J. Strateg. Manag. Stud. Vol. 15, No. 59, Fall 2024, PP: 45-67 ISSN (Print): 2228-6853, (Online): 2676-6744

Journal of Strategic Management Studies

Homepage: https://www.smsjournal.ir/?lang=en

Original Research Article **10.22034/smsj.2023.387893.1810**

Identifying the design requirements of explainable artificial intelligence systems

Zahra Hemmat, PhD student, Faculty of Economics and Administrative Science, Ferdowsi University of Mashhad, Mashhad, Iran
Mohammad Mehraeen*, Professor, Faculty of Economics and Administrative Science, Ferdowsi University of Mashhad, Mashhad, Iran
Rahmat Allah Fattahi, Professor, Faculty of Education and Psychology, Ferdowsi University of Mashhad, Mashhad, Iran
Farhad Shirani, Assistant Professor, Information Technology Research Department, Research Institute for Information Science and Technology, Tehran, Iran

ARTICLE INFO

Article History
Received: 1 March 2023
Revised: 24 September 2023
Accepted: 25 January 2024

Keywords
Ability to understand, Ability to rule, Ability to persuade, Prediction accuracy, Transparency, Usefulness

Corresponding Author
Email: mehraeen@um.ac.ir

ABSTRACT

With the expansion of the use of artificial intelligence algorithms in government organizations, concerns about the social responsibility of using intelligent agents such as transparency, accountability and fairness have been raised in government and academic areas. Accordingly, the research's aim is to explain the structural-interpretive model of the design requirements of artificial intelligence systems with the ability to explain decisions based on human participation and artificial intelligence. The exploratory mixed method of action-based design-fuzzy Delphi and structural-interpretive modeling used. The field of research is the Legal Department of the Judiciary. The participants are professionals from the that centers are 15 people in total. Based on the findings, the model includes five features of understanding ability, governance ability, persuasion ability, prediction accuracy (descriptive), transparency and usefulness. These capabilities classified in two dimensions. The ability dimension includes the ability to understand, the ability to rule, and the ability to persuade. The dimension of research also includes the features of prediction accuracy, transparency and usefulness. In addition, the model can explain the mechanism of enhancing intelligence in human-artificial intelligence interaction.

How to cite this article:
Hemmat, Z., Mehraeen, M., Fattahi, R. A., & Shirani, F. (2024). Identifying the design requirements of explainable artificial intelligence systems, *Journal of Strategic Management Studies*, 59(15), 45-67 (In Persian with English abstract). https://doi.org/10.22034/smsj.2023.387893.1810

©2023 The author(s). This is an open access article distributed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s)and the source.

J. Strateg. Manag. Stud., Vol. 15, No. 59, Fall 2024
Web**:** http://smsjournal.ir
Doi: **10.22034/smsj.2023.387893.1810**

EXTENDED ABSTRACT

Introduction

The use of artificial intelligence algorithms in public sectors is increasingly expanding. At the same time, there are still concerns about the social responsibility of using intelligent agents, such as transparency, accountability and fairness in governmental and academic communities. Based on this, the purpose of this research is to develop a theorical model that includes the requirements for the design of explainable artificial intelligence systems in Human-Artificial Intelligence interaction for decision making. The most important issue is understanding how the system reaches a decision by users, which can be investigated. Because as much as human decision-makers are expected to provide explanations (explain) about their decisions, artificial intelligence systems can also be asked to explain proposed solutions. In turn, this topic provides prescriptive knowledge for system designers to create new insights for users by considering the linkage of user information with various sources through the use of artifacts. Therefore, the scope of this research is the enhancement of intelligence in which artificial intelligence models provide recommendations for human users. According to these cases, the research's goel is to identify the descriptive and prescriptive knowledge required for designing a class of recommender systems based on human-artificial intelligence interaction. To achieve this goal, it is borrowed from the theory of ability to specify the relationships between the extracted categories. Finally, it is discussed how these requirements are used in design cycles.

Methodology

A Mixed methods research designs: Action design rsearch and Fuzzy delphi method and interpretive structural modeling approach used to develop and evaluate the design principles of that. We follow the design research developed by Mlarki et al. The General Department of Revision of Laws and Regulations in the Legal Department of the Judiciary has been selected as the context of research. The participants of this research are professionals from the General Department of Revision of Laws and Regulations and the Information Technology Center, who together with researchers (supervisors and students) constitute the research team, which is a total of 15 people. Using the triangulation technique, data have been collected from different sources. And data analysis was done in two steps: in the first step, by continuous refining the concepts, the extracted components were aggregated in theoretical dimensions. Also, a data structure was created by combining concepts, components and dimensions. In the second step, the research lens was used to develop the theory.

