asociate. În pofida existenței unui interes global crescut privind etica IA, cercetarea noastră indică un peisaj complex, divers, marcat de o serie de provocări. Deși s-au făcut progrese semnificative în definirea liniilor directoare etice pentru IA, există abordări variate notabile, dar care au un puternic caracter normativ, centrat exclusiv pe ceea ce ar trebui să respecte IA din punct de vedere etic.

O provocare semnificativă este dată de implementarea practică a acestor principii. Tehnologiile IA, marcate de complexitatea lor tehnică și de natura adesea opacă, fac dificilă asigurarea aderării la principiile etice convenite, cum ar fi echitatea, transparența, responsabilitatea și confidențialitatea.

O altă problemă este interpretarea subiectivă a principiilor etice. Definiția conceptelor precum „echitatea” poate varia semnificativ în diferite contexte și culturi, adăugând un strat suplimentar de complexitate în alinierea rezultatelor sistemelor IA cu principiile etice.

Lipsa unor norme specifice și verificabile referitoare la modul de aplicabilitate a principiilor etice IA poate afecta eficacitatea unui framework de etică pentru IA. Totodată, implementarea în practică a principiilor etice trebuie să țină cont și de industria în care sunt folosite respectivele tehnologii IA riscurile asociate, tipul de utilizatori finali, aspecte socioculturale. Spre exemplu, tehnologia IA aplicată în domeniul medical ar putea avea mai multe riscuri decât aceeași tehnologie aplicată în industria de retail. De aceea, orice companie care intenționează să utilizeze o tehnologie IA într-un anumit context va trebui să efectueze o verificare inițială pentru a analiza dacă există aspecte etice care trebuie satisfăcute de respectiva tehnologie.

Diferitele valori care pot reprezenta fundamentul a ceea ce numim etica IA fac întotdeauna obiectul unor interpretări diferite, în funcție de contextul în care este implementată o tehnologie IA. Apreciem că tehnologiile IA, în special, în contextul mediului de afaceri, sunt sisteme sociotehnice definite prin interoperabilitatea dintre oameni și tehnologie. Prin urmare, deoarece diverse părți ar putea interacționa cu o tehnologie IA pe durata ciclului său de viață (dezvoltare, implementare sau utilizare), este important să existe o abordare unitară atunci când vine vorba de operaționalizarea eticii IA. Pentru această etapă a cercetării am luat în considerare studiul VDE AI Ethics Group (Hallensleben, 2020), care grupează provocările actuale ale implementării IA în trei categorii principale:

- *Dependența de context* – definirea valorilor/principiilor etice (precizate în secțiunea anterioară), care ar putea valida o tehnologie IA ca fiind etică, depinde de domeniul de aplicare și de contextul cultural.

- *Natura sociotehnică a IA* – tehnologiile IA sunt sisteme sociotehnice. Impactul lor depinde nu numai de tehnologie (date și algoritmi), ci și de obiectivele care stau la baza sistemului și de modul în care IA este încorporată într-o companie.

- *Ușurință în utilizare* – datorită naturii sociotehnice a sistemelor IA, frameworkurile de implementare practică a eticii IA trebuie să ofere instrumente care să ia în considerare diferitele roluri jucate de dezvoltatorii și utilizatorii de sisteme în furnizarea măsurilor necesare.

Pe lângă provocările menționate, adăugăm una suplimentară, care este mai degrabă legată de *governanța generală* a unui framework din punct de vedere funcțional. Este vorba de faptul că unele dintre tehnologiile IA sunt capabile să se adapteze automat sau să învețe din procesele anterioare.

Prin urmare, este important ca un framework de etică aplicabil IA să poată fi continuu actualizat prin modalități de colectare de feedback care să fie disponibile între dezvoltatori, consumatori și alți actori care intervin în ciclul de viață al IA. În plus, o companie care implementează IA trebuie să fie capabilă să măsoare rezultatele generate de IA cu referire

la potențiale consecințe non-etice. Acest lucru poate fi făcut prin utilizarea unui framework de etică aplicabil IA, întrebuițarea acestuia fiind o practică a unui proces continuu, iar în cazul în care apar discrepanțe de natură etică, frameworkul să permită îmbunătățirea modelului inițial propus. Modelul VAM, descris în capitolul trei, poate fi folosit în acest sens.

Apreciem că nu orice tehnologie IA va prezenta aceleași riscuri în funcție de modul în care este utilizat și, prin urmare, un framework de etică pentru IA ar trebui să poată face deosebirea între diferitele categorii de riscuri, sugerând în același timp diverse acțiuni în consecință. În acest sens, tehnologiile IA trebuie clasificate în diferite clase de risc și orice cadru etic al IA ar trebui să fie capabil să le recunoască. În consecință, cercetarea noastră va explora modalități de a oferi îndrumări referitoare la riscuri specifice IA, pentru a ajuta la dezvoltarea unor clasificări de risc interne specifice fiecărei companii.

Cercetarea sintetizează în capitolul trei că procesul de adoptare a unor tehnologii IA etice la nivelul afacerilor, poate fi marcat de o serie de provocări, care pleacă de la înțelegerea conceptuală și până la implementarea practică. În pofida acestor obstacole, este un domeniu care necesită cercetare continuă și dialog interdisciplinar pentru a asigura evoluția responsabilă a tehnologiilor IA. Cercetarea noastră, prin scopul propus, răspunde acestor provocări, în sensul în care propunem mecanisme concrete prin care se pot defini principiile etice funcție de contexte diferite, se pot implementa la nivel algoritmic și se pot identifica, monitoriza și ajusta eventuale abateri etice. Toate acestea, ca parte dintr-un proces continuu care să valideze etica asociată IA.

Capitolul 4. Framework de etică aplicabil tehnologiilor inteligenței artificiale implementate în mediul de afaceri

Capitolul patru este dedicat descrierii detaliate a frameworkului de etică aplicabil tehnologiilor inteligenței artificiale implementate în mediul de afaceri (FEIA) pe care îl propunem.

Prin întrebarea de bază a cercetării, ***Cum putem valida etica aplicată tehnologiilor inteligenței artificiale implementate în mediul de afaceri?***, am căutat răspunsuri privind modul în care se poate asigura un cadru de lucru prin care tehnologiile IA implementate în mediul de afaceri produc rezultate etice într-un context dat. Scopul a fost crearea unui framework operațional care poate fi utilizat în validarea principiilor etice la nivelul inteligenței artificiale pe toată durata ciclului de viață a acestor tehnologii, de la etapa de dezvoltare până la post-implementare. Așa cum am expus în capitolul anterior, literatura științifică surprinde un număr mare de frameworkuri de validare etică a IA, majoritatea fiind limitate din punctul de vedere al operaționalizării practice a principiilor etice în tehnologiile IA.

Metodologia pe care am adoptat-o pentru a realiza frameworkul de etică se regăsește detaliată în subcapitolul 4.2. Aceasta a presupus atât analiza de conținut, cât și observația, respectiv modelarea conceptuală. Totodată, menționăm că în crearea frameworkului de etică propus a contat și experiența personală acumulată, în cei 12 ani de activitate a doctorandului, în calitate de consultant în informatică pentru diferite companii multinaționale.

Ca urmare a analizei literaturii științifice, efectuată în capitolul trei, care descrie și explorează principiile etice și frameworkurile de etică aplicabile IA (folosind metoda PRISMA) am identificat frameworkul de etică VCIO, care descrie practic un mod funcțional de implementare a principiilor etice în tehnologiile IA, în funcție de *Valori, Criterii, Identificatori și Observabile*. Ulterior, pe baza altor lucrări conexe despre etică, am concluzionat că, deși frameworkul de etică VCIO face un pas înainte din perspectiva operaționalizării eticii IA, îi lipsesc elemente de natură tehnică. Astfel, am considerat modelul VCIO ca punct relevant de plecare în realizarea propriului framework de etică, care să abordeze atât aspectele normative ale principiilor etice, cât și elementele tehnice specifice implementării la nivelul tehnologiilor IA.

Prima versiune de framework de etică (FEIA v1, descrisă în secțiunea 4.2.2 a tezei) pe care am propus-o a obținut o validare științifică

în cadrul conferinței CEUR Proceedings I-ESA 2022 (Ciobanu & Meșniță, 2022). Feedbackul obținut în cadrul conferinței s-a referit la elementele funcționale și posibilitățile prin care FEIA v1 poate fi implementat în practică.

Plecând de la acest feedback, într-o etapă ulterioară s-a decis demararea unor discuții cu experți care să poată oferi opinii privind percepția asupra unui framework de etică pentru IA implementat practic în mediul de afaceri. Astfel, s-au ales câte doi experți din domenii care au legătură cu dezvoltarea și implementarea tehnologiilor IA: *dezvoltatori de inteligență artificială, reprezentanți din conducerea unor afaceri care utilizează inteligență artificială, teoreticieni ai eticii aplicate*. Discuțiile avute cu aceștia au avut ca subiect următoarele teme:

- definirea principiilor etice specifice IA;
- nevoia de reglementare etică la nivelul IA;
- modalități de operaționalizare a unui framework de etică pentru IA;
- nevoia mediului de afaceri de a gestiona tehnologia IA;
- riscuri și beneficii ale IA în mediul de afaceri;
- instrumente care să asigure etica la nivelul modelelor de antrenament IA.

În urma acestor discuții, am obținut versiunea FEIA v2, descrisă în secțiunea 4.2.3 a tezei, pe care o considerăm ca o soluție de gestionare a eticii la nivelul tehnologiilor IA implementate în mediul de afaceri.

Ca urmare a abordării metodologice detaliate, versiunea actualizată a frameworkului (FEVIA v2) aduce completări care vizează elementele tehnice de control al modelelor de antrenament, precum și elemente ce țin de modelul decizional referitor la aspectele de etică relevante pentru afacere, ca urmare a implementării IA (vezi figura 3).

