Sciences and Techniques of Information Management
Volume 11, Issue 2, 2025

Demystifying Artificial Intelligence (AI) in Human Resource Management (HRM): A Bibliometric Analysis of Explainable Artificial Intelligence (XAI) (2013-2023)

Ehsan Farat Yazdi
Department of Business Creation entrepreneurship, Faculty of Entrepreneurship, University of Tehran, Tehran, Iran. ehsanforat@ut.ac.ir

Ehsan Chitsaz
Faculty member, Faculty of Entrepreneurship, University of Tehran, Tehran, Iran **(Corresponding author)**. chitsaz@ut.ac.ir

Mohammad Etemadi
Department of entrepreneurship development, Faculty of Entrepreneurship, University of Tehran, Tehran, Iran. mohammad.etemadi@ut.ac.ir

Abstract

Purpose: The integration of Artificial Intelligence (AI) has impacted various sectors, bringing about significant changes. However, these applications come with challenges. One sector that has notably benefited from AI is Human Resource Management (HRM). The adoption of AI in HRM has raised concerns regarding transparency. This study explores the role of Explainable Artificial Intelligence (XAI) within HRM. The importance of XAI is especially critical in sensitive areas like HRM, where AI systems must be not only efficient but also ethically responsible and transparent. Building on this premise, the study further investigates the role and potential of XAI in HRM.

Method: This study conducts a scientometric analysis of research published over the past decade (2013–2023) in the field of Explainable Artificial Intelligence (XAI) and its applications in Human Resource Management (HRM). Scientometric (bibliometric) analysis is employed as a tool to map the XAI research landscape, identify key themes and patterns, and explore its intersection with HRM. The analysis begins by examining general XAI concepts, including its principles, techniques, and potential applications, before focusing on the specific application of XAI in HRM. Additionally, co-occurrence analysis is utilized to detect themes and patterns within the body of research.

Findings: The co-occurrence analysis in this study presents two distinct knowledge maps. The first map focuses exclusively on explainable artificial intelligence (XAI), offering an overview of the field. It reveals that XAI applications are predominantly associated with the medical domain, with comparatively less emphasis on the humanities. This finding suggests that, although XAI holds significant potential for human resource management (HRM), this area remains underexplored and underutilized. The second knowledge map investigates the intersection between XAI and HRM.

Cite this article: Farat Yazdi, E., Chitsaz, E., & Etemadi, M. (2025). Demystifying Artificial Intelligence (AI) in Human Resource Management (HRM): A Bibliometric Analysis of Explainable Artificial Intelligence (XAI) (2013-2023). *Sciences and Techniques of Information Management*, 11(2), 93-132. https://doi.org/10.22091/stim.2024.10488.2075

Received: 2024-03-11 ; **Revised:** 2024-05-08 ; **Accepted:** 2024-05-27 ; **Published online:** 2024-06-05

Article type: Research Article **Published by:** University of Qom.

Interestingly, the integration of HR within AI exhibits a notable lack of correlation with core AI topics. This indicates a research and innovation gap at the convergence of these fields, potentially arising from differing disciplinary backgrounds, perspectives, or HR managers' reluctance to adopt intelligent systems. Additionally, the study identifies explainable machine learning techniques and ontological frameworks as core elements of XAI applications, which support XAI's capacity to deliver transparent and interpretable AI models. However, the limited interaction between technical AI and HR highlights the necessity for interdisciplinary research that combines HR expertise with AI to develop more relevant and effective HR tools.

Conclusion: In conclusion, this study provides an overview of the current state of explainable artificial intelligence (XAI) and its applications in human resource management (HRM). It highlights the significant potential of this field while identifying key challenges and areas for future research. The findings emphasize that AI systems, particularly in sensitive domains such as HRM, must be not only technically efficient but also ethically sound and transparent. This is essential to ensure that AI systems are used responsibly and contribute positively to HRM practices. It is hoped that this work will contribute to the ongoing dialogue about the role of AI in HRM and inspire further exploration and innovation in this dynamic field.

Keywords: Human Resource Management, Explainable Artificial Intelligence, Co-occurrence Analysis, Interpretability, Explainability, Bibliometrics.

علوم و فنون مدیریت اطلاعات
دوره ۱۱، شماره ۲، ۱۴۰۴

تبیین هوش مصنوعی در مدیریت منابع انسانی: تحلیل کتاب‌سنجی هوش مصنوعی توضیح‌پذیر (۲۰۱۳–۲۰۲۳)

احسان فرات‌یزدی
گروه کارآفرینی کسب و کار جدید، دانشکده کارآفرینی، دانشگاه تهران، تهران، ایران. ehsanforat@ut.ac.ir
احسان چیت‌ساز
گروه توسعه کارآفرینی، دانشکده کارآفرینی دانشگاه تهران (**نویسنده مسئول**). chitsaz@ut.ac.ir
محمد اعتمادی
گروه توسعه کارآفرینی، دانشکده کارآفرینی، دانشگاه تهران، تهران، ایران. mohammad.etemadi@ut.ac.ir

چکیده

هدف: هوش مصنوعی (AI) در بخش‌های متنوعی کاربرد پیدا کرده و تغییرات جدیدی را ایجاد کرده است، اما این کاربرد چالش‌هایی نیز دارد. یکی از بخش‌هایی که از AI بهره برده، مدیریت منابع انسانی است. استفاده از هوش مصنوعی در مدیریت منابع انسانی چالش شفافیت مدل را ایجاد کرده است. این مطالعه به بررسی نقش هوش مصنوعی توضیح‌پذیر (XAI) در مدیریت منابع انسانی می‌پردازد. اهمیت هوش مصنوعی توضیح‌پذیر در حوزه‌های حساس مانند مدیریت منابع انسانی برجسته است و ضروری است که سیستم‌های هوش مصنوعی مورد استفاده در HRM نه تنها کارآمد، بلکه از نظر اخلاقی نیز سالم و شفاف باشند.

روش: این مطالعه تجزیه‌وتحلیل علم‌سنجی از تحقیقات انجام شده در دهۀ گذشته (۲۰۱۳–۲۰۲۳) در زمینۀ هوش مصنوعی توضیح‌پذیر و کاربرد آن در مدیریت منابع انسانی انجام می‌دهد. در این مطالعه، تجزیه‌وتحلیل کتاب‌سنجی به‌عنوان ابزاری برای ترسیم چشم‌انداز تحقیقات، شناسایی مضامین و الگوهای کلیدی، و درک تقاطع آن با مدیریت منابع انسانی عمل می‌کند. این تحلیل با کاوش در مفاهیم کلی هوش مصنوعی توضیح‌پذیر آغاز می‌شود و سپس به کاربرد خاص آن در مدیریت منابع انسانی می‌پردازد. همچنین از تجزیه‌وتحلیل هم‌رخدادی برای شناسایی الگوها و مضامین کلیدی در تحقیقات استفاده شده است.

یافته‌ها: تجزیه‌وتحلیل هم‌رخدادی شامل دو نقشۀ دانش مجزا بود: نقشۀ اول نشان داد که برنامه‌های کاربردی هوش مصنوعی توضیح‌پذیر عمدتاً با حوزه پزشکی مرتبط هستند و تأکید کمتری بر علوم انسانی دارند؛ این یافته نشان می‌دهد که پتانسیل قابل‌توجهی برای کاربرد در مدیریت منابع انسانی وجود دارد، اما هنوز به‌طور کامل کاوش نشده است. نقشۀ دوم

استناد به این مقاله: فرات یزدی، الف.، چیت‌ساز، الف.، و اعتمادی، م. (۱۴۰۴). تبیین هوش مصنوعی در مدیریت منابع انسانی: تحلیل کتاب‌سنجی هوش مصنوعی توضیح‌پذیر (۲۰۱۳–۲۰۲۳). *علوم و فنون مدیریت اطلاعات*، ۱۱(۲)، ۹۳-۱۳۲.

https://doi.org/10.22091/stim.2024.10488.2075

تاریخ دریافت: ۱۴۰۲/۱۲/۲۱ ؛ **تاریخ اصلاح:** ۱۴۰۳/۰۲/۱۹ ؛ **تاریخ پذیرش:** ۱۴۰۳/۰۳/۰۷ ؛ **تاریخ انتشار آنلاین:** ۱۴۰۳/۰۳/۱۶

نوع مقاله: پژوهشی **ناشر:** دانشگاه قم

هم‌افزایی بین هوش مصنوعی توضیح‌پذیر و مدیریت منابع انسانی را بررسی کرد و نبود همبستگی قابل‌توجهی را نشان داد که حاکی از شکاف پژوهش و نوآوری در این حوزه است. همچنین، تکنیک‌های یادگیری ماشین قابلیت توضیح و هستی‌شناختی به‌عنوان جنبه‌های اصلی کاربرد شناسایی شدند. یافته‌ها بیانگر حداقل تعامل بین هوش مصنوعی فنی و منابع انسانی و نیاز به تحقیقات بین‌رشته‌ای هستند.

نتیجه‌گیری: این مطالعه یک نمای کلی از وضعیت فعلی هوش مصنوعی توضیح‌پذیر و کاربرد آن در مدیریت منابع انسانی ارائه می‌کند. نتایج نشان می‌دهد که این حوزه دارای پتانسیل قابل‌توجهی برای تغییر شیوه‌های مدیریت منابع انسانی است، در حالی که چالش‌ها و زمینه‌های کلیدی برای تحقیقات آینده نیز شناسایی شدند. تأکید می‌شود که سیستم‌های هوش مصنوعی در حوزه‌های حساس باید نه تنها از نظر فنی کارآمد، بلکه از نظر اخلاقی نیز سالم و شفاف باشند. این امر برای استفاده مسئولانه و ایجاد اثر مثبت در فعالیت‌های مدیریت منابع انسانی ضروری است.

این مطالعه همچنین تکنیک‌های یادگیری ماشین قابلیت توضیح و هستی‌شناختی را به‌عنوان جنبه‌های اصلی کاربرد هوش مصنوعی توضیح‌پذیر شناسایی می‌کند. این جنبه‌ها زیربنای توانایی هوش مصنوعی توضیح‌پذیر برای ارائهٔ مدل‌های هوش مصنوعی شفاف و قابل درک است. با این حال، حداقل تعامل بین هوش مصنوعی فنی و منابع انسانی نیاز به تحقیقات بین‌رشته‌ای را نشان می‌دهد که تخصص منابع انسانی را با هوش مصنوعی ترکیب می‌کند تا ابزارهای منابع انسانی مرتبط‌تر و مؤثرتر را توسعه دهد.

کلیدواژه‌ها: مدیریت منابع انسانی، هوش مصنوعی توضیح‌پذیر، تحلیل هم‌رخدادی، تفسیرپذیری، توضیح‌پذیری، تحلیل کتاب‌سنجی.

۱. مقدمه

فناوری‌ها و روش‌های جدید در چند دهۀ اخیر تغییرات چشمگیری در حوزۀ مدیریت منابع انسانی[۱] به وجود آورده‌اند. یکی از این تغییرات، استفاده از هوش مصنوعی[۲] در این حوزه است که برخی از وظایف تصمیم‌گیرندگان انسانی را برعهده گرفته است (اوپادیای[۳]، ۲۰۲۰؛ اعتمادی و همکاران، ۱۴۰۳). همچنین، پیشرفت فناوری اطلاعات و ارتباطات (ICT) و تولید حجم عظیم داده‌های سازمانی[۴]، نقش هوش مصنوعی در مدیریت منابع انسانی را تقویت کرده است (شولز[۵]، ۲۰۱۹). هوش مصنوعی می‌تواند به سازمان‌ها کمک کند تا دانش بیشتری به دست آورند (حسن‌زاده، ۲۰۲۲) و مزیت رقابتی خود را افزایش دهند (پریفانیس و کیتسیوس[۶]، ۲۰۲۳). در این زمینه، مدیریت دانش (KM) به‌عنوان یک فرایند استراتژیک برای ایجاد، اشتراک‌گذاری، و بهره‌برداری از دانش در سازمان یا جامعه، اهمیت زیادی دارد (جاتوبا و همکاران[۷]، ۲۰۱۹). مدیریت دانش می‌تواند با ایجاد و انتشار محتوای مرتبط و به‌روز، تسهیل همکاری و تعامل بین ذی‌نفعان (سان و همکاران[۸]، ۲۰۱۸)، و فراهم کردن امکان انطباق و نوآوری در رویه‌های سازمانی (شیرعلیان و همکاران، ۲۰۲۴، ۲۰۲۵)، کیفیت و اثربخشی فرایندها و نتایج سازمانی را بهبود بخشد (فریرا و همکاران[۹]، ۲۰۲۰). اما رفتار و تصمیمات مغرضانه هوش مصنوعی و فریب انسان، مانند تشخیص اشتباه سرطان بیماران از سوی واتسون شرکت آی‌بی‌ام (استریکلند[۱۰]، ۲۰۱۹) و توییت‌های نژادپرستانه و جنسیت‌گرایانه مایکروسافت تای[۱۱] (نف و نگی[۱۲]، ۲۰۱۶) یک نگرانی اساسی در مورد اعتماد به تصمیمات مبتنی بر هوش مصنوعی ایجاد می‌کند: قابلیت تفسیر و درک مدل‌های تصمیم‌گیری هوش مصنوعی. در حوزه‌های پیچیده و حساسی که در آن‌ها پتانسیل انسانی و فرهنگ سازمانی مطرح هستند، مانند مدیریت منابع انسانی، نیاز به شفافیت در تصمیمات هوش مصنوعی

1. Human Resource Management (HRM)
2. Artificial Intelligence (AI)
3. Upadhyay et al.
4. Corporate Big Data
5. Scholz
6. Perifanis & Kitsios
7. Jatobá et al.
8. Sun et al.
9. Ferreira et al.
10. Strickland
11. Microsoft Tay
12. Neff & Nagy

بیش از پیش مشهود است. هوش مصنوعی توضیح‌پذیر (XAI) زیرشاخه‌ای از هوش مصنوعی است که هدف آن ارائهٔ تصمیمات قابل درک و تفسیرپذیرتر در حوزه‌های پیچیده و پرمخاطره است (گونینگ و همکاران[۱] ۲۰۱۹؛ هپنستال و مک‌نیش[۲]، ۲۰۲۰؛ سکل و فلچ[۳]، ۲۰۲۰).