Results and Discussion

The research develops a framework that conceptualizes the characteristics of explainable artificial intelligence systems that include of understanding capability, governance capability, persuasion capability, predictive (descriptive) accuracy, transparency and usefulness. These characteristics classified in two dimensions. The affordance dimension includes the ability to understand, the ability to rule, and the ability to persuade. The actualization dimension includes the features of accuracy of prediction, transparency and usefulness. In addition, the model can explain the mechanism of enhancing intelligence in human-artificial intelligence interaction. Therefore, we proposed following propositions in designing explainable artificial intelligence system for humans -artificial intelligence interaction:1) understanding capability of intelligent agent in the human - artificial intelligence interaction lead to intelligence reinforcing; 2) governance capability of intelligent agent in the human - artificial intelligence interaction lead to intelligence reinforcing; 3) persuasion capability of intelligent agent in the human - artificial intelligence interaction lead to intelligence reinforcing; 4) transparency of intelligent agent in the human - artificial intelligence interaction lead to intelligence reinforcing; and 5) the accuracy of prediction of intelligent agent in the human - artificial intelligence interaction lead to intelligence reinforcing .Finally, from the indiscernibility aspect, the findings of this research emphasize the explainable algorithmic activities and increase the understanding and persuasiveness abilities of users through the feature of algorithmic transparency. Since it is difficult to assess indiscernibility. In other words, algorithm-based decision-making process is understandable for some and not understandable for others. The design

requirements of this research are a practical guide for clarifying the algorithmic activity in policy-making according to the user understanding and the purpose of use of artificial intelligence.

Conclusion

We argued that users and artifacts create an affordance in each other that lead to learning. Accordingly, the following can be considered as the theoretical contribution of this research. First, through developing a theoretical model, we established the mechanism of designing systems based on human-artificial intelligence interaction. In addition, from a human-oriented perspective, we identify the characteristics that users need to enhance intelligence which are: capability of understanding, capability of governance, and capability of persuasion; Second, it provides knowledge about the solution space. In other words, the intelligent agent actualized the user's affordance through the accuracy of prediction and transparency; Next, we provide sets of requirements for the implementation of human-artificial intelligence systems. These set of requirements in a theoretical model, constitute a guide for the artificial systems design principles in organizations, which has been the main concern of prior research; and finally, we showed that the design of Human-Artificial Intelligence (AI) interaction systems for decision making is not only technical, but also technical, social and organizational elements are intertwined in different cycles, which correspond to three related and interdependent aspects of AI management, i.e. automation, learning, indiscernibility: In terms of automation, the findings of this research showed that the policy-making functions cannot be coded. As a result, AI automation is constraining and AI should be tools to augmentation. In other hand, augmentation can be automated, so augmentation can become automatic over time. Accordingly, artificial intelligence tools can automate policy-making processes through the features of transparency and predictive accuracy that needed for policy maker's affordance, i.e. the ability to understand, the ability to rule, and the ability to persuade. In terms of learning, the findings of this research have emphasized the capacity of machine learning algorithms for semantic search to increase the accuracy of artificial intelligence prediction.

Keywords: Ability to understand, Ability to rule, Ability to persuade, Prediction accuracy, Transparency, Usefulness

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

مطالعات مدیریت راهبردی، دوره ۱۵، شماره ۵۹، پاییز ۱۴۰۳، صفحه: ۴۵ تا ۶۷ شاپای چاپی: ۶۸۵۳-۲۲۲۸ الکترونیکی: ۶۷۴۴-۲۶۷۶

مطالعات مدیریت راهبردی

Homepage: https://www.smsjournal.ir

مقاله پژوهشی e 10.22034/smsj.2023.387893.1810 OPEN ACCESS

شناسایی الزامات طراحی و قابلیت تبیین سیستم‌های هوش مصنوعی

زهرا همت، دانشجوی دکتری، دانشکده علوم اداری و اقتصادی، دانشگاه فردوسی مشهد، مشهد، ایران

محمد مهرآیین *، استاد، دانشکده علوم اداری و اقتصادی، دانشگاه فردوسی مشهد، مشهد، ایران

رحمت الله فتاحی، استاد، دانشکده علوم تربیتی و روانشناسی، دانشگاه فردوسی مشهد، مشهد، ایران

فرهاد شیرانی، استادیار، پژوهشکده فناوری اطلاعات، پژوهشگاه علوم و فناوری اطلاعات ایران، تهران، ایران