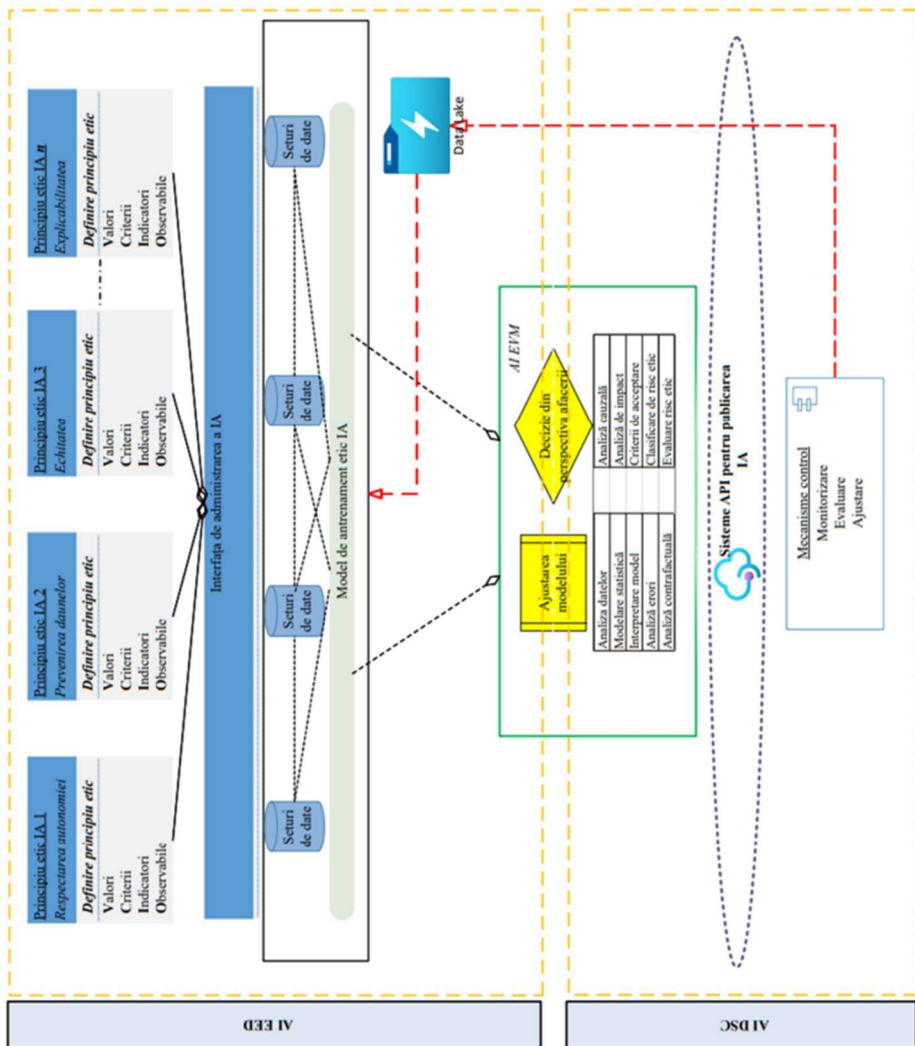


Figura 3. Versiunea a doua a frameworkul de etică aplicabil tehnologiilor IA (FEIA v2)

Frameworkul cuprinde două componente majore, *AI Embedded Ethics by Design (AI EED)* și *AI Desired State Configuration (AI DSC)*, care pot fi practic inter-conectate prin sisteme API (*Application*

Programming Interface), prin care sugerăm o abordare holistică a operaționalizării unei tehnologii IA, care tratează provocările atât din partea producătorilor de IA, cât și a consumatorilor. Acestea sunt descrise în detaliu în capitolul patru al tezei.

Fundamentul dezvoltării acestui framework a fost inspirat din conceptul de *Enterprise Architecture* aplicat industriei IT.

Astfel, în cercetarea de față înțelegem prin *framework* acel set de linii directoare și instrumente care oferă o structură pentru a sprijini dezvoltarea unui anumit sistem informatic bazat pe tehnologii IA. Pe cale de consecință, înțelegem prin *framework de etică aplicabil IA* acel set de linii directoare care au un rol în a preveni, diminua și rezolva eventuale consecințe etice care pot rezulta din dezvoltarea, implementarea și monitorizarea tehnologiilor IA. Astfel, frameworkul de etică aplicabil IA pe care îl propunem definește organizarea, arhitectura și procesele de guvernare pe care companiile le pot avea în vedere atunci când decid utilizarea tehnologiilor IA în scop de optimizare a afacerii. Acest framework aduce o abordare standardizată care oferă o bază pentru utilizarea tehnologiilor IA în mediile de afaceri digitalizate, controlând de o manieră proactivă și reactivă eventualele consecințe non-etice care pot rezulta în urma implementării IA.

Beneficiarii direcți ai acestui framework vor fi:

- companiile care doresc să utilizeze IA, dar controlând, proactiv și reactiv, potențialele efecte non-etice care pot rezulta post-implementare și care pot afecta negativ imaginea;
- experți IA care doresc să dezvolte și să întrețină sisteme informatice bazate pe IA care să aibă decizii considerate etice funcție de criteriile prestabilite (economice, socioculturale etc.);
- departamente specifice din companii care au ca obiect de activitate controlarea aspectelor ce țin de legalitate, securitate și conformitate, pentru ca organizația de afaceri să își desfășoare activitatea conform normelor juridice în vigoare;

- industria IA care, prin punerea la dispoziție a unor tehnologii IA care pot fi gestionate etic printr-un astfel de framework, vor avea o reputație sporită;
- clienții finali, care folosesc serviciile și produsele create de companii prin utilizarea tehnologiilor IA.

În general, în procesul de dezvoltare-implementare a unei soluții IA există șase etape esențiale care au rol de a asigura că valoarea adusă în procesul de afaceri este cea dorită (Rao et al., 2021). Aceste etape răspund fiecare unor întrebări esențiale, și anume:

- Există oportunitatea și nevoia de a implementa o soluție IA sau nu?
- Cum proiectăm (construim sau cumpărăm sau închiriem) soluția AI?
- Modelul de antrenament de test corespunde așteptărilor, priorităților afacerii?
- Dacă proiectul-pilot este acceptat, va fi modelul implementat în producție?
- Este modelul gata să fie transferat pentru funcționare („business-as-usual“)?
- Ar trebui modelul să continue așa cum este sau trebuie recalificat, reproiectat sau retras?

Practic, fiecare întrebare reprezintă o etapă în ciclul de viață al unei tehnologii IA implementate cu scop de a crea valoare prin optimizarea unor procese de afaceri. Frameworkul propus nu contrazice abordarea generală prezentată anterior, însă adresează specific problematica eticii care trebuie satisfăcută de orice tehnologie IA. Acest lucru se datorează faptului că frameworkul pleacă de la ideea deja menționată, conform căreia etica IA nu este doar un *task* care trebuie bifat într-un proces, ci un proces continuu în sine.

Crearea unui framework de etică pentru IA cu aplicabilitate practică reprezintă un proces complex, multidisciplinar și presupune o explorare holistică. Un astfel de demers trebuie să ia în considerare atât aspecte ce țin de identificarea unor norme etice, cât și de caracteristicile unei afaceri și de particularitățile tehnice ale tehnologiilor IA.

Un element esențial pentru dezvoltarea FEIA a fost dat de faptul c, în urma analizei literaturii, s-a constatat un puternic caracter normativ care scoate în evidență mai degrabă *ce trebuie* să respecte IA din punct de vedere etic, și mai puțin *cum trebuie* să respecte.

Astfel, gestionarea aspectelor de etică într-un mod *continuu și comprehensiv* în diferitele faze de adoptare a IA, de la dezvoltare, implementare, până la post-implementare, a fost ideea de bază a creării FEIA.

Frameworkul propus în această cercetare este conceptual și nu pretinde că are răspunsuri *a priori* în garantarea eticii IA. Abordarea eticii IA nu ar trebui să fie referitoare la *cum putem crea a priori tehnologii IA etice și responsabile*. Datorită caracteristicilor tehnice ale acestor tehnologii, considerăm a fi mai potrivită abordarea care pune accent pe *cum se poate crea un proces care să gestioneze etic IA în diferitele etape ale adoptării acesteia*.

Considerăm însă că FEIA, prin cele două componente interoperabile și prin elementele funcționale tehnice prezentate, oferă pentru companii posibilitatea verificării continue a eticii IA. Totodată, FEIA propune mecanisme concrete care să detecteze eventualele abateri etice, ca mai apoi să se poată acționa în vederea corijării acestor abateri direct în modelele de antrenament, inițial diferite.

Faptul că un model de antrenament este *antrenat etic* în faza pilot a adoptării IA nu garantează că datele pe care IA le va primi post-implementare nu vor influența rezultatele aceluia IA, în sens de abatere etică. Așadar, este important ca un framework de etică să asigure că astfel de situații sunt identificate, analizate și reintegrate în modelele inițiale. Crearea și dezvoltarea FEIA s-a bazat pe această abordare, pentru că astfel se poate asigura și *trasabilitatea* eticii la nivelul IA. Această caracteristică a trasabilității la nivelul FEIA permite identificarea momentului care a cauzat abaterea etică și, astfel, separarea datelor care au influențat acel moment, într-o bază de date nouă, în vederea analizării și reantrenării. Acest lucru poate ajuta atât practicienii IA care au

dezvoltat tehnologia, cât și echipele de audit informatic, interne sau externe.

Nu în ultimul rând, cercetarea reține că FEIA, deși este prezentat ca un concept, face un pas înainte din punct de vedere științific, propunând metode concrete prin care normele etice pot fi operaționalizate la nivelul algoritmilor IA, ca parte dintr-un proces continuu care să valideze etica IA.

Concluzii finale

Inteligența artificială a evoluat de la a fi un termen la modă la a deveni o realitate omniprezentă a vieții moderne, cu un potențial uriaș de creștere în sfera afacerilor. Companiile folosesc IA pentru a dezvolta sau maximiza eficiența proceselor de afaceri, dar complexitatea tehnologiilor IA creează adesea un sentiment de „cutie neagră”, chiar și pentru cei care le dezvoltă.