در این پژوهش، نقش هوش مصنوعی در مدیریت منابع انسانی و چالش‌های شفافیت و تفسیرپذیری تصمیمات مبتنی بر هوش مصنوعی بررسی شد. سؤال اصلی این پژوهش این است که چگونه می‌توان تکنیک‌های هوش مصنوعی توضیح‌پذیر را در مدیریت منابع انسانی ادغام کرد تا مدل‌های هوش مصنوعی را قابل درک‌تر کرد. این موضوع از آن رو مهم است که تصمیمات مدیریت منابع انسانی مستقیما بر موفقیت سازمانی و رفاه کارکنان تأثیر می‌گذارند (آرمسترانگ[۴]، ۲۰۱۴) و در عین حال، مدل‌های هوش مصنوعی معمولا به عنوان جعبه‌های سیاه[۵] (رودین[۶]، ۲۰۱۹) عمل می‌کنند که دلایل و مبانی تصمیمات آن‌ها برای انسان‌ها قابل درک و توجیه نیستند (دس و راد[۷]، ۲۰۲۰). این مسئله می‌تواند باعث ایجاد مشکلات اخلاقی، حقوقی، و عملی در مدیریت منابع انسانی شود، مانند تبعیض، نپاسخ‌گو نبودن، کاهش اعتماد و رضایت کارکنان (ادیب‌فر و همکاران، ۲۰۲۴؛ قدرتی‌زاده و همکاران، ۲۰۲۴)، و کاهش کیفیت و اثربخشی فرایندها و نتایج سازمانی (آیزنبرگ و هوون[۸]، ۲۰۲۰؛ فلوریدی و همکاران[۹]، ۲۰۱۸؛ جوبین و همکاران[۱۰]، ۲۰۱۹)

برای پاسخ به سؤل پژوهشی خود، از یک روش تحلیل کتاب‌سنجی برای ساخت دو نقشه دانش، یکی برای هوش مصنوعی توضیح‌پذیر (XAIS) و دیگری برای تعامل بین هوش مصنوعی توضیح‌پذیر و مدیریت منابع انسانی (XHRS) بر اساس ادبیات علمی منتشر شده از سال ۲۰۱۲ تا ۲۰۲۲ استفاده می‌کنیم. نقشه‌های دانش، خوشه‌ها، مضامین، و روندهای اصلی تحقیق در این دو حوزه و همچنین شکاف‌ها و فرصت‌های تحقیقات آتی را نشان می‌دهند. به‌طور ویژه بر الگوهای هم‌رخداد و شبکه بین واژگان در مقالات این زمینه بین‌رشته‌ای تمرکز شد. چنین تحلیلی نه تنها به درک چشم‌انداز

1. Gunning et al.
2. Hepenstal and McNeish
3. Sokol & Flach
4. Armstrong
5. black-box
6. Rudin
7. Das & Rad
8. Aizenberg & Hoven
9. Floridi et al.
10. Jobin et al.

فکری کمک می‌کند، بلکه بینش‌هایی را در مورد حوزه‌های پژوهش، الگوریتم‌ها، و روش‌های تأثیرگذار و مسیر موضوعات تحقیقاتی ارائه می‌دهد.

۲. مبانی نظری و پیشینه‌ی پژوهش

هوش مصنوعی از فناوری‌های نوظهور و تأثیرگذار در دنیای امروز است که در بسیاری از زمینه‌ها و حوزه‌ها خصوصا مدیریت منابع انسانی کاربرد دارد. مدیریت منابع انسانی شامل عملکردها و فعالیت‌های متنوعی است و هوش مصنوعی می‌تواند به بهبود و بهینه‌سازی این عملکردها و فعالیت‌ها کمک کند. پتانسیل‌های هوش مصنوعی موجب کاربردهای گسترده‌ای در مدیریت منابع انسانی ازجمله اتوماسیون وظایف تکراری و پیچیده مانند جذب، توسعه، حفظ، ارزیابی، و مدیریت استعداد (جیا و همکاران[۱]، ۲۰۱۸؛ کزیم و همکاران[۲]، ۲۰۲۱؛ میکام[۳]، ۲۰۲۰؛ ساماراسینگه و مدیس[۴]، ۲۰۲۰؛ واردارلیر و زفر، ۲۰۲۰)، فراهم کردن منابع و روش‌های جدید برای جمع‌آوری، تجزیه‌وتحلیل و تصمیم‌گیری در زمینهٔ مدیریت منابع انسانی با استفاده از داده‌های بزرگ (آلبرت[۵]، ۲۰۱۹)، افزایش کارایی، دقت، نوآوری، و کیفیت در فرایندهای مدیریت منابع انسانی با استفاده از الگوریتم‌های چابک، خودآموز، خودبهبود، و خودکارساز (شولز، ۲۰۱۹؛ اعتمادی و همکاران، ۲۰۲۴)، بهبود رابطه با کارمندان با استفاده از ربات‌ها، دستیاران صوتی، و شبکه‌های اجتماعی (شان موگام و گارگ[۶]، ۲۰۱۵؛ وتو و همکاران[۷]، ۲۰۲۱)، بهبود آموزش و توسعهٔ کارکنان با استفاده از سامانهٔ مدیریت یادگیری، یادگیری آنلاین، یادگیری تطبیقی، و یادگیری نظیر به نظیر شده است.

اما استفاده از هوش مصنوعی چالش‌ها و ریسک‌های فنی (لان[۸]، ۲۰۲۰)، اخلاقی (داونپورت و همکاران[۹]، ۲۰۱۹)، و حقوقی و اجتماعی هانگ و راست[۱۰]، ۲۰۱۸) را به همراه دارد. یکی از چالش‌های کار با سیستم‌های هوش مصنوعی این است که نتایج آن‌ها توضیح‌پذیر یا قابل پیگیری نیست (ویودی و همکاران[۱۱]، ۲۰۲۱)، به‌ویژه آن دسته از مدل‌های هوش مصنوعی مانند شبکه‌های

1. Jia et al.
2. Kazim et al.
3. Meacham
4. Samarasinghe & Medis
5. Albert
6. Shanmugam & Garg
7. Votto et al.
8. Lohn
9. Davenport et al.
10. Huang & Rust
11. Dwivedi et al.

عصبی مصنوعی (ANN) برگرفته از الگوی اتصال نورون‌ها (راسل و نورویگ[1]، ۲۰۱۰) که الگوی مغز انسان را دنبال می‌کنند به‌عنوان یک جعبه سیاه توصیف شده‌اند (اددی و برادا[2]، ۲۰۱۸). جعبه سیاه به‌معنای یک سیستم یا مدل هوش مصنوعی است که خروجی‌های آن برای کاربران و ذی‌نفعان قابل فهم منطقی نباشد (رودین، ۲۰۱۹). با افزایش پیچیدگی الگوریتم، درک فرایندهای چنین جعبه‌های سیاهی سخت‌تر می‌شود. الگوریتم‌ها، نهایتاً منجر به شناسایی الگوهای زیربنایی در یک مجموعه داده می‌شوند؛ لذا نمی‌توانند نتایج متعصبانه را شناسایی کنند و قادر به ارزیابی اخلاقی نیستند (باردو آریتا[3]، ۲۰۲۰).

برخی از افراد به تصمیمات و پیش‌بینی‌های هوش مصنوعی علاقه نشان نمی‌دهند و تصمیم نهایی استخدام را به انسان واگذار می‌کنند (برگر و همکاران[4]، ۲۰۲۱؛ دیورست و همکاران[5]، ۲۰۱۵؛ جوسوپو و همکاران[6]، ۲۰۲۰؛ اوکمن و همکاران[7]، ۲۰۱۱) نامفهوم بودن توصیه‌های هوش مصنوعی (اددی و برادا، ۲۰۱۸)، باعث بی اعتمادی و مقابله با تصمیم درست آن می‌شود (برت[8]، ۲۰۱۸). نیاز به سیستم‌های هوش مصنوعی با توضیحات دلایل تصمیمات آن برای تصمیم‌گیرنده‌ها بیش از پیش احساس می‌شود تا اعتماد به تصمیم این سیستم‌ها را افزایش دهد (آیزنبرگ و هون، ۲۰۲۰). از آنجاکه جعبه سیاه هوش مصنوعی ممکن است باعث تصمیمات متعصبانه، نادرست، یا نامناسب شود (اددی و برادا، ۲۰۱۸) در نتیجه منجر به کاستی در پذیرش و رضایت کارکنان از تصمیمات آن در مدیریت منابع انسانی می‌شود.

هوش مصنوعی توضیح‌پذیر به دنبال افزایش قابل‌فهم بودن، دارای‌منطق‌بودن، و قابل‌اعتمادبودن هوش مصنوعی برای کاربران و ذی‌نفعان آن است (هولستین و همکاران[9]، ۲۰۱۸). هوش مصنوعی توضیح‌پذیر منابع انسانی سعی می‌کند توضیحات شفاف، خلاقانه، تأثیرپذیر، تأثیرگذار و قابل درک برای عملکرد و نتایج هوش مصنوعی ارائه کند (داس و راد[10]، ۲۰۲۰؛ سانتونی د سیو و

1. Russell & Norvig
2. Adadi & Berrada
3. Barredo Arrieta
4. Berger et al.
5. Dietvorst et al.
6. Jussupow et al.
7. Ochmann et al.
8. Baert
9. Holstein et al.
10. Das & Rad

هوون[1]، ۲۰۱۹). هدف این توضیحات افزایش اطمینان، امنیت، عدالت، و اخلاق در سامانه‌های هوش مصنوعی منابع انسانی است (فلوریدی و کولز[2]، ۲۰۱۹).

پژوهش‌های کتاب‌سنجی زیرروندها، الگوها، خلاقیت، تأثیرپذیری، و تأثیرگذاری در این زمینه را شناسایی کردند و کاربردها، چالش‌ها، و راهکارهای این زمینه را در حوزه‌های مختلف نظیر مالی، سلامت، پزشکی، تولید، حمل‌ونقل، و زنجیره تأمین بررسی کردند:

آلونسو و همکاران[3] (۲۰۱۸) با استفاده از پایگاه اسکوپوس[4] و ابزار وی او اس ویور[5]، یک تحلیل کتاب‌سنجی از تحقیقات هوش مصنوعی توضیح‌پذیر در دوره ۲۰۰۰ تا ۲۰۱۷ ارائه دادند. آن‌ها نقش جامعهٔ فازی در پیشبرد هوش مصنوعی توضیح‌پذیر را بررسی کردند.

چن و همکاران (۲۰۲۳) با استفاده از پایگاه وب‌آوساینس و ابزار سایت اسپیس[6]، یک تحلیل کتاب‌سنجی از تحقیقات هوش مصنوعی توضیح‌پذیر در حوزهٔ مالی در دورهٔ ۲۰۱۳ تا ۲۰۲۳ منتشر کردند. آن‌ها چشم‌انداز فعلی و آیندهٔ هوش مصنوعی توضیح‌پذیر در این حوزه را نشان دادند و برخی از کاربردهای آن را معرفی کردند.

عبدالکریم (۲۰۲۱) بررسی منابع از کاربردهای یادگیری ماشین در سیستم‌های اطلاعاتی (IS) و یک تحلیل کتاب‌سنجی از ادبیات هوش مصنوعی توضیح‌پذیر از سال ۲۰۰۰ تا ۲۰۱۹ انجام داد. آن‌ها مسائل باز در این زمینه را شناسایی کردند و گفتند که روش‌های یادگیری ماشین در مجلات سیستم‌های اطلاعاتی با سطح بالا کمتر به چالش کشیده شده‌اند. آن‌ها همچنین چارچوب مفهومی برای هوش مصنوعی توضیح‌پذیر با استفاده از چهار بعد نوع داده، نوع توضیح، نوع ارزیابی، و نوع دامنه پیشنهاد کردند.

ولان و لانگ (۲۰۲۰) با استفاده از روش مرور نظام‌مند، وضعیت فعلی و زمینه‌های آیندهٔ هوش مصنوعی توضیح‌پذیر را بیان کردند. آن‌ها روش‌های هوش مصنوعی توضیح‌پذیر را به‌عنوان ابزار تحقیق و نه به‌عنوان اهداف مطالعه یا کاربرد در نظر گرفتند. آن‌ها همچنین دربارهٔ پیامدهای اخلاقی هوش مصنوعی توضیح‌پذیر برای هوش مصنوعی مسئول را بحث کردند و تعدادی دستورالعمل‌ها و اصول را برای توسعه و استقرار سیستم‌های هوش مصنوعی توضیح‌پذیر ارائه کردند.

1. Santoni de Sio & Hoven
2. Floridi & Cowls
3. Alonso et al.
4. Scopus
5. VOSviewer
6. Citespase

منیر (۲۰۱۹) با استفاده از پایگاه وب‌آوساینس و پایگاه اسکوپوس، یک تحلیل کتاب‌سنجی از کاربردهای یادگیری عمیق در تشخیص سرطان انجام داد. او نمونه‌هایی از کاربردهای جدید و مشهور یادگیری عمیق در این زمینه را معرفی کرد.

دهمینجا و بگ (۲۰۲۰) با استفاده از پایگاه اسکوپوس و ابزار وی او اس ویور، با روش کتاب‌سنجی نقش هوش مصنوعی در محیط عملیات را بررسی کردند. آن‌ها تکنیک‌های هوش مصنوعی به‌طورکلی، نه فقط یادگیری عمیق، در مدیریت عملیات شامل حوزه‌ها و بخش‌های مختلف مرتبط و محیط عملیات، مانند تولید، حمل‌ونقل، زنجیره تأمین، و کیفیت را بررسی کردند.