اطلاعات مقاله

سابقه مقاله

تاریخ دریافت: ۱۴۰۱/۱۲/۱۰

تاریخ بازنگری: ۱۴۰۲/۰۷/۰۲

تاریخ پذیرش: ۱۴۰۲/۱۱/۰۵

واژه‌های کلیدی

قابلیت درک،
قابلیت حکمرانی،
قابلیت اقناع،
دقت پیش‌بینی،
شفافیت،
سودمندی

ایمیل نویسنده مسئول

mehraeen@um.ac.ir

چکیده

با گسترش کاربرد الگوریتم‌های هوش مصنوعی در سازمان‌ها و بخش‌های دولتی، دغدغه‌هایی در مورد مسئولیت اجتماعی کاربست عامل‌های هوشمند همانند شفافیت، پاسخگویی و انصاف در محافل دولتی و دانشگاهی مطرح شده است. بر همین اساس، هدف پژوهش، تبیین مدل ساختاری- تفسیری الزامات طراحی سیستم‌های هوش مصنوعی با قابلیت تبیین در تصمیم‌گیری‌های مبتنی بر مشارکت انسان و هوش مصنوعی است. برای رسیدن به این هدف، از روش آمیخته اکتشافی طراحی مبتنی بر اقدام- دلفی فازی و مدل‌سازی ساختاری- تفسیری برای توسعه و ارزیابی اصول طراحی سیستم هوش مصنوعی با قابلیت تبیین استفاده می‌شود. زمینه پژوهش، ادارهٔ کل تنقیح قوانین و مقررات در معاونت حقوقی قوهٔ قضائیه است. مشارکت‌کنندگان پژوهش، دست اندرکارانی از ادارهٔ کل تنقیح قوانین و مقررات و مرکز فناوری اطلاعات بوده که به همراه پژوهشگران تیم پژوهش را تشکیل می‌دهند و در مجموع ۱۵ نفر هستند. بر اساس یافته‌های پژوهش، مدل، پنج ویژگی قابلیت درک، قابلیت حکمرانی، قابلیت اقناع، دقت پیش‌بینی (توصیفی)، شفافیت و سودمندی را در بر می‌گیرد. این قابلیت‌ها در دو بُعد طبقه‌بندی شدند. بُعد توانش شامل قابلیت درک، قابلیت حکمرانی و قابلیت اقناع است. بُعد محقق‌سازی نیز شامل ویژگی‌های دقت پیش‌بینی، شفافیت و سودمندی است. افزون بر این، مدل می‌تواند سازوکار تقویت هوشمندی در تعامل انسان-هوش مصنوعی را تبیین کند.

استناد به این مقاله: همت، زهرا؛ مهرآیین، محمد؛ فتاحی، رحمت الله؛ شیرانی، فرهاد (۱۴۰۳). شناسایی الزامات طراحی با قابلیت تبیین سیستم‌های هوش مصنوعی. مطالعات مدیریت راهبردی، ۵۹(۱۵)، ۴۵-۶۷.

۱. مقدمه

کاربرد الگوریتم‌های هوش مصنوعی در سازمان‌ها و بخش‌های دولتی بطور فزاینده در حال گسترش است. به طوری که تکیه بر فناوری‌های هوش مصنوعی عامل تحول این بخش معرفی شده است. پذیرش هوش مصنوعی سازمان‌ها و دولت‌ها، مرهون بلوغ الگوریتم‌های یادگیری ماشین است که به سیستم‌های پشتیبان تصمیم این امکان را می‌دهد تا رفتارشان را بر پایه داده‌ها تطبیق دهند. به موجب این پیشرفت‌ها، راه‌حل‌های مدیریتی، بطور بالقوه اثربخش‌تر، کارآمدتر و کم هزینه‌تر می‌شوند. افزون بر این، پیامد استفاده از هوش مصنوعی در سازمان‌ها، تصمیمات عاری از محدودیت‌ها و تعصبات شناختی است. این به نوبهٔ خود، توجیه مهمی برای پذیرش چنین فناوری‌هایی در زمینه‌های بسیار مهم برای حوزه‌هایی همچون قانون‌گذاری است [۲،۲۸،۳۰]. با این وجود، دغدغه‌هایی در مورد مسئولیت اجتماعی تصمیم‌گیری الگوریتمی، همانند شفافیت، پاسخگویی و انصاف در محفل‌های دولتی و دانشگاهی مطرح شده است [۳،۸،۲۸،۳۰].

دغدغه‌ها بر دو عنصر تصمیم‌گیری الگوریتمی متمرکز هستند: نخست، الگوریتم‌ها به مثابهٔ جعبهٔ سیاه عمل می‌کنند؛ این ویژگی، امکان جلوگیری از اشتباهات انسانی در مسائل احتمالی اعتبار و یا سوءگیری داده‌ها را از بین می‌برد؛ دوم، تصمیم‌گیری الگوریتمی، فضای اختیار را برای تصمیم‌گیرندگان انسانی کاهش می‌دهد. در حالی که قدرت اختیار برای تضمین عدالت اداری، چشم‌پوشی از خطاهای احتمالی و تطبیق تصمیمات با شرایط خاص بسیار حیاتی است [۲۸،۳۴] پژوهشگران در پاسخ به این دغدغه‌ها، ساختارهای تصمیم دوسوتوان را پیشنهاد کردند که بطور مؤثر نقاط قوت تصمیم‌گیرندگان انسانی و هوش مصنوعی را تلفیق می‌کنند و در عین حال آسیب‌پذیری‌های هر یک از این روش‌ها را به حداقل می‌رسانند [۱۳،۳۷،۱۶،۳۴،۱۸،۹]. از سوی دیگر، مفهوم سیستم‌های هوش مصنوعی با قابلیت تبیین، پس از دهه‌ها دوباره احیاء شده است و پژوهشگران در پی طراحی و پیاده‌سازی سیستم‌های با قابلیت تبیین هستند که به درک و اعتماد کاربران، نسبت به عامل‌های هوشمند کمک می‌کند[۲۳،۲۶].