Eforturile pentru dezvoltarea și implementarea etică a IA nu sunt noi, iar cercetarea noastră a evidențiat principiile etice și frameworkurile de etică care pot orienta această dezvoltare. Cu toate acestea, există încă discrepanțe în modul în care principiile etice pot fi puse în practică în diferite contexte de aplicabilitate a inteligenței artificiale în mediul de afaceri. Astfel, gestionarea aspectelor de etică aplicabile tehnologiilor IA poate fi influențată de aspecte ce țin de modelul de afaceri în care se adoptă IA, zona geo-culturală, sau specificul utilizatorilor și al consumatorilor acelei afaceri. Pentru adoptarea IA în mediul de afaceri în mod responsabil, nevoile etice ale IA trebuie analizate din perspectiva fiecărui aspect mai sus menționat, ca parte dintr-un proces continuu.

Cercetarea a avut ca obiective, identificarea etapelor în care se justifică aplicarea unui cadru etic în dezvoltarea tehnologiilor IA, identificarea modelelor care pot fi utilizate de companii în vederea adoptării unui framework de etică pentru IA și nu în cele din urmă, identificarea unui mecanism (ca o componentă funcțională a unui framework) de monitorizare a abaterilor etice. Prin aceste obiective, teza

a vizat dezvoltarea unei abordări integrative care să faciliteze analiza și conexiunea dintre dimensiunile conceptuale asociate eticii și ale tehnologiei, în speță ale inteligenței artificiale. Astfel, s-a urmărit interacțiunea dintre principiile etice și aplicabilitatea lor în contextul inteligenței artificiale, având în vedere și implicațiile lor în mediul de afaceri. Acest cadru analitic a permis, prin instrumentele metodologice propuse, analiza literaturii de specialitate, observația, modelarea conceptuală, contextualizarea și delimitarea temei de cercetare, etica tehnologiilor de inteligență artificială în mediul de afaceri și a fundamentat construcția frameworkului de etică pentru IA pe care l-am propus (FEIA).

Teza oferă, pe de o parte, o imagine de ansamblu a ceea ce înseamnă astăzi etica IA, dar și a evoluției eticii în IA raportat la mediul de afaceri. Această abordare a evoluției eticii în raport cu tehnologia, în general, și cu IA, în mod particular, este esențială pentru a înțelege care au fost și sunt metodele prin care putem raporta etica la tehnologie și invers. Tehnologiile, în general, prin caracteristicile inovatoare pe care le-au comportat la momentul apariției lor, au generat, pe lângă oportunități, și provocări la nivelul societății. Cu toate acestea, omenirea s-a adaptat și a găsit mereu un echilibru în a maximiza valoarea adusă de tehnologie și a minimiza eventualele implicații negative.

Pe de altă parte, cercetarea consideră relevantă plasarea temei eticii IA în zona mediului de afaceri. Companiile sunt cele care, prin funcțiile de afaceri specifice, sunt primele care folosesc tehnologiile și le dau sensul valoric. Astfel, este important ca în primul rând companiile să aibă structuri și procese prin care să poată gestiona modul în care o tehnologie este adoptată, validată și gestionată din punct de vedere etic.

Dezvoltarea și implementarea etică a inteligenței artificiale nu reprezintă un concept nou și, așa cum am demonstrat în cadrul tezei, există deja principii de etică bine definite, precum și frameworkuri de etică care să ghideze dezvoltarea etică a IA. Cu toate acestea, persistă o anumită discrepanță între operaționalizarea practică a acestor principii etice și contextul specific în care sunt aplicate.

Pe baza analizei sistematice a literaturii de specialitate, se evidențiază necesitatea integrării considerațiilor etice în cadrul inteligenței artificiale. Este notabil faptul că studiile și publicațiile privind etica în IA tind să se manifeste la un nivel conceptual, caracterizat prin abstracție și generalizare. Predomină o tendință care subliniază imperativul nevoii de etică, în timp ce soluțiile propuse par să se încline mai mult spre teoretic, decât spre abordări pragmatice și aplicabile în situații concrete.

Simultan, se poate constata că, în contextul adoptării inteligenței artificiale de către mediul de afaceri, majoritatea cercetărilor indică profitabilitatea și avantajul competitiv ca fiind principalele motive. Cu toate acestea, succesul adoptării IA în companii depinde de o strategie bine definită, care să coreleze tehnologia cu obiectivele de afaceri și anticiparea potențialelor limitări. Cercetarea sugerează astfel utilizarea unui model de adoptare a tehnologiei (VAM), care să urmărească valoarea totală adăugată. Elemente esențiale pentru adoptarea cu succes a IA includ alinierea la obiectivele strategice, identificarea constrângerilor, evaluarea valorii adăugate și integrarea considerentelor etice. Această abordare este valabilă în cazul adoptării IA, iar frameworkul propus reliefează procese și mecanisme care au rolul de a facilita adoptarea etică a IA, în vederea maximizării valorii adăugate.

Frameworkul propus a fost dezvoltat ca răspuns la întrebarea fundamentală a cercetării și anume, ***Cum putem valida etica aplicată tehnologiilor inteligenței artificiale implementate în mediul de afaceri?*** Frameworkul aduce câteva avantaje esențiale, în ceea ce privește validarea eticii la nivelul tehnologiilor inteligenței artificiale. Pe de o parte, oferă flexibilitate în operaționalizarea principiilor etice într-un mod dinamic, în funcție de cerințele specifice ale fiecărei afaceri. Pe de altă parte, prin cele două componente, AI Embedded Ethics by Design și AI Desired State configuration, asigură controlul implementării principiilor etice în modelele de antrenament IA atât în faza de pre-testare (faza pilot), cât și post-implementare. Totodată, FEIA reliefează metodele prin care principiile etice pot fi operaționalizate la nivelul IA, precum și

modalitățile de monitorizare și detecție a unor posibile abateri etice. De asemenea, FEIA introduce elementul de *trasabilitate*. Este important pentru companii să existe mecanisme prin care să se poată verifica etapa sau chiar funcția algoritmică care a generat o abatere etică.

Necesitatea unui framework care să gestioneze etica IA urmărește stabilirea unui mediu favorabil, unde tehnologiile emergente evoluează în conformitate cu normele etice și nevoile individuale de intimitate, securitate, conformitate. Un astfel de framework nu inhibă progresul tehnologic, ci poate facilita adoptarea tehnologiilor într-un mod responsabil și sustenabil.

Astfel, considerăm că prin atingerea obiectivelor și prin abordarea metodologică urmată, cercetarea răspunde întrebării fundamentale prin dezvoltarea conceptuală a FEIA, care poate gestiona aspectele de etică într-un mod *continuu și comprehensiv* în diferitele faze de adoptare a IA: dezvoltare, implementare, post-implementare.

În concluzie, este necesară o abordare holistică, dar flexibilă, a eticii în IA, care să țină cont de multiplele variabile care pot influența implementarea etică a acestor tehnologii. La momentul realizării acestei cercetări, tehnologiile IA au încă nevoie de inputul uman. Cu toate acestea, este nevoie ca oamenii să definească cadre de lucru prin care puterea inovatoare a IA să fie gestionată și direcționată doar către valoarea percepută.

Comunitatea științifică are nevoie de cercetare pragmatică în domeniul eticii IA care să propună soluții punctuale, concrete și nu doar ghiduri și explicații despre amenințările IA. Riscul real al inteligenței artificiale nu este asociat amenințării, ci competențelor. Un sistem IA superinteligent va fi extrem de bun în a atinge obiective specifice de lucru, dar dacă acele obiective nu sunt aliniate cu obiectivele oamenilor, atunci va exista o problemă (Stephen Hawking, 2015).

Limite și direcții de cercetare

Tehnologiile de inteligență artificială sunt în continuă dezvoltare, iar ultimele inovații în materie, IA generative, precum ChatGPT sau

DALL-E, au schimbat complet paradigma utilizării IA în afaceri. Mai mult, companiile pare că au intrat într-o competiție referitoare la când și cum pot folosi IA în diferite scopuri legate de procesele de afaceri.

Totodată, din punct de vedere tehnic, dezvoltarea unei soluții bazate pe tehnologie IA necesită, în primul rând, o cantitate enormă de date, precum și selecția unor tehnici de antrenament algoritmic aliniate cu obiectivele soluției.

Cercetarea noastră a urmărit realizarea unui framework de etică pentru IA, care să propună metode concrete de operaționalizare a normelor etice la nivel algoritmic. Una dintre limitările acestei cercetări a fost aceea că, dat fiind faptul că frameworkul este unul conceptual, nu a existat posibilitatea experimentării acestuia la nivel practic, pe o tehnologie IA deja implementată sau în curs de implementare. Acest lucru ar fi cerut, în primul rând, acces la o cantitate mare de date pentru a crea un mediu de test care să simuleze diferite scenarii de aplicabilitate IA.

Astfel, din această limitare se desprinde și o posibilă direcție viitoare de cercetare. Și anume: utilizarea FEIA într-un scenariu real, pentru a valida frameworkul din punct de vedere practic. Totodată, alte direcții de cercetare ar putea fi legate de testarea practică, de tip *Proof of Concept*, a elementelor funcționale ale FEIA. De exemplu, s-ar putea măsura gradul de identificare a abaterilor etice prin mecanismul de monitorizare propus și clasificarea rezultatelor fals pozitive sau fals negative.

O altă limitare a acestei cercetări a reprezentat-o faptul că, în România, studiul privind etica IA este încă în faze incipiente, astfel încât materialele bibliografice analizate au fost doar în limba engleză, iar traducerea unor noțiuni tehnice a fost, pe alocuri, dificilă, pentru că s-a dorit păstrarea sensului specific anumitor concepte.

Totodată, considerăm că etica tehnologiilor IA, în sens larg, ar trebui aprofundată în mediul academic din România, iar ca direcții viitoare propunem introducerea de cursuri universitare specifice și centre de cercetare aplicate, în parteneriat cu industria dezvoltatoare de inteligență artificială. Având în vedere capacitatea de forță de muncă din

zona tehnologiei informației din România, un astfel de parteneriat ar putea pune bazele unor echipe multi-disciplinare care să elaboreze proiecte cu potențial inovator în acest domeniu de cercetare.