تران و همکاران (۲۰۱۹) با استفاده از پایگاه وب‌آوساینس و ابزار سایت اسپیس، یک تحلیل منابع جهانی از تحقیقات هوش مصنوعی در سلامت و پزشکی ارائه دادند. آن‌ها نشان دادند که پژوهش در هوش مصنوعی در سلامت و پزشکی به‌ویژه در زمینهٔ پیش‌بینی و درمان بالینی به‌طور قابل‌توجهی در سال‌های اخیر افزایش‌یافته است.

مطالعات هوش مصنوعی توضیح‌پذیر، بیشتر به مفاهیم اصلی آن و ملاحظات هستی‌شناسانه توضیح‌پذیری (گونینگ و همکاران، ۲۰۱۹) پرداخته است که تحقیقات بر اصول اساسی طراحی و توسعه هوش مصنوعی توضیح‌پذیر متمرکز است (باردِو آریتا و همکاران، ۲۰۲۰). مفاهیمی چون اخلاق، عدالت، شفافیت، قابل‌تفسیر بودن، پاسخ‌گویی، و قابل‌اعتماد بودن (جوبین و همکاران، ۲۰۱۹) به‌عنوان پایه‌های اصلی در نظر گرفته شده‌اند. این مفاهیم برای اطمینان از سازگاری سیستم‌های هوش مصنوعی توضیح‌پذیر با ارزش‌ها، حقوق، و مسئولیت‌های انسانی بررسی شده‌اند (دوشی ولز و کیم[1]، ۲۰۱۷).

دیگر مطالعات به مفاهیم و کاربردهای یادگیری ماشین و هوش مصنوعی در حوزه‌های مختلف پرداختند . از اعتبارسنجی و یادگیری ماشین قابل‌تفسیر گرفته تا کاربردهای آن در پزشکی دقیق و بیوانفورماتیک (دینگ و همکاران[2]، ۲۰۱۳)، که این تحقیقات نشان‌دهندهٔ اهمیت فزاینده هوش مصنوعی و یادگیری ماشین در زمینه‌های پزشکی و مراقبت‌های بهداشتی است (هولزینگر و همکاران[3]، ۲۰۱۷).

در حوزهٔ دیگر هوش مصنوعی توضیح‌پذیر بر یادگیری عمیق در پردازش تصویر، گفتار، و سیگنال که به بررسی کاربردهای یادگیری عمیق در پردازش تصویر، گفتار و سیگنال می‌پردازد (گولوم و

1. Doshi-Velez & Kim
2. Ding et al.
3. Holzinger et al.

همکاران، ۲۰۲۱؛ لیژنس و همکاران[1]، ۲۰۱۷) تمرکز شده است. با استفاده از شبکه‌های عصبی کانولوشنال[2] (CNN)، شبکه‌های عصبی بازگشتی[3] (RNN) و شبکه‌های مولد رقیب[4] (GAN) (چنگ و همکاران، ۲۰۲۳)، مدل‌هایی برای تشخیص و تفسیر تصاویر پزشکی، تولید و ترجمه متن و تولید تصاویر واقع‌گرایانه و خلاقانه ایجاد شده‌اند (گودفلو و همکاران[5]، ۲۰۱۴).

در تحقیقات هوش مصنوعی توضیح‌پذیر به روش‌ها و فن‌های علم داده که به بررسی روش‌ها و فن‌های علم داده می‌پردازد و شامل جمع‌آوری، پردازش، تجزیه‌وتحلیل و تفسیر داده‌ها توجه داشته است.تحقیقات در این حوزه به مدل‌های پیش‌بینی، استخراج ویژگی، تجسم، و داده‌کاوی (آگاروال[6]، ۲۰۱۵؛ پرووست فاوت[7]، ۲۰۱۳) می‌پردازند.

دیگر مطالعات هوش مصنوعی توضیح‌پذیر فنون آموزش در مدل‌سازی پیش‌بینی و توضیح‌پذیری که فنون ساخت و یاددهی برای پیش‌بینی و بهینه‌سازی الگوریتم‌های یادگیری را مورد بررسی می‌کنند (هستیه و همکاران[8]، ۲۰۰۹)، همچنین، ارزیابی و تفسیر مدل‌ها با استفاده از معیارها و فن‌های توضیح‌پذیری می‌پردازند (واچر و همکاران[9]، ۲۰۱۷).

حوزهٔ دیگر مطالعات هوش مصنوعی توضیح‌پذیر یادگیری ماشین قابل‌تفسیر به کاربردهای یادگیری ماشین قابل‌تفسیر در علوم داده می‌پردازد (ویتن[10]، ۲۰۱۶). فن‌های یادگیری ماشین قابل تفسیر مانند انتخاب ویژگی، و درخت تصمیم می‌توانند مدل‌های قابل‌تفسیر تولید کنند یا قوانین را از مدل‌های جعبه سیاه (رودین، ۲۰۱۹) استخراج کنند.

همچنین پژوهش‌گران هوش مصنوعی توضیح‌پذیر پژوهش بر حوزه‌ها و فن‌های پیشرفته در هوش مصنوعی و مدل‌سازی محاسباتی که روش‌ها و فن‌های پیشرفته در هوش مصنوعی و مدل‌سازی محاسباتی را به کار گرفته‌اند را بررسی کرده‌اند و می‌توانند عملکرد و تصمیم‌گیری خود را با استفاده از روش‌های یادگیری تقویتی (سوتن و بارتو[11]، ۲۰۱۸)، سری زمانی (باکس و

1. Litjens et al.
2. Convolutional neural network
3. Recurrent neural network
4. Generative adversarial network
5. Goodfellow et al.
6. Aggarwal
7. Provost & Fawcett
8. Hastie et al.
9. Wachter et al.
10. Witten
11. Sutton & Barto

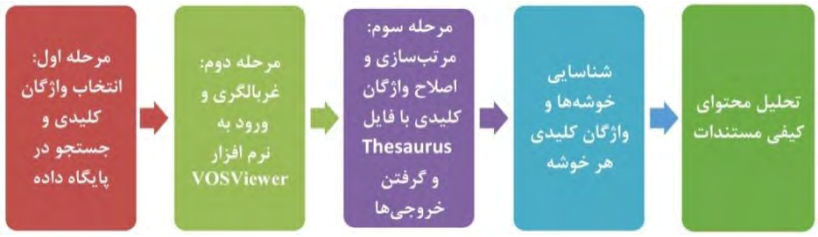
همکاران[1]، ۲۰۱۵)، و یادگیری عمیق (گودفل، ۲۰۱۶) توضیح دهند.

بررسی پژوهش‌های موجود نشان می‌دهد که تاکنون پژوهشی با تکنیک تحلیل هم‌رخدادی بر روی هوش مصنوعی توضیح‌پذیر در پایگاه وب‌آوساینس انجام نشده است. پژوهش‌های پیشین یا از روش علم‌سنجی استفاده نکرده‌اند و یا در صورت استفاده از این روش‌شناسی، سایر تکنیک‌ها همچون تحلیل هم‌استنادی را به کار گرفته‌اند. همچنین، آن دسته از مطالعات که به توضیح‌پذیری هوش مصنوعی پرداخته‌اند عموماً از دریچه فنی و مهندسی نه از دریچهٔ علوم انسانی و به‌طور خاص مدیریت منابع انسانی به مسئله نگاه کرده‌اند؛ لذا تمرکز موضوعی مطالعات پیشین، متفاوت بوده و بازهٔ زمانی متفاوتی را موردتوجه قرار داده‌اند. ازاین‌رو، مطالعه‌ای برای فهم روند مطالعاتی سال‌های اخیر در حوزهٔ هوش مصنوعی توضیح‌پذیر و کاربست آن در مدیریت منابع انسانی با تکنیک‌های علم‌سنجی و همچنین تحلیل محتوا انجام نشده است.

۳. روش‌شناسی پژوهش

این تحقیق از روش علم‌سنجی تحلیل هم‌رخدادی برای مطالعهٔ موضوع هوش مصنوعی توضیح‌پذیر (و ارتباط آن با مدیریت منابع انسانی استفاده کرد. علم‌سنجی کاربرد تکنیک‌های آماری و ریاضی برای تحلیل متون و اسناد است (بوزیدولسکی[2]، ۲۰۱۵). تحلیل هم‌رخدادی نوعی تحلیل محتوا است که می‌تواند الگوها، موضوعات، و گرایش‌ها را در یک رشتهٔ تحصیلی آشکار کند (احمدی و عصاره، ۲۰۱۷). این پژوهش، از نظر ابزار گردآوری داده‌ها نیز، یک پژوهش آمیخته (کمی-کیفی) است، چرا که پس از تحلیل کمّی داده‌ها با استفاده از نرم‌افزار ویواس‌ویوور[3]، به تحلیل محتوای کیفی (واژگان) مستندات پرداخته و سپس، خوشه‌های واژگانی را بحث و بررسی کرده است.

در این مطالعه، از چارچوب اجرایی زیر برای انجام پژوهش استفاده شده است.



شکل ۱. مراحل انجام پژوهش

1. Box et al.
2. Buzydlowski
3. VOSViewer

برای شناسایی مقالات مربوط به موضوع هوش مصنوعی توضیح‌پذیر، از کلیدواژه‌هایی که با الگوهای توضیح‌پذیری مطابقت داشتند، استفاده شد. علاوه‌براین، از عبارات هوش مصنوعی به‌عنوان کلیدواژه‌های مرتبط با موضوع اصلی پژوهش نیز استفاده شد. محدودهٔ زمانی جستجو از تاریخ ۱ ژانویه ۲۰۱۳ تا ۲۰ اکتبر ۲۰۲۳ تعیین شد و نوع سند مورد نظر مقالهٔ علمی بود. با اعمال این معیارها، تعداد ۳۲۳۸ مقاله در پایگاه وب‌آوساینس یافت شد. پس از استخراج مقالات تعداد ۱۱۴۳۱ کلمه در نرم‌افزار ویواس‌ویوور تحلیل شد. برای تهیهٔ فایل مترادف‌ها، واژگانی که حداقل ۵ بار در مقالات رخ داده‌اند، انتخاب شدند. تعداد این واژگان ۷۳۹ واژه بود. سپس، با استفاده از ثزاروس، واژگان مترادف و مشابه به یکدیگر تبدیل شدند. در نتیجه، تعداد کلمات به ۶۱۵ کلمه کاهش یافت. برای تحلیل هم‌رخدادی واژگان، حداقل تعداد رخداد واژگان را ۱۰ رخداد انتخاب کردیم. در این صورت، ۲۶۹ واژه حائز این شرط انتخاب شدند. در تنظیمات آنالایسیس، شروع ۱۰۰، کاهش اندازهٔ گام را ۰٫۲۵ و دانهٔ تصادفی را ۵۰ گذاشتیم. برای خوشه‌بندی واژگان، شروع ۱۰۰، تکرار ۱۰۰۰ و دانه تصادفی را ۳۰ به‌عنوان مقدارهای پارامترها تنظیم کردیم. پس از ترسیم نقشه دانش، به ۸ خوشه رسیدیم.

فرمول ۱: کوئری جستجوی XAIS

Explainabl* OR understandabl* OR comprehensibl* OR interpretabl* (Topic) AND "AI" OR "artificial intelligence"OR "artificial-intelligence" OR "XAI" (Topic) AND 2013-01-01/2023-10-20 (Publication Date) AND Article (Document Type)

برای بررسی رابطه بین هوش مصنوعی توضیح‌پذیر و منابع انسانی، کلیدواژه «human resource*» را به کوئری قبلی اضافه کردیم. با این کار، تعداد ۸ مقاله یافت شد. با افزودن قیدهای «employee، personnel» و «recruitment» به جستجو (کوئری ۲)، تعداد ۴۹ سند یافت شد. از این پس به این حوزهٔ تحقیقاتی با عنوان XHRS اشاره می‌کنیم. با ورود این اسناد به نرم‌افزار ویواس‌ویوور، ۳۲۷ کلمه کلیدی یافت شد که بعد از اعمال ثزاروس به ۳۰۹ کلمه رسیدیم. با گذاشتن هم‌رخدادی روی ۲ کلمه، به ۴۶ کلمه کلیدی رسیدیم.

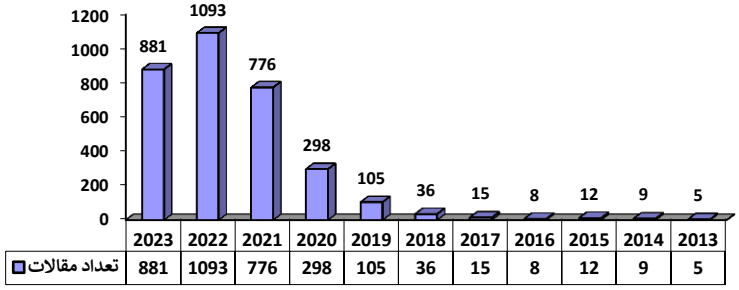
فرمول ۲: کوئری جستجوی XHRS

Explainabl* OR understandabl* OR comprehensibl* OR interpretabl* (Topic) AND "AI" OR "artificial intelligence" OR "artificial-intelligence" OR "XAI" (Topic) AND Article (Document Type) AND 2013-01-01/2023-10-20 (Publication Date) AND "human resource*" OR employee OR personnel OR recruitment (All Fields)

۴. یافته‌های پژوهش

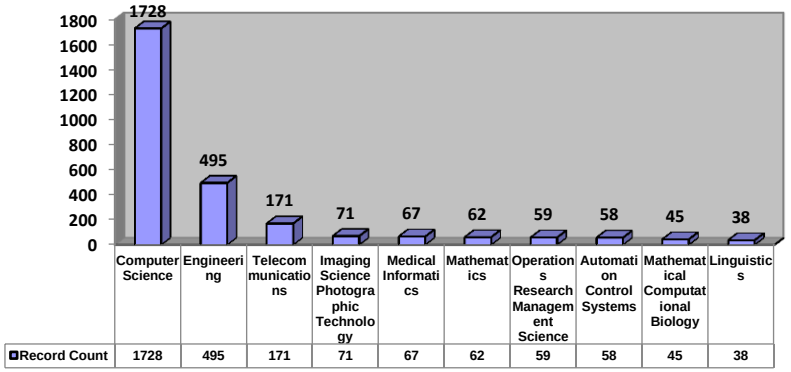
۴-۱. نقشه دانش هوش مصنوعی توضیح‌پذیر

سال انتشار مستندات منتخب (شکل ۲) نشان می‌دهدکه انتشار این مستندات از سال ۲۰۱۸ تا ۲۰۲۳ روند صعودی داشته که این امر، نشان‌دهندهٔ اهمیت موضوع قابلیت توضیح و تمرکز بیشتر بر روی آن در ادبیات هوش مصنوعی دارد. به‌نظر می‌رسد ظهور پاندمی کووید ۱۹ در سال‌های اخیر و افزایش نیازها به استفاده از هوش مصنوعی توضیح‌پذیر در تحقیقات درمانی کووید۱۹، یکی از دلایل افزایش تمرکز بر روی این موضوع باشد.