با وجود پژوهش‌های متنوع در توضیح مدل‌های جعبه سیاه اما هنوز مسائل پژوهشی بدون پاسخ وجود دارد. مهم‌ترین مسأله، درک نحوهٔ رسیدن سیستم به یک تصمیم توسط کاربران است که می‌توان آن را مورد بررسی قرار داد. زیرا به همان اندازه که از تصمیم‌گیرندگان انسانی انتظار می‌رود در مورد تصمیمات خود، توضیحاتی را ارائه و تبیین کنند، از سیستم‌های هوش مصنوعی نیز می‌توان در مورد راه‌حل‌های پیشنهادی توضیح خواست. به نوبهٔ خود، این موضوع دانش تجویزی را برای طراحان سیستم فراهم می‌کند که بتوانند با در نظر گرفتن پیوند اطلاعات کاربر با منابع مختلف از طریق استفاده از مصنوع، بینش‌های جدیدی را برای کاربران ایجاد نمایند [۱۵،۳۴]. با این وجود، پژوهش‌های موجود در زمینهٔ سیستم‌های هوش مصنوعی با قابلیت تبیین، تنها از رویکرد شهودی پژوهشگران در طراحی سیستم‌هایی با ویژگی‌های تبیین خوب استفاده می‌کنند و از چارچوب‌های نظری موجود در علوم اجتماعی در مورد ویژگی‌های تبیین بهره نمی‌گیرند [۲۳]. در بخش دولتی نیز، سیستم‌های هوش مصنوعی از طریق برون‌سپاری، طراحی و پیاده‌سازی می‌شوند که در آن فرایندهای داخلی سیستم برای کاربران به طور تقریبی و ناقص تفسیر می‌شوند. از این جهت، ضرورت دارد که کاربران در توسعهٔ سیستم‌های با قابلیت تبیین مشارکت نمایند و لازمهٔ آن همکاری بین طراحان، کاربران و پژوهشگران است [۳۴]. بنابراین، ضرورت توجه به ویژگی‌های مکانی در طراحی سیستم‌های با قابلیت تبیین احساس می‌شود، زیرا این غفلت، توانایی دست‌اندرکاران برای کسب بینش از داده‌های سازمان را محدود می‌کند و الزامات تعامل انسان– هوش مصنوعی در یک بستر سازمانی معین مشخص نمی‌شود. بر همین اساس، این پژوهش با روش طراحی مبتنی بر اقدام که در آن تعریف نیازمندی‌ها، یکی از اشکال اصلی دانش طراحی است، به دنبال پاسخ به پرسش زیر است:

الزامات تبیین خوب در طراحی سیستم‌های مبتنی بر تعامل انسان– هوش مصنوعی برای تصمیم‌گیری شامل چه مؤلفه‌ها و ابعادی است؟

در این مسیر، ادارهٔ کل تنقیح قوانین و مقررات در قوهٔ قضائیه بستر پژوهش انتخاب شد که به منظور تسریع و نظام‌مند کردن تنقیح قوانین و مقررات در قوهٔ قضائیه، به دنبال طراحی و پیاده‌سازی سیستم‌های الکترونیکی تنقیح قوانین و مقررات قرار گرفته است. افزون بر این، با توجه به انبوه قوانین و مقررات، تعامل انسان–هوش مصنوعی به منظور رفع ابهام، تعارض، تداخل و تکرار از اهمیت زیادی برخوردار است. بنابراین، دامنهٔ این پژوهش، تقویت هوشمندی است که در آن مدل‌های هوش مصنوعی توصیه‌هایی را برای کاربران انسانی فراهم می‌کنند. با توجه به این موارد، هدف این پژوهش، شناسایی دانش توصیفی و تجویزی مورد نیاز طراحی طبقه‌ای از سیستم‌های توصیه‌گر مبتنی بر تعامل انسان– هوش مصنوعی است. برای رسیدن به این هدف، از نظریهٔ توانش برای مشخص کردن روابط میان مقوله‌های استخراج شده وام گرفته می‌شود. در نهایت بحث می‌شود که چگونه این الزمات در چرخه‌های طراحی مورد استفاده قرار می‌گیرند.

۲. مبانی و چارچوب نظری تحقیق

مبانی نظری تعامل انسان- هوش مصنوعی. دانشگاهیان و دست‌اندرکاران بر این موضوع اتفاق نظر دارند که در آینده، سازمان‌ها نیازمند یک مدل همکاری بین انسان و هوش مصنوعی هستند که هم‌افزایی را به ارمغان می‌آورد [۱۸،۱۶،۳۴]. به عبارت دیگر، هوش مصنوعی به افزونهٔ مغز انسان تبدیل می‌شود و این دو کاملاً با هم همکاری می‌کنند [۳۵]. با این وجود، یکی از محدودیت‌های مدل‌های تصمیم‌گیری در حوزهٔ تعامل انسان و هوش مصنوعی، فقدان تبیین در فرایند تصمیم‌گیری است که نگرانی‌هایی را دربارهٔ مسئولیت اجتماعی هوش مصنوعی ایجاد کرده است. بنابراین، تا زمانی که نتوان فرایند تصمیم‌گیری هوش مصنوعی را از جعبه سیاه خارج نمود، امکان توسعهٔ یک نمونه اولیه از سیستم‌های تعامل انسان–هوش مصنوعی وجود ندارد. افزون بر این،کارکرد اصلی تبیین تسهیل یادگیری است [۱۵،۲۳،۳۷]. بر همین اساس، هوش مصنوعی با قابلیت تبیین به رفع این ابهام کمک می‌کند.