O altă direcție de cercetare pe care o considerăm oportună, este aceea a aprofundării studiului cu privire la necesitatea creării unui organism (la nivel național sau internațional) cu rol de verificare sau auditare a aspectelor de etică aplicabile IA. În acest sens, s-ar putea avansa cercetări care să măsoare viabilitatea unor astfel de organisme, fie prin comparație cu firme private de audit (ca PwC, Deloitte, KPMG și altele), fie prin comparație cu organisme guvernamentale de control al activităților informatice. Astfel de cercetări ar putea ajuta toate părțile implicate în dezvoltarea și utilizarea IA (companii care au în uz sau care produc IA, structuri guvernamentale, societate civilă, și nu numai) să aibă o înțelegere mai bună a riscurilor și avantajelor pe care tehnologiile IA le comportă.

Diseminarea activităților de cercetare

Cercetarea doctorală s-a desfășurat pe durata a trei ani, care au presupus atât prezentarea rapoartelor de cercetare intermediare în fața comisiei de îndrumare, cât și inițiative științifice, precum susținerea de seminare în cadrul Școlii Doctorale de Economie și Administrarea Afacerilor a Universității Alexandru Ioan Cuza din Iași, precum și publicarea de articole sau participarea la conferințe relevante pentru tema de cercetare. În rândurile următoare, reliefăm activitățile de cercetare desfășurate în perioada 2020-2023.

Publicații:

1. Ciobanu, C. A. & Meșniță, G., 2021. AI ethics in business – a bibliometric approach. *Review of economic and business studies*, Volume 14,(2), pp. 169-202. ISSN-1843-763X. Disponibil la http://rebs.feaa.uaic.ro/article-_ai_ethics_in_business_—a_bibliometric_approach-320.html .

2. Ciobanu, C. A. & Meșniță, G., 2022. AI Ethics for Industry 5.0 – From Principles to Practice. *Proceedings of the Workshop of I-ESA '22*, March 23–24, 2022, Valencia, Spain. Disponibil la <https://ceur-ws.org/Vol-3214/WS5Paper2.pdf>.
3. Ciobanu, C.A. & Meșniță G., 2023. Operationalizing Artificial Intelligence ethical principles in business – A conceptual framework în *Conference Proceedings of European Finance, Business and Regulation 2023*, Editura Universității Alexandru Ioan Cuza, Iași, România.

Participări la conferințe:

1. *Microsoft Research Summit 2021*, eveniment virtual cu tema "What's Next for Technology and Humanity?", 19-20 octombrie 2021.
2. *Microsoft Research Summit 2022*, eveniment virtual cu tema "What's Next for Technology and Humanity?", 18-20 octombrie 2022.
3. *Globalization and Higher Education in Economics and Business Administration (GEBA 2021)*, Iași, România, 21-23 octombrie 2021.
4. *I-ESA Workshops 2022, Interoperability for Enterprise Systems and Applications Workshops 2022*, Valencia, Spania, 23-25 martie, 2022.
5. *EU-FIRE, European Finance, Regulation and Business – May 2023*, Iași, România, 19-20 mai 2023.

Referințe bibliografice

- Accenture, 2019. *“Responsible AI and robotics: an ethical framework”*. [Interactiv]
Available at: <https://www.accenture.com/gb-en/company-responsible-ai-robotics>
- ACM, 2017. *Statement on algorithmic transparency and accountability*. [Interactiv]
Available at: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
- Afiouni, R., 2019. *Organizational Learning in the Rise of Machine Learning*. Munchen, McGill University.
- Agrawal, A., Gans, J. & Goldfarb, A., 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*. s.l.:Harvard Business Press.
- Ahmed, M., Mahmood, N. & Hu, J., 2016. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, Volumul 60, pp. 19-31.
- AI for Humanity, 2018. *For a meaningful artificial intelligence: towards a French and European strategy*.. [Interactiv]
Available at: <https://www.aiforhumanity.fr/en/>
- AI Now Institute, 2018. *“AI Now report 2018”*. [Interactiv]
Available at: https://ainowinstitute.org/AI_Now_2018_Report.pdf
- Ajzen, I., 1985. From Intentions to Actions: A Theory of Planned Behavior. În: *Action Control*. Berlin: Heidelberg, pp. 11-39.
- Ajzen, I., 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50(2), pp. 179-211.
- Ajzen, I. & Fishbein, M., 1973. Attitudinal and normative variables as predictors of specific behavior. *Journal of Personality and Social Psychology*, 27(1), p. 41–57.
- Alcantara, I., 2023. *The Ethics of Artificial Intelligence in the Workplace: How to Balance Innovation with Responsibility*. [Interactiv]
Available at: <https://www.linkedin.com/pulse/ethics-artificial-intelligence-workplace-how-balance-igor-alcantara/>
[Accesat 16 Mai 2023].

Algo.Rules, 2019. "Rules for the design of algorithmic systems". [Interactiv]
Available at: https://algorules.org/fileadmin/files/alg/Algo.Rules_EN_2.pdf

Alshammari, M. & Simpson, A., 2017. "Towards a Principled Approach for Engineering Privacy by Design.". *Annual Privacy Forum*.

Alsheibani, S., Cheung, Y. & Messom, C., 2018. *Artificial Intelligence Adoption: AI-readiness at Firm-Level.* [Interactiv]
Available at: <https://aisel.aisnet.org/pacis2018/37/>

Alsheibani, S., Cheung, Y., Messom, C. & Alhosni, M., 2020. Winning AI strategy: six-steps to create value from artificial intelligence. *Americas Conference on Information Systems*.

AMA, 2018. *Policy recommendations on augmented intelligence in health care H-480.940.* [Interactiv]
Available at: <https://www.ama-assn.org/press-center/press-releases/ama-passes-first-policy-recommendations-augmented-intelligence>

Amnesty International/Access Now, 2018. "The Toronto declaration: protecting the rights to equality and non-discrimination in machine learning systems". [Interactiv]
Available at: www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf

Amodei, D. și alții, 2016. Concrete Problems in AI Safety. *ArXiv abs/1606.06565*

Anderson, S. & Anderson, M., 2021. AI and ethics.. *AI Ethics* , Volumul 1, p. pp. 27–31.

Angerschmid, A. și alții, 2022. Fairness and Explanation in AI-Informed Decision Making.. *Machine Learning and Knowledge Extraction*. 2022; 4(2):, 4(2), pp. 556-579.

Angwin, J., Larson, J., Matt, S. & Kirchner, L., 2016. *Machine Bias*. [Interactiv]
Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
[Accesat 20 04 2023].

Arntz, M., Gregory, T. & Zierahn, U. (. ", 2016. The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. *OECD Social, Employment and Migration Working Papers*, Volumul 189.

Arrieta, A. B. și alții, 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, oportunități și provocări față de AI responsabilă. *Information Fusion*(58), pp. 82-115.

Asimov, I., 1984. *Foundation; I, Robot*. London: Octopus Books (First published 1950).

Association for Data-driven Marketing and Advertising (ADMA), 2013. *Best practice guideline: Big Data*. [Interactiv]
Disponibil la: <https://silo.tips/download/best-practice-guideline-big-data>

Astapciks, I., 2023. *Why Do Companies Need Digital Transformation?*. [Interactiv]
Disponibil la: <https://www.forbes.com/sites/forbestechcouncil/2023/03/20/why-do-companies-need-digital-transformation/?sh=5e937332bab3>
[Accesat 2 04 2023].

Åström, J., Reim, W. & Parida, V., 2022. Value creation and value capture for AI business model innovation: a three-phase process framework. *Review of Managerial Science*.

Ayling, J. & Chapman, A., 2022. Putting AI ethics to work: are the tools fit for purpose?.. *AI Ethics*, Volumul 2, p. 405–429.

Baier, L., Jöhren, F. & Seebacher, S., 2019. *CHALLENGES IN THE DEPLOYMENT AND OPERATION OF MACHINE LEARNING IN PRACTICE*. Stockholm, s.n.

Barn, B., 2019. Mapping the public debate on ethical concerns: algorithms in mainstream media. *Journal of Information, Communication and Ethics in Society*..

Barocas, S. & Selbst, A., 2016. Big Data's Disparate Impact. *California Law Review*, 104(3), pp. 671-732.

Basoglu, N., Ok, A. & Daim, T., 2017. What will it take to adopt smart glasses: A consumer choice based review?. *Technology and Society* (50), pp. 50-56.

Beauchamp, T. & Childress, J., 1979. Principles of Biomedical Ethics. În: s.l.:Oxford University Press, Oxford..

Bennett, M., 2023. *AI regulation: What businesses need to know in 2023*. [Interactiv]
Disponibil la: <https://www.techtarget.com/searchenterpriseai/feature/AI->

regulation-What-businesses-need-to-know

[Accesat 10 Iunie 2023].

Bessen, J., 2019. Automation and jobs: when technology boosts employment. *Economic Policy, CEPR/CES/MSH*, 34(100), pp. 589-626.

Bigelow, S. J., 2023. *What are the types of APIs and their differences?*. [Interactiv]

Available at: <https://www.techtarget.com/searchapparchitecture/tip/What-are-the-types-of-APIs-and-their-differences>

Blackman, R., 2020. *A Practical Guide to Building Ethical AI*. [Interactiv]

Available at: <https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai>
[Accesat 28 Octombrie 2022].

Blakemore, E., 2023. *New AI may pass the famed Turing Test. This is the man who created it..* [Interactiv]

Available at: <https://www.nationalgeographic.com/science/article/alan-turing-test-artificial-intelligence-life-history>
[Accesat 15 Mai 2023].

Bohai, L., 2021. *Ethical Concerns of Combating Crimes with AI Surveillance and Facial Recognition Technology*. [Interactiv]

Available at: <https://towardsdatascience.com/ethical-concerns-of-combating-crimes-with-artificial-intelligence-surveillance-and-facial-a5eb7a09abb1>

Borges, A. și alții, 2021. The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*(57).

Bostrom, N. & Yudkowsky, E., 2014. The Ethics of Artificial Intelligence. În: K. a. R. W. E. Frankish, ed. *Cambridge Handbook of Artificial Intelligence*. New York: Cambridge University Press, pp. 316-334.