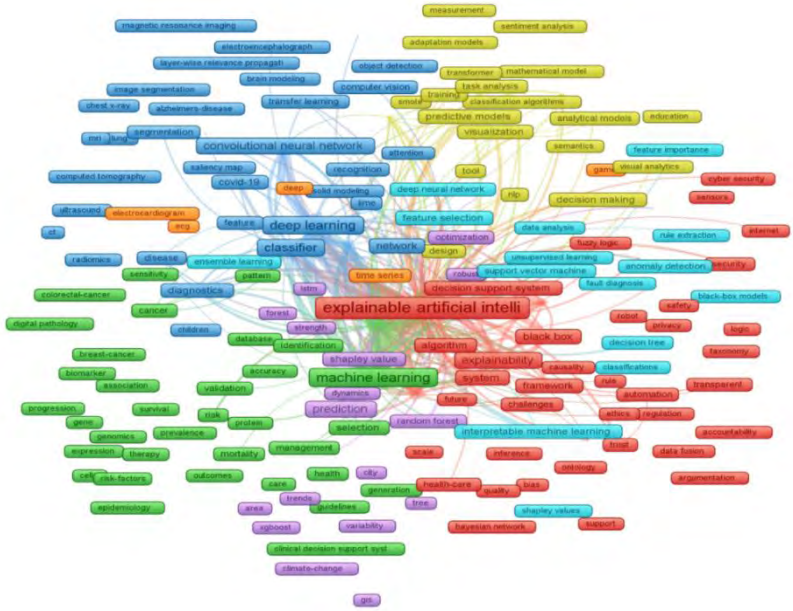
شکل ۲. سال انتشار مستندات منتخب

در طبقه‌بندی موضوعی پایگاه وب‌آو ساینس، رایج‌ترین دسته‌های موضوعی مستندات منتخب به شرح شکل ۳ قابل‌مشاهده هستند. برخی مستندات، هم‌زمان در چند دسته‌بندی موضوعی قرار گرفته‌اند. بالاترین رتبه مربوط به علوم کامپیوتر با ۱۷۲۸ پژوهش است. مهندسی با ۴۹۵ پژوهش با اختلاف بسیار زیادی از علوم کامپیوتر رتبهٔ دوم را کسب کرده است. مدیریت با ۵۹ مقاله جایگاه هفتم را کسب کرده است.



شکل ۳. رایج‌ترین دسته‌های موضوعی مستندات منتخب در پایگاه وب‌آو ساینس

خروجی تحلیل هم رخدادی هوش مصنوعی توضیح‌پذیر با استفاده از نرم‌افزار ویواس‌ویوور در شکل ۴ قابل‌مشاهده است.



شکل ٤. نقشهٔ علمی تحلیل هم‌رخدادی مطالعات حوزه هوش مصنوعی توضیح‌پذیر

تحلیل هم‌رخدادی، همان‌گونه که در نقشهٔ دانش هوش مصنوعی توضیح‌پذیر دیده می‌شود، وجود ۸ خوشه را به شرح جداول ۱، ۲ و ۳ نشان می‌دهد. در این جداول، رخداد، میزان وجود واژگان کلیدی در پژوهش‌ها نشان داده شده است. مجموع قدرت اتصال نیز، تعداد ارتباط هم‌رخدادی هر واژه کلیدی را با سایر واژه‌ها نشان می‌دهد.

در جدول ۱ واژگان سه خوشهٔ قرمز، سبز و آبی قابل‌مشاهده است. اولین و بزرگ‌ترین خوشه (قرمز) به «مفاهیم اصلی هوش مصنوعی توضیح‌پذیر و ملاحظات هستی‌شناسانه توضیح‌پذیری» می‌پردازد و شامل متغیرهای بنیادینی مثل «اخلاق»، «عدالت»، و «قابل‌اعتمادبودن» است که عامل‌های هوش مصنوعی توضیح‌پذیر بر مبنای آن‌ها طراحی، ارزیابی و نظارت می‌شود، به‌طوری‌که با ارزش‌ها، حقوق، و مسئولیت‌های انسانی سازگار باشند. کلمات شفافیت، قابل‌تفسیر بودن، پاسخ‌گویی، عدالت، و قابل‌اعتماد بودن (باردِو آریتا و همکاران، ۲۰۲۰ ؛ دوشی ولز و کیم، ۲۰۱۷؛ جوبین و همکاران، ۲۰۱۹)، اشاره به اصول اساسی طراحی و توسعه هوش مصنوعی توضیح‌پذیر و کلمات انتخاب ویژگی، انتساب ویژگی، بصری‌سازی مدل، توضیح مدل، ارزیابی مدل، و بازخورد

کاربر (اددی و بردا، ۲۰۱۸؛ ریبیرو و همکاران[1]، ۲۰۱۶)، روش‌های ایجاد و ارزیابی هوش مصنوعی توضیح‌پذیر دارد. تولید بینش و دانش از هوش مصنوعی توضیح‌پذیر از رهگذر تحلیل، فرایند جمع‌آوری، پردازش، و تفسیر داده‌ها می‌گذرد که واژگان اکتشاف، کاوش، پیش‌پردازش، ترکیب، و ارزیابی کیفیت داده، به‌ترتیب، به مراحل آن اشاره دارند. همچنین اخلاق از مهمترین مباحث زمینه‌ای و هستی‌شناسانه در تصمیم‌گیری است. کلمات تجزیه‌وتحلیل ذی‌نفعان، شناسایی معضلات اخلاقی، انتخاب چارچوبی اخلاقی، استدلال، و توجیه اخلاقی (میلستات[2]، ۲۰۱۶)، به ملاحظات اخلاقی تصمیمات سامانه‌های هوش مصنوعی توضیح‌پذیر اشاره دارد.

خوشهٔ دوم (سبز)، «مفاهیم و کاربردهای یادگیری ماشین و هوش مصنوعی» در حوزه‌های مختلف است که کلمات اعتبارسنجی، یادگیری ماشین قابل‌تفسیر، شبکه عصبی گرافی، پزشکی دقیق، و بیوانفورماتیک، نشان‌دهندهٔ اهمیت در زمینهٔ پزشکی و مراقبت‌های بهداشتی، و کلمات تولید، کشف، بیان، کشف دارو، آسیب‌شناسی دیجیتال، سرطان، سلامت، مرگ‌ومیر، بقا و افسردگی نشان‌دهندهٔ کاربرد در ژنتیک، پزشکی، توسعهٔ اجتماعی، و سلامت روان هستند (چن و همکاران، ۲۰۱۸؛ دینگ و همکاران، ۲۰۱۳) و کلمات خطر، مدیریت، تصمیم، راهنماها، و دخالت نشان‌دهندهٔ کاربرد برای پشتیبانی و بهبود مدیریت راهبردی در کسب‌وکارها، دولت‌ها، و سازمان‌های اجتماعی هستند (کوکبرن و همکاران[3]، ۲۰۱۸).

خوشهٔ سوم (آبی) مشخص شده، «مفاهیم، روش‌ها، و کاربردهای یادگیری عمیق در پردازش تصویر، گفتار، و سیگنال» است که با شبکه‌های عصبی کانولوشنال، تصاویر پزشکی را دسته‌بندی می‌کنند، تشخیص می‌دهد، و تفسیر می‌کنند و با شبکه‌های عصبی بازگشتی، مدل‌های زبانی را برای تولید و ترجمه متن ایجاد و با شبکه‌های مولد رقیب، تصاویر واقع‌گرایانه و خلاقانه تولید می‌کنند. در این خوشه کلمات یادگیری عمیق، شبکهٔ عصبی، شبکه عصبی کانولوشنال، توجه، یادگیری انتقالی به جنبه‌های کلیدی یادگیری عمیق و کلمات مدل، بهینه‌سازی، خطا، معیار، آزمون به توسعه و بهینه‌سازی مدل اشاره دارند. کاربردهای یادگیری عمیق در پردازش سیگنال شامل کلمات دسته‌بندی، تشخیص، تفسیر، الکتروانسفالوگرافی[4] (کلیفورد و همکاران[5]، ۲۰۱۷؛ لیژنس و همکاران[6]، ۲۰۱۷؛

1. Ribeiro et al.
2. Mittelstadt et al.
3. Cockburn et al.
4. Electroencephalogram (EEG)
5. Clifford et al.
6. Litjens et al.

شیرمیسر و همکاران[1]، ۲۰۱۷) و کاربردهای یادگیری عمیق در پردازش تصویر کلمات تشخیص، تفسیر، تقطیع، تصویربرداری پزشکی، شبکه‌های مولد رقیب و کاربردهای یادگیری عمیق در پردازش زبان شامل کلمات مدل زبانی، ترجمه، تولید متن، تشخیص گفتار، تحلیل احساسات (دولین و همکاران[2]، ۲۰۱۹؛ هینتون و سالاخودینوس[3]، ۲۰۰۶) هستند.

جدول ۱. فراوانی واژگان کلیدی هوش مصنوعی توضیح‌پذیر و قدرت اتصال آن‌ها در سه خوشه قرمز و سبز و آبی

خوشه اول قرمز	مجموع قدرت اتصال	رخداد	خوشه دوم سبز	مجموع قدرت اتصال	رخداد	خوشه سوم آبی	مجموع قدرت اتصال	رخداد
XAI	5653	1662	machine learning	2925	727	deep learning	2323	565
model	1447	330	artificial intelligence	2852	722	classifier	1607	348
explainability	992	233	selection	320	65	neural network	1255	260
interpretable	864	187	risk	251	61	CNN	870	199
system	662	154	cancer	276	57	network	487	119
black box	750	134	management	238	51	diagnostics	584	118
algorithm	570	115	validation	165	45	covid-19	403	101
framework	460	90	identification	132	36	lime	236	56
decision support system	414	79	explainable machine learning	126	35	segmentation	283	56
knowledge	240	61	mortality	138	35	disease	266	54
trust	266	53	prognosis	135	30	transfer learning	200	49
big data	275	52	health	122	29	computer vision	136	32
behavior	170	40	pattern	120	28	recognition	152	32
IOT	153	40	association	124	26	feature	147	30
transparent	216	39	biomarker	120	26	EEG	111	28
information	170	37	accuracy	121	24	attention	136	27
automation	171	35	survival	100	23	fusion	124	25
uncertainty	176	33	breast-cancer	95	22	image classification	93	25
challenges	181	32	care	95	22	ct	96	22
reliability	194	32	discovery	76	22	mri	90	22
representation learning	132	32	generation	101	22	alzheimers-disease	120	20
security	162	32	outcomes	78	22	computer-aided diagnosis	107	20

1. Schirrmeister et al.
2. Devlin et al.
3. Hinton & Salakhutdinov

خوشه اول قرمز	مجموع قدرت اتصال	رخداد	خوشه دوم سبز	مجموع قدرت اتصال	رخداد	خوشه سوم آبی	مجموع قدرت اتصال	رخداد
health-care	190	31	prevalence	87	22	imaging	101	20
fuzzy logic	107	28	sensitivity	113	20	grad-cam	90	19
cyber security	134	25	time	66	20	object detection	100	19
future	133	25	clinical decision support system	81	19	children	58	17
inference	111	25	database	100	19	chest x-ray	82	16
internet	126	25	graph neural network	69	19	visual explanation	61	16
ontology	70	23	depression	81	18	medical diagnostic imaging	109	14
ethics	97	22	risk-factors	69	18	computed tomography	91	13
HCI	111	22	gene-expression	76	17	dementia	67	13
rule	135	22	genomics	83	16	image processing	57	13
ayesian network	71	20	medicine	93	16	mild cognitive impairment	76	13
intelligence	66	20	precision medicine	65	15	pneumonia	73	13
privacy	130	20	therapy	62	15	radiomics	72	13
quality	84	19	gene	62	14	saliency map	48	13
accountability	108	18	age	52	13	brain modeling	84	12
industry 4.0	75	18	guidelines	69	13	image segmentation	72	12
support	80	18	bioinformatics	82	12	MRI	67	12
recommender system	58	17	cells	36	12	brain	62	11
intrusion detection	74	16	digital pathology	34	12	Generative adversarial networks	55	11
logic	53	16	epidemiology	50	12	ultrasound	56	11
technology	75	16	expression	46	12	layer-wise relevance propagation	51	10
metaanalysis	59	15	growth	55	12	lesions	66	10
robot	77	15	progression	69	12	lung	64	10
safety	83	15	colorectal-cancer	29	11	medical imaging	44	10
taxonomy	77	15	drug discovery	47	11	solid modeling	70	10
analytics	63	14	score	44	11			
autonomous vehicles	67	13	electronic health records	40	10			
fairness	64	13	intervention	39	10			
adversarial attacks	49	12	personalized medicine	51	10			

خوشه اول قرمز	رخداد	مجموع قدرت اتصال	خوشه دوم سبز	رخداد	مجموع قدرت اتصال	خوشه سوم آبی	رخداد	مجموع قدرت اتصال
data fusion	12	70	protein	10	49			
bias	11	48	resistance	10	42			
case-based reasoning	11	65						
causality	11	56						
scale	11	50						
sensors	11	63						
argumentation	10	28						
human-ai interaction	10	32						
ontologies	10	47						
regulation	10	64						
standards	10	47						

خوشهٔ چهارم (زرد) شامل کلمات روش‌ها و فن‌های علم داده، تجزیه‌وتحلیل و حل مسائل محاسباتی و کاربردهای آن در حوزه‌های مختلف است. این خوشه شامل سه بخش است.