هوش مصنوعی با قابلیت تبیین. درمبانی نظری داده‌کاوی و یادگیری ماشین به این مسألهٔ مهم پرداخته می‌شود که ماشین‌ها و الگوریتم‌ها اغلب نمی‌توانند بینش‌هایی در مورد رفتارشان و فرایندهای فکری ارائه نمایند. بر همین اساس، هوش مصنوعی با قابلیت تبیین معرفی شده است تا این امکان را برای عامل‌های هوشمند و قسمت‌هایی از فرایندهای داخلی آن فراهم نماید تا بتوانند شفاف‌تر باشند و جزئیات تصمیم‌گیری خود را توضیح دهند [۱۴]. بنابراین، قابلیت تبیین به توانایی مدل هوش مصنوعی در شرح دلایل رفتار و ارائهٔ بینش‌هایی در مورد علت‌های تصمیمات گرفته شده به انسان اطلاق می‌شود [۲۶]. کاربردهای تبیین در بسیاری از قلمروهای فرعی هوش مصنوعی، نظیر توجیه رفتار عامل هوشمند، اشکال‌زدایی مدل‌های یادگیری ماشین، توضیح تصمیم‌گیری در حوزهٔ پزشکی و توضیح پیش–بینی‌های طبقه‌بندی‌کننده‌ها مورد بررسی قرار گرفته است [۲۳]. بطورکلی، قابلیت تبیین هوش مصنوعی با دو معیار قابلیت تفسیر و جامعیت مورد ارزیابی قرار می‌گیرد. قابلیت تفسیر بدین معنا است که فرایندهای داخلی یک سیستم برای انسان قابل درک باشد. این معیار به شناخت، دانش و تعصبات کاربر بستگی دارد. به عبارت دیگر، برای اینکه یک سیستم از قابلیت تفسیر برخوردار باشد، توصیف‌ها باید به اندازه‌ای ساده باشند که برای کاربر غیرمتخصص، معنادار باشند؛ جامعیت به معنای میزان دقت در توصیف عملکرد سیستم است. به عبارت دیگر تبیین، زمانی کامل می‌شود که بتوان رفتار سیستم را در شرایط مختلف پیش‌بینی کرد [۱۴،۱۵] به تازگی تمرکز بر رویکردهای انسان‌محور در این قلمرو پژوهشی افزایش یافته است. با این وجود، دانش توصیفی محدودی در مورد ادغام عوامل انسانی در فرایند توسعهٔ تبیین‌های ایجاد شده با هوش مصنوعی وجود دارد که پیامد آن تقویت هوشمندی است [۱۲،۳۱]. بررسی مبانی نظری، بیانگر این است که استدلال از طریق بهترین تبیین، رهنمون‌های ارزشمندی برای طراحی سیستم‌های با قابلیت تبیین ارائه می‌کند. زیرا، تبیین، تنها بیان روابط علت و معلولی نیست، بلکه نقش زمینه را هم باید در نظر گرفت [۲۳،۲۶]. بر همین اساس، نظریه توانش در چارچوب نظری پژوهش استفاده می‌شود.