Brookhouse, O., 2022. *Can Artificial Intelligence understand emotions?*. [Interactiv]

Available at: <https://business.blogthinkbig.com/can-artificial-intelligence-understand-emotions/>

Brundage, M. e. a., 2018. *The malicious use of artificial intelligence : forecasting, prevention, and mitigation*. Oxford: Future of Humanity Institute.

Brynjolfsson, E. & McAfee, A., 2017. Artificial intelligence, for real. *Harvard business review* (1), pp. 1-31.

Bughin, J. & van Zeebroeck, N., 2017. "The best response to digital disruption." *MIT Sloan Management Review*, 58(4), pp. 80-86.

BUITEN, M., 2019. Towards Intelligent Regulation of Artificial Intelligence.. *European Journal of Risk Regulation*, 10(1), pp. 41-59.

Buolamwini, J. & Gebru, T., 2018. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research* 81:77-91.. [Interactiv] Available at: <https://proceedings.mlr.press/v81/buolamwini18a.html>. [Accesat 25 Noiembrie 2021].

Burt, A., 2020. *Ethical Frameworks for AI Aren't Enough*. [Interactiv] Available at: <https://hbr.org/2020/11/ethical-frameworks-for-ai-arent-enough>

Bytniewski, A. și alții, 2020. The functionalities of cognitive technology in management control system.. *Intelligent Information and Database Systems. Part II. DOI: http://doi.org/10.1007/978-3-030-42058-1_19*, pp. 230-240.

Caliskan, A., Bryson, J. & Narayanan, A., 2017. Semantics derived automatically from language corpora contain human-like biases.. *Science. Apr 14;*, 356(6334), pp. 183-186.

Calo, R. & Citron, D. K., 2020. The Automated Administrative State: A Crisis of Legitimacy. *Emory Law Journal, Forthcoming*.

Calvo, R. & Peters, D., 2019. *Design for Wellbeing - Tools for Research, Practice and Ethics*.. New York, NY, In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19). Association for Computing Machinery.

Campolo, A., Sanfilippo, M., Whittaker, M. & Crawford, K., 2017. *AI now 2017 report*.. [Interactiv] Available at: https://assets.ctfassets.net/8wprhvnpc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/AI_Now_Institute_2017_Report_.pdf [Accesat 14 04 2022].

Canadei, G., 2015. *What is ethics?*. [Interactiv] Available at: <https://www.canada.ca/en/treasury-board-secretariat/services/values-ethics/code/what-is-ethics.html> [Accesat 12 02 2021].

Cath, C., 2018. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Phil. Trans. R. Soc. A.*, 376(2133).

Cerna Collectif, ,, 2018. “*Research ethics in machine learning*”.. [Interactiv]
Available at: http://cerna-ethics-allistene.org/digitalAssets/54/54730_cerna_2017_machine_learning.pdf

Chandola, V., Banerjee, A. & Kumar, V., 2009. Anomaly detection: A survey.. *ACM Computing Surveys* 41(3), pp. 1-58.

Charlesworth, A., 2021. Regulating Algorithmic Assemblages: Looking beyond Corporatist AI Ethics.. În: U. K. & J. Eisler, ed. *Data-Driven Personalisation in Markets, Politics and Law*. Cambridge : Cambridge University Press, pp. 243-262.

Chow, S., 2020. *Ethical AI starts with Good Data: Running a Data Design Sprint with your Team*.. [Interactiv]
Available at: <https://towardsdatascience.com/good-ai-starts-with-good-data-running-a-data-design-sprint-with-your-team-b478e0faa3b>
[Accesat 28 Mai 2022].

Chuah, S. și alții, 2016. Wearable technologies: The role of usefulness and visibility in smartwatch adoption. *Computers in Human Behavior*(65), pp. 276-284.

Chui, M., Manyika, J. & Miremadi, M., 2016. *Where machines could replace humans—and where they can't (yet)*. [Interactiv]
Available at: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet>
[Accesat 17 Iulie 2022].

Cioara, I., 2013. *Etica I-II. Note de curs*. [Interactiv]
Available at: <https://www.scribd.com/doc/137978264/Hedonism-etica>
[Accesat 10 3 2021].

Ciobanu, C. A. & Meșniță, G., 2021. AI ethics in business - a bibliometric approach. *Review of economic and business studies*, Volume 14,(2), pp. pp. 169-202.

Ciobanu, C. A. & Meșniță, G., 2022. *AI Ethics for Industry 5.0 – From Principles to Practice*. Valencia, Spain, CEUR Workshop Proceedings (CEUR-WS.org) .

Clark, R., 2019. *Principles for AI: a SourceBook*”.. [Interactiv]
Available at: <http://www.rogerclarke.com/EC/GAIP.html>

Comisia Europeană, E., 2015. *Autonomous systems*. [Interactiv]
Available at: <https://europa.eu/eurobarometer/surveys/detail/2018>
[Accesat 29 Iunie 2022].

Comptia, 2023. *ARTIFICIAL INTELLIGENCE IN BUSINESS: TOP CONSIDERATIONS BEFORE IMPLEMENTING AI*. [Interactiv]
Available at: <https://connect.comptia.org/content/guides/business-considerations-before-implementing-ai>
[Accesat 23 iunie 2023].

Conitzer, V., Sinnott-Armstrong, M. V. & Borg, S., 2017. *Moral Decision Making Frameworks for Artificial Intelligence. Proceedings of the AAAI Conference on Artificial Intelligence vol. 31, pp. 4831..* Palo Alto, California, AAAI Press.

Constantinescu, M., 2021. *Cât de inteligent putem utiliza IA. Inteligență Artificială, externalități morale și reglementare permisivă..* [Interactiv]
Available at: https://www.cnr-unesco.ro/uploads/media/f951_etica_ai.pdf.pdf
[Accesat 4 10 2022].

Coombs, C., Hislop, D., Taneva, S. K. & Barnard, S., 2020. The strategic impacts of Intelligent Automation for knowledge and service work: An interdisciplinary review. *The Journal of Strategic Information Systems* (24)4.

Cooper, R. B. & Zmud, R. W., 1990. Information technology implementation research: A technological diffusion approach.. *Management Science*, 36(2), pp. 123-139.

Council of Europe, 2019. *“European ethical charter on the use of artificial intelligence in judicial systems and their environment”..* [Interactiv]
Available at: <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>

Council of Europe, 2022. *AI and control of Covid 19 coronavirus*. [Interactiv]
Available at: <https://www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus>

Craigon, P., Sacks, J. & Brewer, S., 2023. Ethics by design: Responsible research & innovation for AI in the food sector. *Journal of Responsible Technology*, Volumul 13.

Crawford, K. & Calo, R., 2016. There is a blind spot in AI research.. *Nature*, Volumul 538, p. 311–313 .

Cummings, M., 2006. Integrating ethics in design through the value-sensitive design approach.. *SCI ENG ETHICS*, Volumul 12, p. 701–715.

Danks, D. & London, A., 2017. *Algorithmic Bias in Autonomous Systems. International Joint Conference on Artificial Intelligence*.. Melbourne, s.n.

Darbinyan, R., 2022. *What Are AI APIs, and How Do They Work?*. [Interactiv] Available at: <https://www.dataversity.net/what-are-ai-apis-and-how-do-they-work/#>

Datatilsynet, 2018. “*Artificial intelligence and privacy*”.. [Interactiv] Available at: <https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/ai-and-privacy/>

Davenport, T. H. & Ronanki, R., 2018. Artificial intelligence for the real world. *Harvard Business Review*, 96(1), pp. 108-116.

Davis, F., 1985. *A Technology Acceptance Model for Empirically Testing New End-User Information Systems*. [Interactiv] Available at: https://www.researchgate.net/publication/35465050_A_Technology_Acceptance_Model_for_Empirically_Testing_New_End-User_Information_Systems

Davis, F., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), pp. 319-340.

Dawson D. et al., 2019. “*Artificial intelligence: Australia’s ethics framework, Australian Government*”.. [Interactiv] Available at: <https://webarchive.nla.gov.au/awa/20220603081853/https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework>

De Hert, P. & Papaconstantinou, V., 2016. The New General Data Protection Regulation: Still a Sound System for the Protection of Individuals?. *Computer Law and Security Review*, Volumul 32, pp. 179-194.

Demiaux, V. & Abdallah, Y., 2017. “*How can humans keep the upper hand? the ethical matters raised by algorithms and artificial intelligence*”., Paris: Report on the public debate led by the French Data Protection Authority (CNIL) as part of the ethical discussion assignment set by the digital republic bill.

Demlehner, Q. & Laumer, S., 2020. “*Shall We Use It or Not? Explaining the Adoption of Artificial Intelligence for Car Manufacturing Purpose*”. Marakech, s.n.

Dhar, P., 2020. The carbon impact of artificial intelligence.. *Nat Mach Intell* , Volumul 2, pp. 423-425.

Dignum, V., Baldoni, M. & Baroglio, C., 2018. *Ethics by Design: Necessity or Curse? In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18)*.. New York, Association for Computing Machinery, p. 60–66.

Doshi-Velez, F. & Kim, B., 2017. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv: Machine Learning*.

Ecoetica, 2019. *Curs - 4 teorii etice*. [Interactiv] Available at: <https://www.studocu.com/ro/document/universitatea-din-pitesti/ecoetica/c4-ecoetica-curs-4-teorii-etice-10-pag-pag-29-38/6509861> [Accesat 14 12 2023].

Eloundou, T., Manning, S. & Mishkin, P. R. D., 2023. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. *arXiv*.

Endsley, M. R. & Kaber, D. B., 1999. Level of automation effects on performance, situation awareness and workload in a dynamic control task.. *Ergonomics*, 42(3), p. 462–492.

Enholt, I. M., Papagiannidis, E., Mikalef, P. & Krogstie, J., 2021. Artificial Intelligence and Business Value: a Literature Review. *Inf Syst Front*.

Etică, C. d., 2020. *What is ethics?*. [Interactiv] Available at: <https://ethics.org.au/about/what-is-ethics/> [Accesat 12 2 2021].

Etzioni, A. & Etzioni, O., 2017. Incorporating Ethics into Artificial Intelligence.. *The Journal of Ethics*, 21(4), pp. 403-418.