بخش اول: شیوهٔ جمع‌آوری، پردازش، تجزیه‌وتحلیل، و تفسیر داده‌ها. کلمات مدل‌های پیش‌بینی، استخراج ویژگی، تجسم، داده‌کاوی، مدل‌های تحلیلی، پیش‌بینی، مدل‌های ریاضی، و الگوریتم‌های طبقه‌بندیکه به علم داده اشاره دارد.

بخش دوم: فرایندها و مراحل تجزیه‌وتحلیل و حل مسائل محاسباتی که شامل شناسایی، فرمول‌بندی، پیاده‌سازی و ارزیابی راه‌حل‌های مسائل مبتنی بر داده و واژه‌هایی مانند تصمیم‌گیری، تحلیل کار، طراحی، آموزش، نظارت، و مدل‌های انطباق (انت[1]، ۲۰۰۳؛ کراس[2]، ۲۰۱۱؛ هالند[3]، ۱۹۹۲؛ مایکل[4]، ۱۹۹۷؛ سایمون[5]، ۱۹۷۹) است.

بخش سوم: کاربردهای حل مسئلهٔ محاسباتی در حوزه‌های مختلف مثل پردازش زبان طبیعی، علوم اعصاب، تحلیل شبکه‌های اجتماعی، یادگیری نامتعادل، نظریهٔ اندازه‌گیری، روش علمی، افزایش داده‌ها، آموزش و منطق فازی. این بخش شامل کلمات نورون‌ها، رسانه‌های اجتماعی، اندازه‌گیری، علم، تقویت داده‌ها، آموزش، و سامانه‌های فازی تجزیه‌وتحلیل احساسات،

1. Annett
2. Cross
3. Holland
4. Mitchell
5. Simon

ترانسفورماتور، سازوکار توجه، شبکه عصبی مکرر، معناشناسی، شناخت و ادراک است.

خوشهٔ پنجم (بنفش)، فنون آموزش در مدل‌سازی پیش‌بینی، توضیح‌پذیری، و کاربرد در تغییرات آب‌وهوا.

بخش اول: فنون ساخت و یاددهی برای پیش‌بینی و بهینه‌سازی الگوریتم‌های یادگیری تحت نظارت و بدون نظارت شامل کلمات رگرسیون، جنگل تصادفی، شبکهٔ عصبی مصنوعی، برنامه‌ریزی ژنتیکی و انتخاب متغیر است.

بخش دوم: ارزیابی و تفسیر مدل‌ها با استفاده از معیارها و فن‌های توضیح‌پذیری شامل مفاهیمی مانند عملکرد، تأثیر، قوی، پویایی، تغییرپذیری، روندها و قدرت (بیشاپ، ۲۰۱۶) و فنون توضیح‌پذیری شامل ارزش شیپلی[1]، توضیحات افزودنی شیپلی، توضیح خلاف واقع و تحلیل حساسیت است.

بخش سوم: کاربردهای مدل‌ها در تغییرات آب‌وهوا و سنجش‌ازدور. این بخش شامل تحلیل و مدل‌سازی پدیده‌های پیچیده و پویا مانند گرمایش جهانی، تغییر کاربری زمین، شهرنشینی، جنگل‌زدایی، و داده‌های مکانی با استفاده از مفاهیمی مانند تغییر اقلیم، سنجش‌ازدور، شبیه‌سازی، دما، شهر، جنگل، و جی آی اس است.

خوشهٔ ششم (فیروزه‌ای) شامل کلمات روش‌ها و فن‌های یادگیری ماشین قابل‌تفسیر (IML) و کاربردهای آن در علوم داده است. این کلمات مفاهیم و فنون اساسی در یادگیری ماشین و علم داده را نشان می‌دهند که می‌توانند نتایج و تصمیمات قابل‌درک و شفافی را ارائه دهند.

فن‌های یادگیری ماشین قابل‌تفسیر مانند انتخاب ویژگی، درخت تصمیم، استخراج قانون، مقادیر شیپلی، و خلاف واقع می‌توانند مدل‌های قابل‌تفسیر تولید کنند یا قوانین را از مدل‌های جعبه سیاه استخراج کنند.

کاربردهای یادگیری ماشین قابل‌تفسیر مانند مدل پیش‌بینی، تشخیص ناهنجاری، تشخیص خطا، طبقه‌بندی، و کاهش ابعاد می‌توانند پیامدها و ناهنجاری‌های مختلف را با استفاده از مدل‌های یادگیری ماشین، پیش‌بینی کنند و تشخیص دهند.

روش‌ها و فن‌های یادگیری ماشینی مانند شبکهٔ عصبی عمیق، یادگیری مجموعه، ماشین بردار پشتیبان، رمزگذار خودکار و خوشه‌بندی می‌توانند مدل‌های قدرتمند و پیچیدهٔ یادگیری ماشین قابل‌تفسیر را ارائه دهند.

فرایندها و رشته‌های علم داده مانند تجزیه‌وتحلیل داده‌ها، تجسم داده‌ها، یادگیری بدون نظارت، و

1. Shapley value

علم داده می‌توانند داده‌ها را با استفاده از مدل‌ها و فن‌های یادگیری ماشین جمع‌آوری، پردازش، تجزیه‌وتحلیل، و تفسیر کنند.

جدول ۲. فراوانی واژگان کلیدی هوش مصنوعی توضیح‌پذیر و قدرت اتصال آن‌ها در خوشه‌های زرد بنفش و آبی

رخداد	مجموع قدرت اتصال	خوشه ششم آبی کم‌رنگ	رخداد	مجموع قدرت اتصال	خوشه پنجم بنفش	رخداد	مجموع قدرت اتصال	خوشه چهارم زرد
96	380	interpretable machine learning	198	917	prediction	101	678	predictive models
75	367	feature selection	95	421	shapley value	97	585	feature-extraction
55	281	prediction model	77	328	performance	70	372	Visualization
46	212	decision tree	68	290	regression	69	464	data models
45	178	anomaly detection	62	278	random forest	68	355	decision making
44	204	deep neural network	61	254	optimization	61	378	task analysis
35	141	ensemble learning	51	256	impact	57	365	computational modeling
33	204	support vector machine	37	156	genetic programming	56	181	Nlp
28	106	interpretable model	36	149	artificial neural network	54	204	Design
25	134	fault diagnosis	30	104	shapley additive explanations	43	259	Training
25	89	feature importance	28	98	counterfactual explanation	32	217	analytical models
21	107	rule extraction	26	113	index	31	134	data mining
20	73	data analysis	21	129	robust	30	162	Tool
20	59	shapley values	20	60	dynamics	26	94	Transformer
15	64	autoencoder	20	96	simulation	22	136	Cognition
12	45	clustering	19	81	xgboost	22	133	visual analytics
12	45	counterfactuals	17	63	climate-change	21	100	Semantics
12	82	data visualization	17	83	remote sensing	18	81	Forecasting
12	54	unsupervised learning	16	61	trends	18	108	mathematical model
11	59	data science	16	53	variability	16	68	sentiment analysis
10	69	black-box models	15	58	area	15	54	attention mechanism
10	64	Classifications	15	66	lstm	15	64	recurrent neural network
10	43	dimensionality reduction	14	49	strength	14	79	Monitoring
			14	85	tree	13	90	adaptation models
			13	51	climate	12	83	classification algorithms

خوشه چهارم زرد	مجموع قدرت اتصال	رخداد	خوشه پنجم بنفش	مجموع قدرت اتصال	رخداد	خوشه ششم آبی کم‌رنگ	مجموع قدرت اتصال	رخداد
Neurons	70	12	extreme gradient boosting	52	12			
Smote	44	12	long short-term memory	38	11			
social media	45	12	sensitivity-analysis	46	11			
Measurement	58	11	temperature	45	11			
Science	44	11	city	52	10			
Buildings	54	10	forest	46	10			
data augmentation	39	10	gis	43	10			
Education	41	10						
fuzzy systems	39	10						
Perception	32	10						
social networking (online)	49	10						

خوشهٔ هفت (نارنجی)، حوزه‌ها و فن‌های پیشرفته در هوش مصنوعی و مدل‌سازی محاسباتی هستند که می‌توانند عملکرد و تصمیم‌گیری خود را با استفاده از روش‌های یادگیری تقویتی، رشتهٔ زمانی، یادگیری عمیق، نظریهٔ بازی، مدل‌سازی سامانه‌های زیستی و نظریهٔ بازی (go) توضیح دهند و شامل واژه‌های مرتبط با روش‌ها و فن‌های هوش مصنوعی قابل‌تفسیر[1] (IAI) به‌کار رفته در یادگیری عمیق و یادگیری تقویتی است. کلمات هوش مصنوعی قابل تفسیر، یادگیری عمیق قابل‌تفسیر، توضیح و سطح محلی) باردو آریتا و همکاران، ۲۰۲۰؛ دوشی ولز و کیم، ۲۰۱۷؛ ریبیرو و همکاران، ۲۰۱۶؛ ژانگ و ژو، ۲۰۱۸) به روش‌ها و فن‌های هوش مصنوعی قابل‌تفسیر که می‌توانند نتایج و تصمیمات قابل‌درک و شفاف را از مدل‌های یادگیری عمیق و یادگیری تقویتی ارائه دهند، هستند.

جدول ۳. فراوانی واژگان کلیدی هوش مصنوعی توضیح‌پذیر و قدرت اتصال آن‌ها در خوشهٔ هفتم

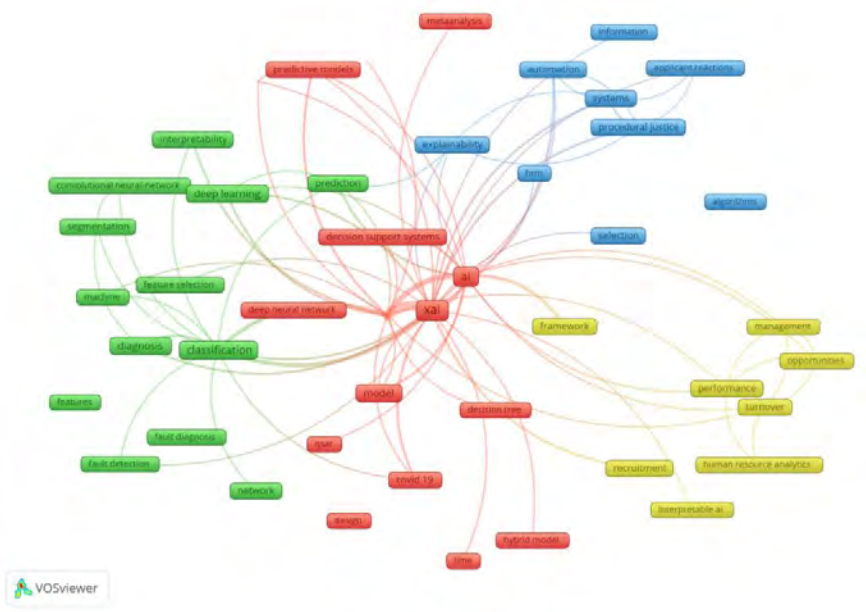
خوشه هفتم نارنجی	مجموع قدرت اتصال	رخداد	خوشه هفتم نارنجی	مجموع قدرت اتصال	رخداد
IAI	283	70	interpretable deep learning	62	15
reinforcement learning	133	40	deep reinforcement learning	41	12
time series	165	38	Electrocardiogram	62	11

1. Interpretable AI

خوشه هفتم نارنجی	مجموع قدرت اتصال	رخداد	خوشه هفتم نارنجی	مجموع قدرت اتصال	رخداد
Deep	109	23	local explanation	42	11
Game	87	20	atrial-fibrillation	48	10
biological system modeling	143	18	ECG	52	10
Go	72	17	Level	53	10

۴-۲. نقشهٔ دانش هوش مصنوعی توضیح‌پذیر در مدیریت منابع انسانی

با اضافه‌شدن واژگان مربوط به مدیریت منابع انسانی به کوئری جست‌وجو، نقشهٔ دانشی شامل ۴ خوشه (شکل ۵) حاصل شد. بیشترین پژوهش‌ها را محققان حوزهٔ علوم کامپیوتر، مهندسی، و بعد از آن اقتصاد کسب‌وکار انجام داده‌اند.



شکل ۵. نقشهٔ علمی تحلیل هم‌رخدادی مطالعات حوزه هوش مصنوعی توضیح‌پذیر و منابع انسانی

خوشهٔ اول (قرمز) به هوش مصنوع توضیح‌پذیر، مدل‌های یادگیری ماشین و تکنیک‌های توضیح‌پذیرکردن هوش مصنوعی اشاره دارد. مدل‌های پیش‌بینی و مدل‌سازی محاسباتی می‌توانند رفتار یا پویایی یک سیستم یا یک پدیده را شبیه‌سازی کنند (ژو و همکاران[1]، ۲۰۲۳). در این خوشه به

1. X. Zhu et al.

بررسی تفاوت عملکرد و خصوصیات مدل‌های ترکیبی، درخت تصمیم، شبکهٔ عصبی عمیق، طراحی مدل‌های جدید مثل QSAR (یان و همکاران[۱]، ۲۰۲۱) برای وظایف مختلف و شفافیت پرداخته شده است (آلتوف و همکاران[۲]، ۲۰۲۱؛ آلوس و همکاران[۳]، ۲۰۲۱) تا بتوان با تکنیک‌هایی مانند LIME[۴]، قابلیت توضیح و اطمینان مدل‌های هوش مصنوعی را افزایش داد (سوسژاک و مادیجن[۵]، ۲۰۲۳).