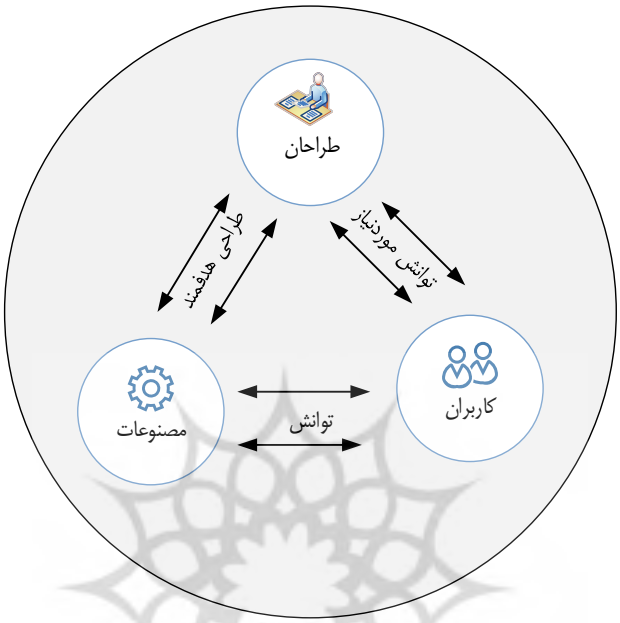
اخیرا نظریهٔ توانش[1]مورد توجه بسیاری از پژوهشگران سیستم‌های اطلاعاتی قرار گرفته است [۱۱،۲۷،۳۶،۴]. این نظریه، از دیدگاه گیبسون، مفهوم توانش به هر خصیصه‌ای اشاره می‌کند (چه خوب یا بد) که محیط آن را به انسان اعطاء می‌کند و بر مکمل بودن انسان و محیط دلالت دارد. برای مثال، کوه، جنگل، صحرا و دریا موجب رفتارهای متفاوت گونه‌های جانوری می‌شود. بنابراین، این فرایند ادراکی را می‌توان همچون یک سیستم تعاملی نگریست که در آن دو موجودیت وجود دارد: حیوان و محیط زیست. توانش نیز یک مفهوم رابطه‌ای است که به هر دو موجودیت بستگی دارد. نخستین ویژگی مهم توانش که خاصیت رابطه‌ای آن را تعریف می‌کند، مکملیت است. دلالت این ویژگی در پژوهش سیستم اطلاعاتی بسیار مهم است. به عبارت دیگر، روابط در هم تنیدهٔ بین افراد (طراحان و کاربران) و اشیاء خاص در محیط‌شان (مصنوعات) موضوع مورد علاقهٔ پژوهشگران این حوزه است. گیبسون با بیان اینکه توانش می‌تواند خوب یا بد باشد، همچنین در مورد قطبیت[2] توانش بحث کرده است. توانش مثبت به طور بالقوه برای کاربر مفید هستند، در حالی که توانش منفی به طور بالقوه مضر هستند. به عبارت دیگر، طراحان باید سیستم‌هایی با توانش مثبت مطلوب بدون معرفی توانش منفی نامطلوب طراحی کنند. افزون بر این، سیستم‌ها باید از توانش متعددی برخوردار باشند. برای مثال، گیبسون مطرح می‌کند که که آب امکان نوشیدن، ریختن، شستشو و استحمام را فراهم می‌کند [۲۲]. به طور ویژه، در پژوهش سیستم‌های اطلاعاتی مدیریت، بیشتر پژوهشگران، توانش را رفتارهای بالقوه‌ای مفهوم‌سازی کرده‌اند که از روابط بین مصنوع و کاربر شکل می‌گیرند. این نوع مفهوم‌سازی از توانش هم ویژگی رابطه‌ای (یعنی، توانش رابطهٔ بین مصنوع و کاربر است) و هم ویژگی ذاتی (یعنی، توانش، یک ویژگی سیستم است) را در بر می‌گیرد [۲۷].

---

[1] Theory of affordances

[2] Polarity

مطابق شکل ۱، دیدگاه نظری، از مدل طراحی رابطه‌ای برای پرداختن به تعاملات میان طراحان، مصنوعات و کاربران استفاده می‌کند. مدل سه نوع تعامل را نشان می‌دهد: نخست، ماهیت تعامل بین طراحان و کاربران به گونه‌ای است که طراحان نیاز دارند مجموعه‌ای از توانش‌های مورد نیاز کاربر را شناسایی کنند. دوم، ماهیت تعامل بین طراحان و مصنوعات به گونه‌ای است که طراحان ویژگی‌های مصنوع را توسعه می‌دهند که توانش مورد نیاز را فراهم کند. سوم، ماهیت تعامل بین کاربران و مصنوعات به گونه‌ای است که توانش مصنوع نحوهٔ استفاده از آن را برای کاربران تعیین می‌کند. این مدل همراستا با رویکرد پژوهش علم طراحی است که در آن مطرح می‌شود هنگام طراحی یک سیستم اطلاعاتی در عمل، پژوهشگران سیستم اطلاعاتی به شناسایی همهٔ توانش‌های موردنیاز و طراحی هدفمندشان نیاز دارند که با کاربران ویژه در موارد استفاده خاص قابل تحقق است [۲۷].

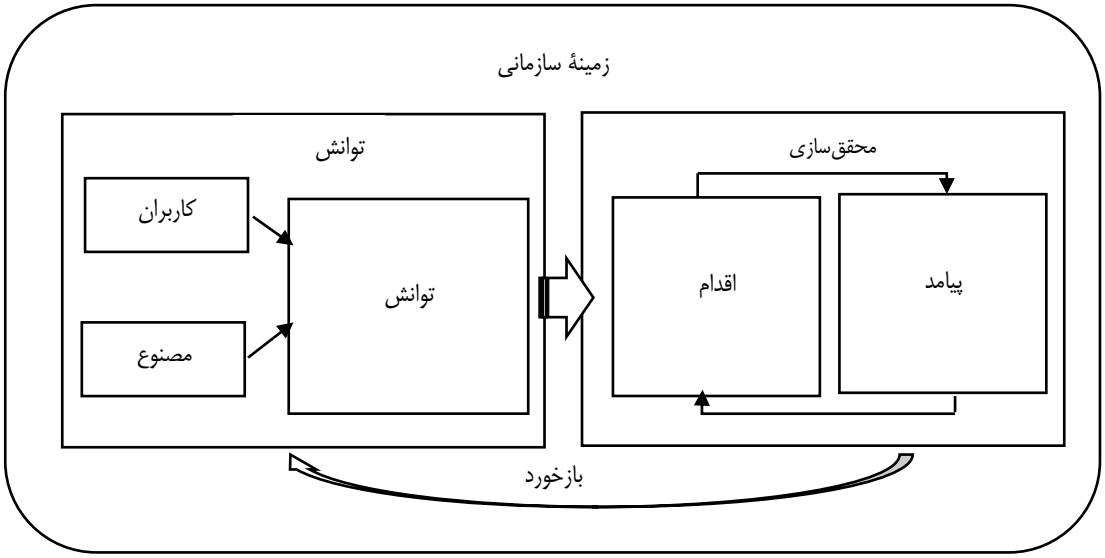


شکل ۱. تعامل وابسته به توانش در سیستم‌های طراح- مصنوع- کاربر [۲۷]