European Commission, 2019. *European Commission*. [Interactiv] Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

European Parliament, 2017. “*Report with recommendations to the commission on civil law rules on robotics*”.. [Interactiv] Available at: https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html

Europeană, C., 2020. *White Paper on Artificial Intelligence: a European approach to excellence and trust*. [Interactiv] Available at: <https://commission.europa.eu/publications/white-paper-artificial->

intelligence-european-approach-excellence-and-trust_en
[Accesat 23 09 2021].

European, P., 2023. *Ce ste Inteligența Artificială și cum este utilizată?*. [Interactiv]
Available at: <https://www.europarl.europa.eu/news/ro/headlines/society/20200827STO85804/ce-este-inteligența-artificiala-si-cum-este-utilizata>
[Accesat 29 Iunie 2023].

European Commission, 2021. *European Commission*. [Interactiv]
Available at: <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>

Fahmideh, M. și alții, 2022. *Ethics of AI: A Systematic Literature Review of Principles and Challenges*. [Interactiv]
Available at: https://www.researchgate.net/publication/354651489_Ethics_of_AI_A_Systematic_Literature_Review_of_Principles_and_Challenges
[Accesat 12 02 2023].

FATML, 2016. *“Principles for accountable algorithms and a social impact statement for algorithms”*. [Interactiv]
Available at: <https://www.fatml.org/resources/principles-for-accountable-algorithms>

Finch, G., Goehring, B. & Marshall, A., 2017. The enticing promise of cognitive computing: high-value functional efficiencies and innovative enterprise capabilities. *Strategy & Leadership*, 45(6), p. 26–33.

Floridi, L., 2019. Establishing the rules for building trustworthy AI. *Nature Machine Intelligence* 1, pp. 261-262.

Floridi, L., 2019. Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, 1(6), pp. 261-262.

Floridi, L. & Cows, J., 2019. A united framework of five principles for AI in society. *Harvard Data Science Review* 1(1).

Floridi, L., Cows, J. & al., e., 2018. AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), pp. 689-707.

Fountaine, T., McCarthy, B. & Saleh, T., 2019. Building the AI-powered organization.. *Harvard Business Review*, 97(4), p. 62–73.

Franzke, A., 2022. An exploratory qualitative analysis of AI ethics guidelines. *Journal of Information, Communication and Ethics in Society*, Vol. 20 No. 4, pp. 401-423..

Future Advocacy, 2019. "*Ethical, social, and political challenges of artificial intelligence in health*". [Interactiv]
Available at: <https://futureadvocacy.com/publications/ethical-social-and-political-challenges-of-artificial-intelligence-in-health/>

Future of Life Institute, 2017. "*Asilomar AI principles*". [Interactiv]
Available at: <https://futureoflife.org/2017/08/11/ai-principles/>

Garcia, L., 2020. *Publish your models with a standard based REST API*. [Interactiv]
Available at: <https://towardsdatascience.com/publish-your-models-with-a-standard-rest-api-94de1e1da682>

Garimella, V. & Weber, I., 2017. A Long-Term Analysis of Polarization on Twitter. *ArXiv*, abs/1703.02769..

Gartner, 2019. *Report*. [Interactiv]
Available at: <https://www.gartner.com/en/information-technology/insights/artificial-intelligence>.

Giachetti, R. E., 2010. "*Design of Enterprise Systems, Theory, Architecture, and Methods*". Boca Raton : CRC PRes.

Gilburt, B., 2019. "*Women leading in AI: 10 principles of responsible AI*". [Interactiv]
Available at: <https://towardsdatascience.com/women-leading-in-ai-10-principles-for-responsible-ai-8a167fc09b7d>

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning*. s.l.:The MIT Press.

GREEN, B. P., 2018. Ethical Reflections on Artificial Intelligence.. *Scientia et Fides [online]*, 6(2), p. 9–31.

Gregory, R. W., Henfridsson, O., Kaganer, E. & Kyriakou, H., 2021. The Role of Artificial Intelligence and Data Network Effects for Creating User Value. *Academy of Management Review* (46), pp. 534-551.

Gruber, G., 2015. *Generalități despre etică, morală și deontologie. Un scurt istoric..* [Interactiv]

Available at: https://www.researchgate.net/publication/311406664_Generalitati_despre_etica_morala_si_deontologie_Un_scurt_istoric
[Accesat 3 3 2021].

Haan, K., 2023. *Forbes.* [Interactiv]

Available at: <https://www.forbes.com/advisor/business/ai-statistics/>
[Accesat 12 May 2023].

Haas, L., Gießler, S. & Thiel, V., 2020. *In the realm of paper tigers – exploring the failings of AI ethics guidelines.* [Interactiv]

Available at: <https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/>

Hagendorff, T., 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*,30(1), pp. 99-120.

Hallensleben, S. & Hustdet, C., 2020. *From Principles to Practice:An interdisciplinary framework to operationalise AI ethics.* [Interactiv]

Available at: <https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73ae4a489d6249f5ee2b9f/aieig--report---download-hb-data.pdf> 2019
[Accesat 15 February 2022].

Hanna, R. & Kazim, E., 2021. Philosophical foundations for digital ethics and AI Ethics: a dignitarian approach.. *AI Ethics*, 1(4), pp. 405-423.

Hardt, M., Price, R. & Srebro, N., 2016. Equality of Opportunity in Supervised Learning. *arXiv*.

Hengstler, M., Enkel, E. & Duelli, S., 2016. Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*(105), pp. 105-120.

High-Level Group of Experts on AI, 2019. *The Ethics Guidelines for Trustworthy Artificial Intelligence (AI).* [Interactiv]

Available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

Hoffmann, D. & Masucci, R., 2018. *Intel's AI Privacy Policy.* [Interactiv]

Available at: <https://www.intel.com/content/dam/www/public/us/en/ai/documents/Intels-AI-Privacy-Policy-White-Paper-2018.pdf>

Holmes, F., 2019. *AI Will Add \$15 Trillion To The World Economy By 2030*. [Interactiv]

Available at: <https://www.forbes.com/sites/greatspeculations/2019/02/25/ai-will-add-15-trillion-to-the-world-economy-by-2030/?sh=23feca071852>

[Accesat 23 10 2021].

Hsieh, P., 2015. Healthcare professionals' use of health clouds: Integrating technology acceptance and status quo bias perspectives. *International Journal of Medical Informatics* 84(7), pp. 512-523.

Hsieh, P., 2016. An empirical investigation of patients' acceptance and resistance toward the health cloud: The dual factor perspective. *Computers in Human Behavior*, pp. 959-969.

Hurbean, L. D. O., Munteanu, V. & Danaiaata, D., 2022. Effects of Instant Messaging Related Technostress on Work Performance and Well-Being. *Electronics*, 11(6).

IBM, 2018. *"Everyday ethics for artificial intelligence"*. [Interactiv]

Available at: <https://www.ibm.com/design/ai/ethics/everyday-ethics/>

IEEE, 2019. "Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," in *Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems* ,. pp. 1-294.

IIIM, 2015. *Ethics policy*. [Interactiv]

Available at: <https://www.iiim.is/2015/08/ethics-policy/>

Inglis, A., 2008. Approaches to the Validation of Quality Frameworks for E-Learning, 16(4),. *Quality Assurance in Education: An International Perspective*, pp. 347-362.

Institute of Business Ethics, 2018. *"Business ethics and artificial intelligence"*. [Interactiv]

Available at: <https://www.ibe.org.uk/resource/ibe-briefing-58-business-ethics-and-artificial-intelligence-pdf.html>

J.R. Searle, 1980. Minds, Brains, and Programs. *Behavioral and Brain Sciences*, vol. 3, no. 3, p. 417–457.

Jarrahi, M., 2018. Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making.. *Business Horizons*, Volumul 61, pp. 577-586.

Jobin, A., Ienca, M. & Vayena, E., 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9), pp. 389-399.

Kazim, E. & Koshyiama, A., 2021. A high-level overview of AI ethics. *Patterns*, 2(9).

Keding, C., 2020. Understanding the interplay of artificial intelligence and strategic management: four decades of research in review.. *Management Review Quarterly*, 71(1), p. 91–134.

Kesidou, S. & Sovacool, B. K., 2019. Supply chain integration for low-carbon buildings: A critical interdisciplinary review. *Renewable and Sustainable Energy Reviews, Elsevier*, 113(C), pp. 1-11.

Key, C., 1997. *Notes on Deontology*. [Interactiv] Available at: <http://sites.wofford.edu/kaycd/deontology/> [Accessed 10 3 2021].

Kieslich, K., Keller, B. & Starke, C., 2022. (2022). Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence.. *Big Data & Society*, , 9(1).

Kim, H., Chan, C. & Gupta, S., 2007. Value-based Adoption of Mobile Internet: An empirical investigation. *Decision Support Systems* 43(1), pp. 111-126.

Kim, L., 2018. *Human Ethics for Artificial Intelligent Beings*. [Interactiv] Available at: <https://aistrategyblog.com/2018/05/31/human-ethics-for-artificial-intelligent-beings/>

Kim, Y., Park, Y. & Choi, J., 2017. A study on the adoption of IoT smart home service: using Value-based Adoption Model. *Total Quality Management & Business Excellence*, 28:9-10, pp. 1149-1165.

Kleinberg, J., Mullainathan, S. & Raghavan, M., 2017. *Inherent Trade-Offs in the Fair Determination of Risk Scores*. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67, of *Leibniz International Proceedings in Informatics (LIPIcs)*. s.l., s.n.

Koski, O. & Husso, K., 2018. *Work in the age of artificial intelligence: Four perspectives on the economy, employment, skills and ethics*. [Interactiv] Available at: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/TEMjul_21_2018_Work_in_the_age.pdf

Krkač, K., 2018. Corporate social irresponsibility: humans vs artificial intelligence.. *Social Responsibility Journal*, 15(6), pp. 786-802.

Lacurezeanu, R., Tiron-Tudor, A. & Bresfelean, V. P., 2020. Robotic Process Automation in Audit and Accounting. *Audit Financiar*, Volumul vol. XVIII, no. 4(160), pp. . 752-770.

Lahlali, M., Berbiche, N. & Alami, J. E., 2021. How Enterprise Must Be Prepared To Be “Ai First”?.. *IJACSA*, 12(5), pp. 346-351.