خوشهٔ دوم (سبز) شامل کلماتی است که نشان‌دهندهٔ رابطه بین یادگیری عمیق و هوش مصنوعی و کاربردهای آن در تشخیص و پیش‌بینی است که نشان می‌دهد چگونه شبکه‌های عصبی عمیق می‌توانند ویژگی‌های پنهان را استخراج کنند و برای دسته‌بندی و پیش‌بینی از آن استفاده کنند. همچنین این کلمات نشان می‌دهند که چگونه وظایفی مانند طبقه‌بندی، تقسیم‌بندی، یا پیش‌بینی می‌توانند به تمایز بین انواع مختلف داده‌ها کمک کنند. کلمات تقسیم‌بندی، تفسیرپذیری، انتخاب ویژگی، تشخیص عیب مربوط به حوزهٔ تفسیرپذیری و انتخاب ویژگی هستند که با استفاده از روش‌های مختلف، توضیحات قابل‌فهمی برای عملکرد و نتایج شبکه‌های عصبی پیچشی ارائه می‌دهند (یانگ و همکاران[۶]، ۲۰۲۰).

خوشهٔ سوم (آبی) به توانایی توضیح هوش مصنوعی در انتخاب پرسنل با استفاده از سیستم‌های هوشمند در مدیریت منابع انسانی اشاره دارد. در انتخاب و گزینش کارمندان، سیستم‌های هوشمند می‌توانند از داده‌ها و تجزیه‌وتحلیل برای غربالگری، ارزیابی، و رتبه‌بندی متقاضیان استفاده کنند (لانگر و همکاران[۷]، ۲۰۲۱؛ وشه و همکاران[۸]، ۲۰۲۲)، اما این سیستم‌ها نیاز به توضیح‌پذیری دارند، که می‌تواند ادراک متقاضیان از عدالت و شفافیت فرایند انتخاب را بهبود بخشد (لانگر و کونیگ[۹]، ۲۰۲۳). عدالت و شفافیت در انتخاب پرسنل از جنبه‌های مهم مدیریت منابع انسانی هستند که می‌توانند تأثیرات مختلفی بر واکنش‌های متقاضیان و مدیریت منابع انسانی و بر فرصت‌ها و عدالت اجتماعی داشته باشند. «عدالت رویه‌ای» نسبت به «اطلاعات» اهمیت بیشتری دارد (لانگر و همکاران، ۲۰۲۱) اشاره به این موضوع دارد که توانایی توجیه تفسیر تصمیمات ممکن است

1. Yan et al.
2. Althoff et al.
3. Alves et al.
4. Local Interpretable Model-agnostic Explanation (LIME)
5. Susnjak & Maddigan
6. Yang et al.
7. Langer et al.
8. Wesche et al.
9. Langer & König

اهمیت بیشتری نسبت به توضیح یا تفسیر آن‌ها داشته باشد.

خوشهٔ چهارم (زرد) شامل تحلیل منابع انسانی و ایجاد چارچوب تفسیر پذیر است. این خوشه به طراحی چارچوب‌های تفسیر پذیر هوش مصنوعی در تحلیل منابع انسانی برای پیش‌بینی و کاهش جابه‌جایی، افزایش عملکرد و رضایت کارکنان (اعتمادی و همکاران، ۲۰۲۳)، بهبود فرایند استخدام، توسعهٔ کارکنان، و ایجاد مزیت رقابتی برای سازمان‌ها اشاره می‌کند (شوداری[1]، ۲۰۲۳؛ هاریهاران[2]، ۲۰۲۲). پیش‌بینی موفقیت استخدام و طراحی یک مدل بهینه‌سازی عمومی برای جذب کارکنان با استفاده از این چارچوب‌ها امکان‌پذیر است (پساش و همکاران[3]، ۲۰۲۰). قرارگرفتن کلیدواژهٔ هوش مصنوعی تفسیرپذیر در این خوشه نشان می‌دهد این کلیدواژه نسبت‌به هوش مصنوعی توضیح‌پذیر در بین متخصصان حوزه منابع انسانی از رواج بیشتری برخوردار است.

جدول ٤. فراوانی واژگان کلیدی XHRS و قدرت اتصال آن‌ها

خوشه اول قرمز	رخداد	مجموع قدرت اتصال	خوشه دوم سبز	رخداد	مجموع قدرت اتصال	خوشه سوم آبی	رخداد	مجموع قدرت اتصال	خوشه چهارم زرد	رخداد	مجموع قدرت اتصال
xai	27	85	classification	9	39	explainability	4	20	framework	4	25
machine learning	20	72	deep learning	6	21	procedural justice	3	18	turnover	3	19
ai	18	63	diagnosis	5	15	systems	3	18	performance	3	19
model	7	15	prediction	4	23	automation	3	10	recruitment	3	16
covid 19	4	12	interpretability	3	15	selection	3	12	hr analytics	3	13
predictive models	3	13	segmentation	3	14	algorithms	2	12	opportunities	2	13
decision support system	3	7	network	3	9	applicant reactions	2	12	management	2	11
computational modeling	2	9	machine	2	14	information	2	7	interpretable ai	2	11
prevalence	2	8	CNN	2	13	hrm	2	10			
metaanalysis	2	8	fault detection	2	10						
decision tree	2	7	fault diagnosis	2	8						
qsar	2	6	feature selection	2	8						
lime	2	5	features	2	7						
hybrid model	2	5									
deep neural network	2	5									
design	2	4									

1. Chowdhury et al.
2. Hariharan et al.
3. Pessach et al.

۵. بحث و نتیجه‌گیری

هدف از این مطالعه بررسی و تحلیل تعامل، ارتباط، و اشتراک بین دو حوزه هوش مصنوعی توضیح‌پذیر و منابع انسانی بود که برای این منظور، دو نقشهٔ دانش، یکی فقط هوش مصنوعی توضیح‌پذیر (XAIS به‌عنوان نقشهٔ آنالیز اول) و دیگری از تعامل بین هوش مصنوعی توضیح‌پذیر و مدیریت منابع انسانی (XHRS به‌عنوان نقشهٔ آنالیز دوم) ارائه شد.

نقشهٔ هوش مصنوعی توضیح‌پذیر شامل هفت خوشهٔ «اصول و تکنیک‌های اساسی در هوش مصنوعی قابل‌تفسیر: از تحلیل داده تا تصمیم‌گیری اخلاقی»، «چارچوب یکپارچهٔ یادگیری ماشینی و هوش مصنوعی در مدیریت استراتژیک، نوآوری فناورانه، و توسعهٔ اجتماعی»، «جنبه‌ها و کاربردهای کلیدی یادگیری عمیق در پردازش سیگنال، تصویر، و زبان با توسعه و بهینه‌سازی مدل»، «تکنیک‌ها و فرایندهای یکپارچه در علم داده، تحلیل و حل مسئلهٔ محاسباتی»، «عناصر جامع تحلیل داده، یادگیری ماشینی، و تکنیک‌های مدل‌سازی پیش‌بینی»، «مفاهیم و تکنیک‌های اساسی در یادگیری ماشینی و علم داده» و «حوزه‌ها و تکنیک‌های پیشرفته در هوش مصنوعی و مدل‌سازی محاسباتی» که عموماً نشان‌دهندهٔ مفاهیم نظری، تکنیک‌ها و کاربردهای مختلف هوش مصنوعی توضیح‌پذیر است، می‌شود

این خوشه‌ها اگر چه نشان می‌دهند که چه پتانسیل و محدودیت‌هایی در کاربردهای مختلف با تحلیل و مدل‌سازی پدیده‌های پیچیده و پویا در حوزه‌های مختلف مانند سلامت، داروسازی، بیوانفورماتیک، پاتولوژی، اپیدمیولوژی، سلامت روان، تغییرات آب‌وهوا و سنجش از دور کمک کند و با ارزش‌ها، حقوق، و مسئولیت‌های انسانی سازگار باشد، و توضیحاتی برای پیش‌بینی‌ها و تصمیمات خود ارائه دهد؛ با این حال و با وجود همهٔ این واژگان هم‌رخداد متعدد مربوط به کاربردهای هوش مصنوعی، این کاربردها عموماً به حوزهٔ پزشکی مرتبط بود و مسائل حوزه علوم انسانی را کمتر درگیر می‌کرد.

کاربردهای هوش مصنوعی توضیح‌پذیر که در پایین‌ترین قسمت نقشهٔ هوش مصنوعی توضیح‌پذیر، متعلق به خوشهٔ بنفش و بخشی هم در خوشهٔ سبز، بین خوشه‌های قرمز و سبز واقع شده است که نشان می‌دهد یادگیری ماشین در حوزهٔ فنی و توضیح‌پذیری در حوزهٔ هستی‌شناسی، دو رکن اصلی کاربست هوش مصنوعی است.

نقشهٔ هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانیشامل چهار خوشه است که دو خوشهٔ اول به مفاهیم و تکنیک‌های هوش مصنوعی توضیح‌پذیر و دو خوشهٔ دیگر به حوزه‌های مختلف مدیریت منابع انسانی اشاره دارند. این نقشه با اضافه‌شدن منابع انسانی به کوئری ایجاد شد که تعداد مقالات

به‌شدت کاهش یافت و هم‌رخدادی کلمات نیزکمتر شد. علی‌رغم اضافه شدن خوشه‌های مربوط به منابع انسانی در نقشهٔ هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانیهیچ‌گونه ارتباط قابل تأملی با خوشه‌های مربوط به هوش مصنوعی مشاهده نشد. قرار گرفتن دو خوشهٔ مطالعات مدیریت و منابع انسانی در سمت راست نقشه و خوشهٔ فنیهوش و تخصصی هوش مصنوعی در سمت چپ و با فاصلهٔ زیاد مؤید فاصلهٔ زیاد تحقیقات و نبود همکاری و ارتباط معنادار بین محققان و دست‌اندرکاران این دو حوزه است. این شکاف ممکن است به دلیل پیش‌زمینه‌ها، دیدگاه‌ها، اهداف، و روش‌های متفاوت بین این دو حوزه و یا به دلیل نبودِاعتماد و تمایل مدیران منابع انسانی به استفاده از سیستم‌های هوشمند باشد. خوشهٔ قرمز در نقشهٔ هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانی که واژگان عمومی‌تر و مفهومی حوزهٔ هوش مصنوعی توضیح‌پذیر با نگاه سیستمی و مدلی آن را تشکیل می‌دهد، باعث اتصال دو حوزهٔ فنی و منابع انسانی شده است. به‌طور خاص «سیستم‌های پشتیبان تصمیم» ارتباط معنادارتری بین دو حوزه برقرار می‌کنند.

در نقشهٔ هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانی، لغات تخصصی حوزهٔ علوم کامپیوتر بسیار پر رنگ بود و الگوریتم‌های برنامه‌نویسی و روش‌های محاسباتی بسیاری به چشم می‌خورد که نقش پر رنگ محققان این حوزه که به صورتی عمیق در حال توسعهٔ این علم جدید هستند را نشان می‌داد. اما با اضافه شدن قید منابع انسانی در نقشهٔ هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانی، هم‌رخدادی چنین کلماتی مشاهده نشد و صرفا روش‌های کلی برنامه‌نویسی مثل یادگیری عمیق، از فیلتر هم‌رخدادی گذشتند. این موضوع نشان می‌دهد تعامل حوزهٔ منابع انسانی و هوش مصنوعی و به شدت کم و به شکلی سطحی است، به‌طوری که در نقشهٔ هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانی کلمات تخصصی هیچ کدام از دو حوزه وجود ندارد و صرفا کلمات بسیار عمومی و کلی هستند که آن هم ارتباطات معناداری ندارد که بتوان از آن رهگیری مطالعاتی استخراج کرد.

نقشهٔ هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانی نشان می‌دهد که تحقیقات و نوآوری‌های ناچیزی در تعامل بین هوش مصنوعی و منابع انسانی وجود دارد، لذا نیاز به توسعه و به‌کارگیری سیستم‌های هوش مصنوعی که بتوانند مدیریت، انتخاب، جذب، و حفظ سرمایهٔ انسانی در سازمان‌ها را پشتیبانی و بهبود بخشند، به‌شدت احساس می‌شود.

۶. پیشنهادها

یافته‌های این پژوهش هم به محققان و هم به فعالین حوزهٔ مدیریت منابع انسانی کمک شایانی می‌می‌کند. از این‌رو، براساس خوشه‌بندی‌های تعیین‌شده، که به تقسیم‌بندی موضوعات و کلیدواژه‌ها

در حوزه‌های مختلف هوش مصنوعی و یادگیری ماشین می‌پردازند، پیشنهادهایی ارائه می‌شود.

توسعهٔ مدل توضیحی به افزایش شفافیت و کارایی منجر می‌شود. برای افزایش شفافیت و کارایی سیستم‌های هوش مصنوعی توضیح‌پذیر، محققان باید بر طراحی مدل‌هایی متمرکز شوند که درخت‌های تصمیم‌گیری و شبکه‌های عصبی عمیق را در بر می‌گیرند. این مدل‌ها باید تصمیم‌گیری اخلاقی را تضمین کرده و با ارزش‌های سازمانی هماهنگ باشند. همچنین استفاده از فراتحلیل‌ها در هوش مصنوعی توضیح‌پذیر مدیریت منابع انسانی ارزیابی جامعی از مدل‌های پیش‌بینی‌کننده، شناسایی روندها، و آشکارسازی اثربخشی مدل در مجموعه داده‌ها و شرایط مختلف ارائه می‌دهد. براساس اطلاعات خوشهٔ اول، پیشنهاد می‌شود تا با طراحی نوین مدل‌هایی که درخت تصمیم و شبکه‌های عصبی عمیق را در بر می‌گیرند، شفافیت و کارایی سیستم‌ها افزایش داده شود. همچنین، استفاده از تکنیک‌های فراتحلیل در هوش مصنوعی توضیح‌پذیر می‌تواند به ارزیابی وسیع‌تر و دقیق‌تر اثربخشی مدل‌های پیش‌بینی کمک کند و با استفاده از روش‌هایی مانند LIME و QSAR، توانایی مدل‌ها در توضیح‌پذیری و اطمینان‌بخشی به نتایج را بهبود دهد.