از این‌رو،کاربست نظریهٔ توانش در حوزهٔ پژوهش‌های طراحی و پیاده‌سازی سیستم‌های مبتنی بر هوش مصنوعی، از این جهت مهم است که به بررسی همزمان جنبه‌های اجتماعی و فنی در طراحی سیستم می‌پرازد [۱۱]. مزیت این رویکرد در طراحی سیستم‌های مبتنی بر هوش مصنوعی با قابلیت تبیین، این است که هرچه کاربران تعامل بیشتری با طراحان در فرایند توسعهٔ سیستم داشته باشند، نیاز به تبیین از سوی سیستم کمتر احساس خواهد شد؛ زیرا کاربران در ساختن مدل ذهنی سیستم نقش دارند و به راحتی رفتار آن را می‌آموزند [۲۳]. بر همین اساس، از نظریهٔ توانش در بنیان نظری پژوهش بهره گرفته می‌شود. برای این منظور، از چارچوب راهنمای توانش-محقق‌سازی[1] استفاده می‌شود. آنچه که کاربرد این چارچوب را برای پژوهش طراحی مبتنی بر اقدام متناسب می‌سازد، بررسی توانش‌ها و محقق‌سازی آنها بصورت جداگانه است (شکل۲).

پژوهش‌های قبلی، بطور ضمنی فرض می‌کردند که اگر کاربران متناسب وجود داشته باشند، توانش در عمل محقق می‌شود. بر همین اساس، محقق‌سازی توانش را می‌توان از اقدامات هدفمندکاربران برای رسیدن به یک خروجی از طریق استفاده از فناوری تعریف نمود. همچنین، اقدامات و پیامدها، بازخوردی را برای توانش فراهم می‌آورند؛ این بازخوردها منجر به حذف توانش‌های منفی و اصلاح پیامدهای غیرمنتظرهٔ سیستم می‌شوند. افزون بر این، محقق‌سازی توانش‌های اولیه، بر دانش‌کاربران می‌افزاید و آنها را قادر می‌سازد تا از فناوری به شیوه‌ای سودمندتر استفاده کنند. مهم‌تر از همه، نقش زمینهٔ سازمانی در محقق‌سازی توانش بسیار مهم است، زیرا یک زمینهٔ سازمانی مناسب، تسهیلگر محقق‌سازی توانش است، در حالیکه زمینهٔ نامناسب موجب محدود شدن محقق‌سازی می‌شود [۱۱]

[1] Affordance-actualization theory



شکل ۲. چارچوب توانش- محقق‌سازی[۱۱]

پیشینهٔ پژوهش

گرچه تعامل انسان- هوش مصنوعی در حال تبدیل شدن به جریان غالب پژوهش در قلمرو تصمیم‌گیری سازمانی است [۱۸،۳۴،۳۷،۲]. اما بطور کلی، دانش توصیفی و تجویزی کمی در مورد اصول طراحی و پیاده‌سازی سیستم‌های هوش مصنوعی با قابلیت تبیین در تعامل انسان-هوش مصنوعی، برای تقویت هوشمندی وجود دارد. اسکوندرورد، جاریتسما، نرینکس و ون دن بوش[۱][۳۱] یک روش‌شناسی ارائه کردند که مشتمل بر سه مرحلهٔ تحلیل قلمرو، به منظورتعریف مفاهیم و زمینهٔ تبیین؛ استخراج و ارزیابی الزامات، به منظور صورت‌بندی موارد استفاده و نیازمندی‌های تبیین و سپس طراحی و ارزیابی تعاملی چندوجهی، برای ایجاد فهرستی از الگوهای طراحی برای تبیین است. آنها همچنین استدلال کردند برای توسعه موفقیت‌آمیز سیستم‌های هوش مصنوعی با قابلیت تبیین، نیاز به تیمی متشکل از کاربران نهایی و متخصصانی از هوش مصنوعی، عوامل انسانی و طراحی تعامل انسان- کامپیوتر وجود دارد. ولف[۲][۴۰] رویکرد مبتنی بر سناریو را پیشنهاد نمود که در شناسایی نیازهای کاربران مورد استفاده قرار می‌گیرد و استدلال نمود که این رویکرد می‌تواند مبنایی برای توسعهٔ بیشتر تبیین‌ها باشد.