Latonero, M., 2018. *Governing Artificial Intelligence: upholding human rights & dignity*.. [Interactiv]
Available at: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

Lee, J., Suh, T., Roy, D. & Baucus, M., 2019. Emerging Technology and Business Model Innovation: The Case of Artificial Intelligence... *Journal of Open Innovation: Technology, Market, and Complexity*. 5(3):44.

Lee, Y., Kozar, K. & Larsen, K. T. T. A. M. P. P. a. F., 2003. The Technology Acceptance Model: Past, Present. *Communications of the Association for Information Systems* (12), pp. 751-782.

Leung, L., 2015. Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care* 4(3);, pp. 324-327.

Levin, B. & Downes, L., 2023. Microsoft, Google, and a New Era of Antitrust (February 17, 2023).. *Harvard Business Review*, , Volumul Disponibil la SSRN: <https://ssrn.com/abstract=4381807> or <http://dx.doi.org/10.2139/ssrn.4381807>.

Levine, J., 2022. *Combining Intelligence: How People And AI Can Collaborate*. [Interactiv]
Available at: <https://www.forbes.com/sites/forbesbusinesscouncil/2022/04/26/combining-intelligence-how-people-and-ai-can-collaborate/?sh=376e59e8143d>

Li, B. și alții, 2021. *Trustworthy AI: From Principles to Practices*. [Interactiv]
Available at: <https://arxiv.org/abs/2110.01167>

Lin, T., Wu, S., Hsu, J. & Chou, Y., 2012. The integration of value-based adoption and expectation–confirmation models: An example of IPTV continuance intention. *Decision Support Systems* 54(1), pp. 63-75.

Liu, H. & Lang, B., 2019. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey.. *Appl. Sci.* , 9(20).

Loyola-González, O., 2019. "Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View,". *IEEE Access*(7), pp. 154096-154113.

Lunney, A., Cunningham, N. & Eastin, M., 2016. Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. *Computers in Human Behavior*(65), pp. 114-120.

Madaio, M., Stark, L., Vaughan, J. & Wallach, H., 2020. *Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI*. s.l., s.n., pp. 1-14.

Makarius, E. E., Mukherjee, D., Fox, J. D. & Fox, A. K., 2020. Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization.. *Journal of Business Research*, 120, pp. 262-273.

Mallat, N., 2007. Exploring consumer adoption of mobile payments – A qualitative study. *The Journal of Strategic Information Systems* 16(4), pp. 413-432.

Manyika, J. e. a., 2017. *A Future That Works: Automation, Employment and Productivity*. [Interactiv]
Available at: <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages>
[Accesat 18 Iulie 2022].

Mathieson, K., 1991. Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior. *Information Systems Research* 2(3), pp. 173-191.

McCarthy, J. și alții, 2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955.. *AI Magazine*, 27(4).

McKinsey, 2021. *McKinsey Report*. [Interactiv]
Available at: <https://www.mckinsey.com/business-functions/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021>

McKinsey-Global-Institute, 2018. *Notes from the AI frontier: Modeling the impact of AI on the world economy*. [Interactiv]
Available at: <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the->

world-economy

[Accesat 12 6 2020].

Mechkane, I., Mehra, M., Laurent, A. & Ross, E., 2022. *Artificial Intelligence and Machine Learning: Enhancing Human Effort with Intelligent Systems*. [Interactiv]

Available at: <https://www.isa.org/intech-home/2022/august-2022/features/artificial-intelligence-and-machine-learning-enhan>

[Accesat 12 Ianuarie 2023].

MI Garage, 2019. *Ethics framework*. [Interactiv]

Available at: <https://migarage.digicatapult.org.uk/ethics/ethics-framework/>

Microsoft, 2017. *Responsible AI*. [Interactiv]

Available at: <https://www.microsoft.com/en-us/ai/responsible-ai>

Mikalef, P. & Gupta, M., 2021. Artificial Intelligence Capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance.. *Information & Management, Online*. <https://doi.org/10.1016/j.im.2021.103434>.

Miller, C. & Coldicott, R., 2023. *People, power and technology: The tech workers' view..* [Interactiv]

Available at: <https://doteveryone.org.uk/report/workersview/>

[Accesat 18 February 2023].

Minevich, M., 2022. *How is Artificial Intelligence Transforming Humanity in Every Dimension*. [Interactiv]

Available at: <https://www.digitalfirstmagazine.com/how-is-artificial-intelligence-transforming-humanity-in-every-dimension/>

[Accesat 26 Iunie 2023].

Mishra, A., 2018. *Metrics to evaluate your machine learning algorithm*. [Interactiv]

Available at: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>

Mitchell, M. și alții, 2019. *Model cards for model reporting*. In *Proceedings of the Conference on Fairness, Accountability, and Transparency..* Atlanta, GA, USA,, s.n.

Mittelstadt, B., 2019. Principles alone cannot guarantee ethical AI.. *Nat Mach Intell*, Volumul 1, p. pp. 501–507.

Mona, A., Madan, R. & Sivarajah, U., 2022. Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information Management*, 62.

Moor, J., 2006. The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems*, pp. 18-21.

Morley, J., Elhalal, A. & Garcia, F., 2021. Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds & Machines*, Volumul 31, p. pp. 239–256.

Morley, J., Floridi, L. & Kinsey, L., 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices.. *Sci Eng Ethics*, Volumul 26, p. 2141–2168.

Morley, J., Kinsey, L. & Elhalal, A., 2023. Operationalising AI ethics: barriers, enablers and next steps.. *AI & Soc* , Volumul 38, p. pp. 411–423.

Moustafa, N., Hu, J. & Slay, J., 2019. *Journal of Network and Computer Applications*, Volumul 128, pp. 33-55.

Nicol, A. A. și alții, 2018. Comparison of high-technology active learning and low-technology active learning classrooms.. *Active Learning in Higher Education*, 19(3), p. 253–265.

O'Neill, C., 2016. *O'Neil, Cathy. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown .

O~,î, fără an s.l.:s.n.

OECD, 2019. *Forty-two countries adopt new OECD Principles on Artificial Intelligence*. [Interactiv]

Available at: <https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm>

OECD, 2019. *Recommendation of the Council on Artificial Intelligence - OECD*. [Interactiv]

Available at: legalinstruments.oecd.org/api/print?ids=648&lang=en

Oladele, S., 2023. *A Comprehensive Guide on How to Monitor Your Models in Production*. [Interactiv]

Available at: <https://neptune.ai/blog/how-to-monitor-your-models-in-production-guide>

OpenGroup, 2008. *TOGAF® Standard, Version 9.2.* [Interactiv]
Available at: <https://pubs.opengroup.org/architecture/togaf9-doc/arch/index.html>

O'Reilly, 2020. *O'Reilly Report.* [Interactiv]
Available at: https://get.oreilly.com/ind_ai-adoption-in-the-enterprise-2020.html

Orell, B., 2023. *How Workers Can Prepare for the AI Economy.* [Interactiv]
Available at: <https://www.aei.org/podcast/brent-orrell-how-workers-can-prepare-for-the-ai-economy/>
[Accesat 24 Mai 2023].

Page, M. J. și alții, 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 :n71.

Panetta, K., 2019. *Gartner.* [Interactiv]
Available at: <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018>

Parasuraman, R. & Riley, V., 1997. Humans and Automation: Use, Misuse, Disuse, Abuse.. *Human Factors: The Journal of Human Factors and Ergonomics Society*, Volumul 39, pp. 230 - 253.

Parasuraman, R., Thomas, S. & Christopher, W., 2000. A model for types and levels of human interaction with automation.. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* , 30(3), pp. 286-297.

Parry, K., Cohen, M. & Bhattacharya, S., 2016. Rise of the Machines: A Critical Consideration of Automated Leadership Decision Making in Organizations.. *Group & Organization Management*, 41(5), p. 571–594.

Passanen, J., 2022. *AI Ethics: How To Act Ethically in an Automated World.* [Interactiv]
Available at: <https://www.ultimate.ai/blog/collaborations/ai-ethics>
[Accesat 5 May 2023].

Perifanis, N.-A. & Kitsios, F., 2023. Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review.. *Information* , 14(85), pp. 1-42.

Polachowska, K., 2019. *Neoteric.* [Interactiv]
Available at: <https://neoteric.eu/blog/12-challenges-of-ai-adoption/>

Price, W. & Cohen, I., 2019. Privacy in the age of medical big data. *Nature Medicine*, Volumul 25,, pp. 37-43.

Pumplun, L., Tauchert, C. & Heidt, M., 2019. *A new organizational chassis for artificial intelligence-exploring organizational readiness factors*. [Interactiv]
Available at: https://aisel.aisnet.org/ecis2019_rp/106

Ramzai, J., 2020. *Clearly Explained: 4 types of Machine learning algorithms*. [Interactiv]
Available at: <https://towardsdatascience.com/clearly-explained-4-types-of-machine-learning-algorithms-71304380c59a>
[Accesat 17 Iulie 2022].

Rao, S., Golbin, I. & Tembhurnikar, V., 2021. *Six stage gates to a successful AI governance*. [Interactiv]
Available at: <https://towardsdatascience.com/six-stage-gates-to-a-successful-ai-governance-14ab0787a380>

Rathenau Institute, 2017. *“Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality”*.. [Interactiv]
Available at: <https://www.rathenau.nl/en/digitale-samenleving/human-rights-robot-age>

Resseguier, A. & Rodrigues, R., 2020. AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, 7(2).

Ribeiro, M., Singh, S. & C., G., 2016. *“why should I trust you?”: Explaining the predictions of any classifier*. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*.. San Francisco, CA, USA, s.n.

Roff, H. M. & Moyes, R., 2016. Meaningful Human Control. În: *Artificial Intelligence and Autonomous Weapons. Article36 briefing paper prepared for the CCW informal meeting of experts on lethal autonomous weapons systems*.. s.l.:s.n.

Ross, J., P., W. & Robertson, D., 2006. *Enterprise Architecture as Strategy*. Boston: Harvard Business Press.

Royakkers, L., Timmer, J. & Kool, L. e. a., 2018. Societal and ethical issues of digitization. *Ethics Inf Technol* 20, p. 127–142.