از سوی دیگر، توسعهٔ مدل توضیح‌پذیر به تشخیص عیوب کمک شایانی می‌کند. توسعه و اصلاح الگوریتم‌های یادگیری عمیق مانند شبکه‌های عصبی کانولوشن به‌طور قابل‌توجهی تشخیص خطا را بهبود می‌بخشد. فیلتر کردن داده‌های پیشرفته می‌تواند تشخیص بهتری را به‌ویژه در زمینه‌های حساس مانند مراقبت‌های بهداشت روان منابع انسانی و سایر کاربردهای صنعتی امکان‌پذیر کند. همچنین انتخاب ویژگی باید بر تفسیرپذیری سیستم‌های هوش مصنوعی، به‌ویژه در خروج منابع انسانی و آسیب‌شناسی تأکید کند، تا اطمینان حاصل شود که الگوهای حیاتی نادیده گرفته نمی‌شوند. بر اساس اطلاعات خوشهٔ دوم پیشنهاد می‌شود که با توسعه و اصلاح الگوریتم‌های یادگیری عمیق مانند شبکه‌های عصبی پیچشی، کارایی در تشخیص و تعیین عیب افزایش یابد. همچنین، توسعهٔ روش‌هایی برای انتخاب ویژگی قوی برای تقویت تفسیرپذیری سیستم‌های هوش مصنوعی در حوزه‌های حساس مانند خروج منابع انسانی توصیه می‌شود.

همچنین، عدالت رویه‌ای در مدیریت منابع انسانی نگرانی‌های زیادی را برانگیخته است. هوش مصنوعی توضیح‌پذیر باید در استخدام و گزینش برای تقویت عدالت استفاده شود. مدل‌ها باید معیارهای تصمیم‌گیری را برای متقاضیان توضیح دهند و سوگیری‌هایی را که ممکن است ایجاد شود، بررسی کنند. برای درک تأثیر ابزارهای مبتنی بر هوش مصنوعی بر ادراک و واکنش داوطلبان، تحقیقات بیشتری لازم است، و بینش‌هایی در مورد اینکه چگونه هوش مصنوعی بر عدالت و اعتماد تأثیر می‌گذارد، مورد نیاز جدی است. بر اساس اطلاعات خوشهٔ سوم، پیشنهاد می‌شود که سیستم‌های

هوشمند توضیح‌پذیر در فرایندهای استخدام و انتخاب کارمند به کار گرفته شوند تا درک متقاضیان از عدالت و شفافیت فرایند بهبود یابد. از سوی دیگر، پیشنهاد می‌شود از استفاده از ابزارهای هوش مصنوعی که قابلیت توضیح‌پذیری ندارند در بخش‌های حیاتی منابع انسانی و مدیریت اطلاعات جدا خودداری شود. همچنین پیشنهاد می‌شود مطالعاتی برای ارزیابی وضعیت فعلی کاربرد هوش مصنوعی در مدیریت منابع انسانی و تأثیرات آن بر عکس‌العمل‌های متقاضیان انجام گیرد.

در پایان می‌توان گفت که چارچوب‌های توضیح‌پذیر در هوش مصنوعی توضیح‌پذیر می‌توانند به مدیریت استراتژیک و تجزیه‌وتحلیل منابع انسانی با ارائهٔ بینشی در مورد گردش مالی و عملکرد کارکنان کمک کنند. این مدل‌ها به سازمان‌ها کمک می‌کنند تا استراتژی‌های حفظ بهتری را طراحی کنند. بر اساس خوشهٔ چهارم، توسعهٔ چارچوب‌های هوش مصنوعی تفسیرپذیر که از مدیریت استراتژیک و تحلیل منابع انسانی پشتیبانی کند، توصیه می‌شود. این چارچوب‌ها باید در پیش‌بینی و مدیریت جابه‌جایی کارکنان و بهبود عملکرد و رضایت کارکنان (اعتمادی و همکاران، ۲۰۲۳) کاربرد داشته باشند. استفاده از هوش مصنوعی در تحلیل منابع انسانی برای پیش‌بینی موفقیت استخدام و بهینه‌سازی استراتژی‌های حفظ نیروی کار پیشنهاد می‌شود.

۷. تضاد منافع

هیچ تضاد منافعی وجود ندارد.

منـابـع

اعتمادی، محمد، چیت‌ساز، احسان، کوشکی، سحر، و جعفری، سیدمحمدعلی. (۱۴۰۳). هوش مصنوعی در مقابل روش‌های هدایت انسانی در ارزیابی استخدام منابع انسانی: فراترکیب مزایا و معایب. *مدیریت منابع انسانی پایدار*، ۶(۱۱)، ۲۱۴-۱۹۱. https://doi.org/10.22080/shrm.2024.5100

حسن‌زاده، م. (۲۰۲۲).عامل‌های هوشمند و تسهیلات مدیریت دانش: چت جی پی تی و بعد از آن. *علوم و فنون مدیریت اطلاعات*، (۴)۸، ۷-۲۲. https://doi.org/10.22091/stim.2023.2421

مهدی‌پور، محمدعلی، چیت‌ساز، احسان و اعتمادی، محمد. (۱۴۰۳). چشم‌اندازی بر تعامل انسان و هوش مصنوعی: تحلیل کتاب‌سنجی با تکنیک هم‌رخدادی. *فصلنامه علمی کارافن*، (۳)۲۱، ۱۳-۳۳.

https://doi.org/10.48301/kssa.2024.428257.2778

References

Abdel-Karim, B. M., Pfeuffer, N., & Hinz, O. (2021). Machine learning in information systems - a bibliographic review and open research issues. *Electronic Markets*, 31(3). https://doi.org/10.1007/s12525-021-00459-2

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adibfar, P. , Chitsaz, E. and Etemadi, M. (2024). Investigating the role of age and working memory on the relationship between motivation and risk-taking of potential entrepreneurs. *Journal of Entrepreneurship and Innovation Research*, 2(4), 1-15. **doi:** 10.22034/eir.2024.183611

Ahmadi, H., & Asare, F. (2017). Co-word Analysis Concept, Definition and Application. *Librarianship and Informaion Organization Studies*, 28(1), 125–145.

Aizenberg, E., & Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society*. https://doi.org/10.1177/2053951720949566

Albert, E. T. (2019). AI in talent acquisition: A review of AI-applications used in recruitment and selection. *Strategic HR Review*, 18. https://doi.org/10.1108/SHR-04-2019-0024

Alonso, J. M., Castiello, C., & Mencar, C. (2018). A bibliometric analysis of the explainable artificial intelligence research field. *Communications in Computer and Information Science*, 853. https://doi.org/10.1007/978-3-319-91473-2_1

Althoff, D., Rodrigues, L. N., & da Silva, D. D. (2021). Addressing hydrological modeling in watersheds under land cover change with deep learning. *ADVANCES IN WATER RESOURCES*, 154. https://doi.org/10.1016/j.advwatres.2021.103965

Alves, M. A., Castro, G. Z., Oliveira, B. A. S., Ferreira, L. A., Ramirez Jaime Arturod and Silva, R., & Guimaraes, F. G. (2021). Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. *COMPUTERS IN BIOLOGY AND MEDICINE*, 132. https://doi.org/10.1016/j.compbiomed.2021.104335

Anderson, J. R. (2005). *Cognitive Psychology and its implications*. Macmillan.

Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and critique of techniques for extracting

rules from trained artificial neural networks. *Knowledge-Based Systems*, 8(6), 373–389. https://doi.org/10.1016/0950-7051(96)81920-4

Annett, J. (2003). *Hierarchical Task Analysis* (pp. 17–35). https://doi.org/10.1201/9781410607775.ch2

Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* (pp. 63–77). Springer International Publishing. https://doi.org/10.1007/978-3-319-71153-9_3

Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural Machine Translation by Jointly Learning to Align and Translate. In International Conference on Learning Representations (ICLR).* https://arxiv.org/abs/1409.0473

Bankins, S., & Formosa, P. (2021). Ethical AI at work: The social contract for artificial intelligence and its implications for the workplace psychological contract. In M. Coetzee & A. Deas (Eds.), *Redefining the psychological contract in the digital era: Issues for research and practice*. Springer. https://doi.org/10.1007/978-3-030-63864-1_4

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58. https://doi.org/10.1016/j.inffus.2019.12.012

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve—Algorithm Aversion and Demonstrating the Ability to Learn. *Business & Information Systems Engineering*, 63(1), 55–68. https://doi.org/10.1007/s12599-020-00678-5

Bishop, C. M. (2016). *Pattern recognition and machine learning*. Springer.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series analysis*. John Wiley & Sons.

Bramer, M. (2016). *Principles of Data Mining*. Springer London. https://doi.org/10.1007/978-1-4471-7307-6

Buzydlowski, J. W. (2015). Co-occurrence analysis as a framework for data mining. *Journal of Technology Research*, 6.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection. *ACM Computing Surveys*, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

Change, I. P. O. C. (2018). *Global warming of 1.5°C.*

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, X.-Q., Ma, C.-Q., Ren, Y.-S., Lei, Y.-T., Huynh, N. Q. A., & Narayan, S. (2023). Explainable artificial intelligence in finance: A bibliometric review. *Finance Research Letters*, 56, 104145. https://doi.org/10.1016/j.frl.2023.104145

Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2023). Embedding

transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover. *INTERNATIONAL JOURNAL OF HUMAN RESOURCE MANAGEMENT*, 34(14), 2732–2764. https://doi.org/10.1080/09585192.2022.2066981

Clifford, G., Liu, C., Moody, B., Lehman, L., Silva, I., Li, Q., Johnson, A., & Mark, R. (2017, September 14). *AF Classification from a Short Single Lead ECG Recording: the Physionet Computing in Cardiology Challenge 2017*. https://doi.org/10.22489/CinC.2017.065-469

Cockburn, Iain. M., Henderson, R., & Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *National Bureau of Economic Research* (Issue May).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Cross, N. (2011). *Design thinking: Understanding how designers think and work.*

Das, A., & Rad, P. (2020). *Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey*.

Davenport, T., Guszcza, J., Smith, T., & Stiller, B. (2019). *Insight-driven organization*. Deloitte Insights.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). <title/>. *Proceedings of the 2019 Conference of the North*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dhamija, P., & Bag, S. (2020). Role of artificial intelligence in operations environment: a review and bibliometric analysis. *The TQM Journal*, 32(4), 869–896. https://doi.org/10.1108/TQM-10-2019-0243

Dietvorst, B., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144. https://doi.org/10.1037/xge0000033

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

Duda, R. O. (2022). *Pattern Classification, Third Edition*. John Wiley & Sons.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Etemadi, M., Chitsaz, E., & Abolghasemi Dehaqani, M. (2023a). The Myth of Rewards and Creative Performance: Should Companies Use Incentives to Boost Creativity in Personnel Performance? *3rd Iran Business Watch Conference 1402*. https://doi.org/10.2139/SSRN.4550849

Etemadi, M., Chitsaz, E., & Abolghasemi Dehaqani, M. (2023b). Unveiling the Complexity of the Reward, Creativity, and Performance Relationship: When Does Behavioral Theories Reward Backfire? *3rd Iran Business Watch Conference, 2023*. https://doi.org/10.2139/SSRN.4550749

Etemadi, M., Chitsaz, E., Abolghasemi Dehaqani, M., & Ghodratizadeh, F. (2024). The Interplay of Monetary Rewards, Expectations, and Ideation Quality: An Empirical Analysis. *Journal of Entrepreneurship Development*, 16(4), 116–142.

https://doi.org/10.22059/JED.2023.360337.654210

Etemadi, M., Chitsaz, E., & Ghodratizahed, F. (2024). The Paradox of Rewards: Reconsidering Employee Satisfaction and Ideation Performance. *Journal of Sustainable Human Resource Management*, 6(10). https://doi.org/10.22080/SHRM.2024.4599

Etemadi, M. , Chitsaz, E. , Koushki, S. , & Jafari, S. M. (2024). Artificial Intelligence (AI) vs. Human-Led Approaches in Human Resource Recruitment Assessment: A Meta-Synthesis of Advantages and Disadvantages. *Journal of Sustainable Human Resource Management*, 6(11), 214-191. **doi:** 10.22080/shrm.2024.5100 https://doi.org/10.22080/shrm.2024.5100

Etemadi, M. , Chitsaz, E. , Koushki, S. , & Jafari, S. M. (2024). Analyzing the Efficacy of Artificial Intelligence in Human Resource Recruitment: A Meta-analysis Approach. *Journal of Sustainable Human Resource Management*, 6(11), 214-191. **doi:** 10.22080/shrm.2024.5100 https://doi.org/10.22080/shrm.2024.5100

Ferreira, J., Mueller, J., & Papa, A. (2020). Strategic knowledge management: theory, practice and future challenges. *Journal of Knowledge Management*, 24(2), 121–126. https://doi.org/10.1108/JKM-07-2018-0461

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4). https://doi.org/10.1007/s11023-018-9482-5

Ghodratizadeh, F., Chitsaz, E., & Rostami, R. (2022). Investigating on the effects of continuous financial rewards on the performance of employees'idea generation in the ict industry of china. *Organizational Resources Management Researchs*. http://dorl.net/dor/20.1001.1.22286977.1401.12.3.7.1

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist*. McGraw-Hill Education (UK).

Goodfellow, I. (2016). *Deep Learning*. MIT Press.

Goodfellow, I. J., Vinyals, O., & Saxe, A. M. (2014). *Qualitatively characterizing neural network optimization problems.*

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, 4(37). https://doi.org/10.1126/scirobotics.aay7120

Guyon, I., & Elisseeff, A. (2003). *Feature extraction: Foundations and applications (Studies in Fuzziness and Soft Computing)* (Vol. 207). Springer Science & Business Media.