Ryan, M. & Stahl, B., 2021. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society* 19(1), pp. 61-86.

Saadat, S., 2023. *AI vs. Humans: Who's More Productive?*. [Interactiv]
Available at: <https://www.linkedin.com/pulse/ai-vs-humans-whos-more-productive-sakib-saadat/>
[Accesat 9 Iunie 2023].

Sage, 2017. *The ethics of code: Developing AI for business with five core principles*.. [Interactiv]
Available at: <https://www.sage.com/en-au/blog/the-ethics-of-code-developing-ai-for-business-with-five-core-principles-2/>

Sanderson, C. și alții, 2022. *AI Ethics Principles in Practice: Perspectives of Designers and Developers*. [Interactiv]
Available at: <https://arxiv.org/abs/2112.07467>

Sarter, N. B. & Schroeder, B., 2011. Supporting decision making and action selection under time pressure and uncertainty: the case of in-flight icing. *Human factors*, 43(4), p. 573–583.

Sayak, P., 2018. *Turning Machine Learning Models into APIs in Python*. [Interactiv]
Available at: <https://www.datacamp.com/tutorial/machine-learning-models-api-python>

Schmidt, R., Zimmermann, A., Moehring, M. & Keller, B., 2020. *Value creation in connectionist artificial intelligence—A research agenda*. *AMCIS*. [Interactiv]
Available at: https://aisel.aisnet.org/amcis2020/ai_semantic_for_intelligent_info_systems/ai_semantic_for_intelligent_info_systems/14/

Schneier, B., 2018. *Machine Learning Will Transform How We Detect Software Vulnerabilities*. [Interactiv]
Available at: https://www.schneier.com/essays/archives/2018/12/machine_learning_wil.html
[Accesat 25 Noiembrie 2022].

Setiawan, A., Agiwahyunto, F. & Arsiwi, P., 2019. A Virtual Reality Teaching Simulation for Exercise During Pregnancy. *International Journal of Emerging Technologies in Learning (iJET)*, 14(01), p. 34–48.

Singer, P., 2021. *Enciclopedia Britannica - Ethics*. [Interactiv]
Available at: <https://www.britannica.com/topic/ethics-philosophy>
[Accesat 23 4 2022].

Singh, S., 2020. *Machine Learning Model Deployment as REST API in Four Easy Steps*. [Interactiv]
Available at: <https://towardsdatascience.com/machine-learning-model-deployment-as-rest-api-in-four-easy-steps-e65764f7c5c>

Sjödín, D., Parida, V., Jovanovic, M. & Visnjic, I., 2020. Value creation and value capture alignment in business model innovation: a process view on outcome-based business models.. *J Prod Innov Manag* 37(2), p. 158–183.

Smith, P. J., McCoy, C. E. & Layton, C., 1997. Brittleness in the design of cooperative problem-solving systems: The effects on user performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Human*, 27(3), pp. 360-371.

Sohn, K. & Kwon, O., 2020. Technology Acceptance Theories and Factors Influencing Artificial Intelligence-based Intelligent Products. *Telematics and Informatics* 47.

Soni, N., Sharma, E., Singh, N. & Kapoor, A., 2020. Artificial Intelligence in Business: From Research and Innovation to Market Deployment. *Procedia Computer Science*, Volumul 167, pp. 2200-2210.

St. George, B., 2023. *Turing Test*. [Interactiv]
Available at: <https://www.techtarget.com/searchenterpriseai/definition/Turing-test>
[Accesat 14 Mai 2023].

Stephanidis, C., 2019. Seven HCI Grand Challenges.. *International Journal of Human-Computer Interaction*, Volumul 35, pp. 1229-1269.

Strubell, E., Ganesh, A. & McCallum, A., 2020. *Energy and Policy Considerations for Modern Deep Learning Research*. *Proceedings of the AAAI Conference on Artificial Intelligence*,. Palo Alto, California USA, AAAI Press.

Susarla, A., 2018. *How artificial intelligences can detect and create fake news*. [Interactiv]
Available at: <https://msutoday.msu.edu/news/2018/how-artificial-intelligence-can-detect-and-create-fake-news>

Taddeo, M., McNeish, D. & Blanchard, A. e. a., 2021. Ethical Principles for Artificial Intelligence in National Defence. *Philos. Technol.* 34., p. 1707–1729.

Talgala, N., 2022. *AI Ethics: What It Is And Why It Matters*. [Interactiv] Available at: <https://www.forbes.com/sites/nishatalagala/2022/05/31/ai-ethics-what-it-is-and-why-it-matters/?sh=246b92e73537> [Accesat 12 04 2023].

Tarafdar, M., Beath, C. M. & Ross, J. W., 2019. *Using Artificial Intelligence to Enhance Business Operations*. [Interactiv] Available at: <https://sloanreview.mit.edu/article/using-ai-to-enhance-business-operations/>

Taylor, S. & Todd, P., 1995. Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research* 6(2), pp. 144-176.

Tieto, 2018. *Tieto's AI ethics guidelines*. [Interactiv] Available at: <https://www.tietoevry.com/en/newsroom/all-news-and-releases/press-releases/2018/10/tieto-strengthens-commitment-to-ethical-use-of-ai/>

Torre, D. L., Colapinto, C., Durosini, I. & Triberti, S., 2018. "Team Formation for Human-Artificial Intelligence Collaboration in the Workplace: A Goal Programming Model to Foster Organizational Change,". *IEEE Transactions on Engineering Management*, pp. 1-11.

Tseng, K., Hsu, C. & Chuang, Y., 2013. Designing an intelligent health monitoring system and exploring user acceptance for the elderly. *J. Med. Syst.* 37(6).

Turilli, M., 2007. Ethical protocols design.. *Ethics and Information Technology*, 9(1), p. pp. 49–62.

Turing, A., 1950. COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, Volume LIX, Issue 236, p. 433–460.

UNESCO, 2017. "Report of COMEST on robotics ethics". [Interactiv] Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000253952.locale=en>

Uren, V. & Edwards, J., 2023. Technology readiness and the organizational journey towards AI adoption: An empirical study.. *International Journal of Information Management*, Volumul 68.

Velasquez, M., Andre, C., Shanks, T. & Meyer, M., 2010. *What is Ethics?*. [Interactiv]

Available at: <https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/>
[Accesat 23 4 2022].

Venkatesh, V., Morris, M., Davis, G. & Davis, F., 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27(3), pp. 425-478.

Venkatraman, V., 2017. *The Digital Matrix–New Rules for Business Transformation Trough Technology*. s.l.:LifeTree Press.

Verma, P. și alții, 2021. A Novel Intrusion Detection Approach Using Machine Learning Ensemble for IoT Environments.. *Appl. Sci.*, 11(21).

Vidgen, R., Hindle, G. & Randolph, I., 2020. Exploring the ethical implications of business analytics with a business ethics canvas. *European Journal of Operational Research*, 281(3), pp. pp. 491-501.

Ville, V., Kai-Kristian, K. & Pekka, A., 2019. *AI ethics in industry: A research framework*. [Interactiv]
Available at: <https://arxiv.org/abs/1910.12695>

Vinuesa, R., Azizpour, H. & Leite, I., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals.. *Nat Commun*, 11(233).

Vizard, M., 2023. *Cybersecurity AI arms race is underway*. [Interactiv]
Available at: <https://blog.barracuda.com/2023/04/03/cybersecurity-AI-arms-race>
[Accesat 16 Mai 2023].

Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K. & Wanko, C. E. T., 2020. Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects.. *Business Process Management Journal*, 26(7), p. 1893–1924.

Wang, C., 2016. A novel approach to conduct the importance-satisfaction analysis for acquiring typical user groups in business-intelligence systems. *Computers in Human Behavior*(54), pp. 673-681.

Wang, H., Huang, J. & Zhang, Z., 2019. *The impact of deep learning on organizational agility*. [Interactiv]
Available at: https://aisel.aisnet.org/icis2019/governance_is/governance_is/26/

Wanner, J. și alții, 2020. *White, Grey, Black: Effects of XAI Augmentation on the Confidence in AI-based Decision Support Systems*. India, s.n.

- Weill, P., 2007. *“Innovating with Information Systems: What Do the Most Agile Firms in the World Do?”*. Barcelona, s.n.
- Weill, P. & Woerner, S. L., 2015. Thriving in an increasingly digital ecosystem.. *MIT Sloan Management Review*, 56(4), 27, pp. 24-37.
- Wiener, N., 1954. *The human use of human beings*.. New York: Doubleday.
- Williams, M. D., Rana, N. P. & Dwivedi, Y. K., 2015. The unified theory of acceptance and use of technology (UTAUT): a literature review. *Journal of enterprise information management*., pp. 217-229.
- Wilson, J. & Daugherty, P. R., 2018. *Harvard Business Review*. [Interactiv] Available at: <https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces>
- Yang, H., Lee, H. & Zo, H., 2017. User acceptance of smart home services: an extension of the theory of planned behavior. *Industrial Management & Data Systems*, pp. 68-89.
- Yang, H., Yu, J., Zo, H. & Choi, M., 2016. User acceptance of wearable devices: An extended perspective of perceived value. *Telematics and Informatics*(33), pp. 256-269.
- Yang, K. & Jolly, L., 2009. The effects of consumer perceived value and subjective norm on mobile data service adoption between American and Korean consumers. *Journal of Retailing and Consumer Services* 16(6), pp. 502-508.
- Yao, M., Zhou, A. & Jia, M., 2018. *Applied Artificial Intelligence: A Handbook For Business Leaders*. s.l.:TOPBOTS.
- Zapanta, T., 2023. *The impact of AI in business*. [Interactiv] Available at: <https://www.microsourcing.com/learn/blog/the-impact-of-ai-on-business/#stats>
[Accesat 17 Iunie 2023].
- Zarsky, T., 2016. The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making.. *Science, Technology, & Human Values*, 41(1), p. 118–132.
- Zhou, J. și alții, 2020. A Survey on Ethical Principles of AI and Implementations. *IEEE SYMPOSIUM SERIES ON COMPUTATIONAL INTELLIGENCE (SSCI)*.