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2008). *Feature extraction*. Springer.

Hariharan, S., Robinson, R. R. R., Prasad, R. R., Thomas, C., & Balakrishnan, N. (2022). XAI for intrusion detection system: comparing explanations based on global and local scope. *JOURNAL OF COMPUTER VIROLOGY AND HACKING TECHNIQUES*. https://doi.org/10.1007/s11416-022-00441-2

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hattie, J. (2012). *Visible Learning for Teachers*. Routledge. https://doi.org/10.4324/9780203181522

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59–67. https://doi.org/10.1145/1743546.1743567

Hepenstal, S., & McNeish, D. (2020). *Explainable Artificial Intelligence: What Do You Need to Know?* (pp. 266–275). https://doi.org/10.1007/978-3-030-50353-6_20

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786), 504–507. https://doi.org/10.1126/science.1127647

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems*. The MIT Press. https://doi.org/10.7551/mitpress/1090.001.0001

Holstein, K., Vaughan, J. W., Daumé, H., Dudík, M., & Wallach, H. (2018). *Improving fairness in machine learning systems: What do industry practitioners need?* https://doi.org/10.1145/3290605.3300830

Huang, M. H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, 21(2). https://doi.org/10.1177/1094670517752459

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jain, A. K., & Li, S. Z. (2011). *Handbook of face recognition* (Vol. 1). Springer.

Jaladati, H. M., & Chitsaz, E. (2023). Unraveling the secrets to startup crowdfunding: Cognitive legitimacy in Initial Coin Offerings (ICOs). *Journal of Entrepreneurship Research*, 2(3), 1-22. https://doi.org/10.22034/jer.2023.2006614.1047

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

Jatobá, M., Santos, J., Gutierriz, I., Moscon, D., Fernandes, P. O., & Teixeira, J. P. (2019). Evolution of Artificial Intelligence Research in Human Resources. *Procedia Computer Science*, 164. https://doi.org/10.1016/j.procs.2019.12.165

Jensen, J. R. (2015). *Introductory Digital Image Processing*. Pearson.

Jessell, T. M., Siegelbaum, S. A., & Kandel, E. R. (2021). *Principles of Neural Science, Sixth Edition*. McGraw-Hill Education / Medical.

Jia, Q., Guo, Y., Li, R., Li, Y., & Chen, Y. (2018). A conceptual artificial intelligence application framework in human resource management. *Proceedings of the International Conference on Electronic Business (ICEB), 2018-December*.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*. https://doi.org/10.1038/s42256-019-0088-2

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing*. Prentice Hall.

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *Twenty-Eigth European Conference on Information Systems (ECIS2020), June*.

Kavousi, A., Chitsaz, E., Vahabie, A., & Lotfi, M. (2024). Exploring the effect of repetitive transcranial magnetic stimulation and coaching on ideation. *Psychological Researches in Management*, 10(2),

9-29. https://doi.org/10.22034/jom.2024.2022827.1167

Kazim, E., Koshiyama, A. S., Hilliard, A., & Polle, R. (2021). Systematizing audit in algorithmic recruitment. *Journal of Intelligence*, 9(3). https://doi.org/10.3390/jintelligence9030046

Kelleher, J. D., Mac Namee, B., & Arcy, A. (2018). *Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies.* MIT Press.

Kibert, C. J. (2016). *Sustainable construction: Green building design and delivery*. John Wiley & Sons.

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization.*

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

Koza, J. R., & Koza, J. R. . (1992). *Genetic Programming*. MIT Press.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web*, 591–600. https://doi.org/10.1145/1772690.1772751

langer m. (2022). Handbook of Research on Artificial Intelligence in Human Resource Management. In *Figures*. Edward Elgar Publishing. https://doi.org/10.4337/9781839107535

Langer, M., Baum, K., Koenig, C. J., Haehne, V., Oster, D., & Speith, T. (2021). Spare me the details: How the type of information about automated interviews influences applicant reactions. *INTERNATIONAL JOURNAL OF SELECTION AND ASSESSMENT*, 29(2), 154–169. https://doi.org/10.1111/ijsa.12325

Langer, M., & König, C. J. (2023). Introducing a multi-stakeholder perspective on opacity, transparency and strategies to reduce opacity in algorithm-based human resource management. *Human Resource Management Review*, 33(1), 100881. https://doi.org/10.1016/j.hrmr.2021.100881

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu, H., & Motoda, H. (2007). *Computational methods of feature selection*. CRC Press.

Lohn, A. J. (2020). *Estimating the Brittleness of AI: Safety Integrity Levels and the Need for Testing Out-Of-Distribution Performance. (arXiv:2009.00802). arXiv.* https://doi.org/10.48550/arXiv.2009.00802

Long, J., Shelhamer, E., & Darrell, T. (2014). *Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern*

Recognition (CVPR) (pp. 3431–3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Science and Systems*. John Wiley & Sons.

Lotfi, M., Chitsaz, E., & Rostami Asrabadi, S. (2023). Investigating the factors affecting the occurrence of law breaking in E-business entrepreneurs. *Education and Management of Entrepreneurship*, 1(1), 71-86. https://doi.org/10.22126/eme.2023.2506

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Manning, C., & Schutze, H. (1999). *Foundations of Statistical natural language processing*. MIT Press.

Markus Langer and Cornelius König. (n.d.). Explainability of artificial intelligence in human resources. In *Handbook of Research on Artificial Intelligence in Human Resource Management* (pp. 285–302).

Meacham, M. (2020). *AI in Talent Development: Capitalize on the AI Revolution to Transform the Way You Work, Learn, and Live*. Association for Talent Development. https://books.google.com/books?id=bOILEAAAQBAJ

Mehdipour, M., Chitsaz, E. and Etemadi, M. (2024). A Perspective on Human Interaction and Artificial Intelligence: Bibliometric Analysis with Co-occurrence Technique. *Karafan Journal*, 21(3), 13-33. **doi:** 10.48301/kssa.2024.428257.2778

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, 39(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Michael Armstrong, S. T. (2014). Armstrong`s Handbook Of Human Resource Management Prractice. In *Progress in Human Geography* (Vol. 13, Issue 1).

Mitchell, T. M. (1997). *Machine learning*. McGraw Hill series in computer science.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967. https://doi.org/10.1177/2053951716679679

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. https://doi.org/10.1038/nature14236

Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers*, 11(9), 1235. https://doi.org/10.3390/cancers11091235

Nambisan, S., Wright, M., & Feldman, M. (2019). The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes. *Research Policy*, 48(8), 103773.

https://doi.org/10.1016/j.respol.2019.03.018

Neff, G., & Nagy, P. (2016). Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, 10, 4915–4931. https://ijoc.org/index.php/ijoc/article/view/6277

Ochmann, J., Zilker, S., Michels, L., Tiefenbeck, V., & Laumer, S. (2021). The influence of algorithm aversion and anthropomorphic agent design on the acceptance of AI-based job recommendations. *International Conference on Information Systems, ICIS 2020 - Making Digital Inclusive: Blending the Local and the Global*.

Palmer, F. R., & Palmer, F. R. (1981). *Semantics*. Cambridge University Press.

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(2), 85. https://doi.org/10.3390/info14020085

Pessach, D., Singer, G., Avrahami, D., Ben-Gal, H. C., Shmueli, E., & Ben-Gal, I. (2020). Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *DECISION SUPPORT SYSTEMS*, 134. https://doi.org/10.1016/j.dss.2020.113290

Popper, K. (2005). *The Logic of Scientific Discovery*. Routledge. https://doi.org/10.4324/9780203994627

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking.*

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *Model-Agnostic Interpretability of Machine Learning*.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. In *Nature Machine Intelligence* (Vol. 1, Issue 5). https://doi.org/10.1038/s42256-019-0048-x

Russell, S., & Norvig, P. (2010). *Intelligence artificielle: Avec plus de 500 exercices*. Pearson Education France.

Saltelli, A., Chan, K., & Scott, E. M. (2009). *Sensitivity analysis*. Wiley.

Samarasinghe, K. R., & Medis, Dr. A. (2020). Artificial Intelligence Based Strategic Human Resource Management (AISHRM) For Industry 4.0. *Global Journal of Management and Business Research*. https://doi.org/10.34257/gjmbrgvol20is2pg7

Santoni de Sio, F., & Hoven, J. (2019). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*. https://doi.org/10.3389/frobt.2018.00015

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). *Deep learning with convolutional neural networks for EEG decoding and visualization*. https://doi.org/10.1002/hbm.23730

Scholz, T. M. (2019). Big data and human resource management. In J. S. Pederson & A. Wilkinson (Eds.), *Big data: Promise, application and pitfalls*. Edward Elgar. https://doi.org/10.4337/9781788112352.00008

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2016). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. https://doi.org/10.1007/s11263-019-01228-7

Seto, K. C., Fragkias, M., Güneralp, B., & Reilly, M. K. (2011). A Meta-Analysis of Global Urban Land Expansion. *PLoS ONE*, 6(8), e23777. https://doi.org/10.1371/journal.pone.0023777

Shanmugam, S., & Garg, L. (2015). Model employee appraisal system with artificial intelligence capabilities. *Journal of Cases on Information Technology*, 17(3). https://doi.org/10.4018/JCIT.2015070104

Shiralian, S., Ghodratizadeh, F., Talebi, K. and Chitsaz, E. (2024). The innovation equation: Understanding the connection between team cohesion, motivation, and design thinking mindset in boosting employee's innovative performance. *Journal of Entrepreneurship Development*, 17(2), 188-210. https://doi.org/10.22059/jed.2023.360821.654213

Shiralian, S., Talebi, K., Ghodratizadeh, F., & Chitsaz, E. (2025). The role of design thinking in promoting innovative behavior: An exploration of team cohesion and motivation. *Journal of General Management*. https://doi.org/10.1177/03063070251322856

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Simon, H. A. (1979). *Models of Thought*. Yale University Press.

Sokol, K., & Flach, P. (2020). Explainability fact sheets. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 56–67. https://doi.org/10.1145/3351095.3372870

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, 56(4). https://doi.org/10.1109/MSPEC.2019.8678513

Sun, Y., Jiang, H., Hwang, Y., & Shin, D. (2018). Why should I share? An answer from personal information management and organizational citizenship behavior perspectives. *Computers in Human Behavior*, 87, 146–154. https://doi.org/10.1016/j.chb.2018.05.034

Susnjak, T., & Maddigan, P. (2023). Forecasting patient demand at urgent care clinics using explainable machine learning. *CAAI TRANSACTIONS ON INTELLIGENCE TECHNOLOGY*, 8(3), 712–733. https://doi.org/10.1049/cit2.12258

Tan, Z., Wang, M., Xie, J., Chen, Y., & Shi, X. (2018). Deep Semantic Role Labeling With Self-Attention. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). https://doi.org/10.1609/aaai.v32i1.11928

Tran, B., Vu, G., Ha, G., Vuong, Q.-H., Ho, M.-T., Vuong, T.-T., La, V.-P., Ho, M.-T., Nghiem, K. -C., Nguyen, H., Latkin, C., Tam, W., Cheung, N.-M., Nguyen, H.-K., Ho, C., & Ho, R. (2019). Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. *Journal of Clinical Medicine*, 8(3), 360. https://doi.org/10.3390/jcm8030360

Upadhyay, A. K., Khandelwal, K., & Iyengar, J. (2022). AI Revolution in HRM: The New Scorecard. In *AI Revolution in HRM: The New Scorecard*. https://doi.org/10.4135/9789354792861

Vardarlier, P., & Zafer, C. (2020). Use of Artificial Intelligence as Business Strategy in Recruitment Process and Social Perspective. In *Contributions to Management Science*.

https://doi.org/10.1007/978-3-030-29739-8_17

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*.

Vilone, G., & Longo, L. (2020). *Explainable Artificial Intelligence: a Systematic Review*.

Votto, A. M., Valecha, R., Najafirad, P., & Rao, H. R. (2021). Artificial Intelligence in Tactical Human Resource Management: A Systematic Literature Review. *International Journal of Information Management Data Insights*, 1(2). https://doi.org/10.1016/j.jjimei.2021.100047

Wachter, S., Mittelstadt, B., & Russell, C. (2017). *Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR.*

Wesche, J. S., Hennig, F., Kollhed, C. S., Quade, J., Kluge, S., & Sonderegger, A. (2022). People's reactions to decisions by human vs. algorithmic decision-makers: the role of explanations and type of selection tests. *EUROPEAN JOURNAL OF WORK AND ORGANIZATIONAL PSYCHOLOGY*. https://doi.org/10.1080/1359432X.2022.2132940

Yang, Y., Yin, J., Zheng, H., Li, Y., Xu, M., & Chen, Y. (2020). Learn Generalization Feature via Convolutional Neural Network: A Fault Diagnosis Scheme Toward Unseen Operating Conditions. *IEEE Access*, 8, 91103–91115. https://doi.org/10.1109/ACCESS.2020.2994310

Yan, J., Yan, X., Hu, S., Zhu, H., & Yan, B. (2021). Comprehensive Interrogation on Acetylcholinesterase Inhibition by Ionic Liquids Using Machine Learning and Molecular Modeling. *ENVIRONMENTAL SCIENCE & TECHNOLOGY*, 55(21, SI), 14720–14731. https://doi.org/10.1021/acs.est.1c02960

Zhang, Q., & Zhu, S. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 27–39. https://doi.org/10.1631/FITEE.1700808

Zhou, C., & Paffenroth, R. C. (2017). Anomaly Detection with Robust Deep Autoencoders. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 665–674. https://doi.org/10.1145/3097983.3098052

Zhu, J., Liapis, A., Risi, S., Bidarra, R., & Youngblood, G. M. (2018). Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 1–8. https://doi.org/10.1109/CIG.2018.8490433

Zhu, X., Wang, D., Pedrycz, W., & Li, Z. (2023). Fuzzy Rule-Based Local Surrogate Models for Black-Box Model Explanation. *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, 31(6), 2056–2064. https://doi.org/10.1109/TFUZZ.2022.3218426