افزون بر این، میلر[۳] [۲۳] در مطالعه‌ای نشان داد که بیشتر پژوهش‌های انجام شده در زمینهٔ هوش مصنوعی با قابلیت تبیین، صرفا از شهود پژوهشگران در پیکربندی ویژگی‌های یک "تبیین خوب" استفاده می‌کنند. درحالیکه، بینش‌های ارزشمندی از پژوهش‌های حوزهٔ فلسفه، روانشناسی و علوم شناختی در مورد نحوهٔ تعریف، ایجاد، انتخاب، ارزیابی و ارائهٔ تبیین توسط افراد وجود دارد که استدلال می‌کنند افراد از سوءگیری‌های شناختی و انتظارات اجتماعی خاصی برای فرآیند تبیین استفاده می‌کنند. گیدوتی[۴] و همکاران[۱۵] نیز نشان دادند، گرچه رویکردهای بسیاری برای تبیین جعبه سیاه پیشنهاد شده است، اما مهمترین مسأله‌ای که تاکنون بدون پاسخ مانده، عدم توافق در مورد تعریف تبیین است. به عبارت دیگر، برخی از پژوهشگران، تبیین را مجموعه‌ای از قواعد، گروهی نیز به درخت تصمیم و سایرین نیز یک نمونهٔ اولیه تعریف کرده‌اند. اما ساختار کلی و مشترکی برای تعریف تبیین ارائه نشده است که در آن مشخص شود که تبیین باید متضمن چه ویژگی‌هایی باشد. کولزا، بارنت، وانگ و استامپ[۵][۲۰] در پژوهشی با هدف تأثیرگذاری کاربران نهایی بر پیش‌بینی‌های سیستم‌های یادگیری ماشین، رویکرد اشکال‌زدایی تبیینی[۶] را مطرح کردند. نتایج مطالعهٔ آنها نشان داد، این رویکرد در مقایسه با برچسب- گذاری جعبه سیاه، رویکردی قابل کنترل و رضایت‌بخش است که پیامد آن بهره‌مندی کاربران نهایی معمولی از عامل‌های هوشمند است که به آنها وابسته هستند.

---

[1] Schoonderwoerd, Jorritsma, Neerincx and van den Bosch
[2] Wolf
[3] Miller
[4] Guidotti
[5] Kulesza, Burnett, Wong and Stumpf
[6] Explanatory Debugging

پیشینهٔ پژوهش در عرصهٔ حقوقی و بطور خاص در قلمرو قانون‌گذاری بر طبقه‌بندی متون قانونی به منظور شناسایی قلمرو موضوعی قوانین، شناسایی احکام، شناسه‌کاوی و پیش‌بینی تصمیم دادگاه تمرکز کرده‌اند. برای مثال، چن[1] و همکاران[۶]، با تمرکز بر طبقه‌بندی متون حقوقی ایالت متحده، اثربخشی تکنیک‌های مختلف یادگیری ماشین را با هم مقایسه کرده است. کاستانو[2] و همکاران[۵] نیز مطرح کردند به منظور غلبه بر حجم فزاینده، پیچیدگی و صورت‌بندی متون قانونی برای غلبه بر محدودیت‌های وظایف دستی استخراج اصطلاحات قوانین، به تکنیک‌های استخراج خودکار اطلاحات موردنیاز است. سیل و آبیشک[3][۳۳] نیز، مدل طبقه‌بندی قضاوت مربوط به"قانون خشونت خانگی علیه زنان" را ارائه کردند. مدل پیشنهادی آنها به تصمیم‌گیری متخصصان حقوقی در مورد حل و فصل پرونده‌ها به شیوه‌ای ساختاریافته از طریق طبقه‌بندی و شناسایی مجموعه داده‌های حقوقی برای شناسایی فرد متهم کمک می‌کند. با این وجود، رهبری و شعبانپور[۲۹] چالش‌های بکارگیری هوش مصنوعی را قاضی دادگاه مورد مطالعه قرار داده‌اند و دریافتند که هوش مصنوعی در زمینهٔ استدلال حقوقی، رعایت بی‌طرفی و پذیرش عمومی با چالش‌های جدی مواجه است. بنابراین، هوش مصنوعی با قابلیت‌های فعلی خود نمی‌تواند جایگزین کاملی برای منصب قضاوت تلقی شود و می‌تواند ابزاری برای کمک به رسیدگی و حل و فصل سریع‌تر و دقیق‌تر اختلافات باشد. بطور خلاصه، پژوهش‌های موجود بر متون حقوقی بین المللی همانند کشورهای اروپایی، استرالیا، آمریکا، چین و هند متمرکز شده‌اند و توجه کمی به طبقه‌بندی متن برای متون حقوقی ایران شده است.

این پژوهش‌ها، رهنمودهایی را به پژوهشگران و دست‌اندرکاران در مورد طراحی سیستم‌های با قابلیت تبیین ارائه می‌کنند که نیازهای کاربران را برآورده می‌کنند. همچنین برخی از آنها کاربرد نظام‌مند طراحی انسان محور را در فرایند توسعه و طراحی سیستم‌های هوش مصنوعی با قابلیت تبیین به منظور درک بهتر الزامات فنی و اجتماعی تعامل انسان–هوش مصنوعی توصیف کرده‌اند. با این وجود آنچه که کاربرد این روش‌شناسی‌ها را محدود می‌سازد، این است که آنها در زمینهٔ کشورهای پیشرفته از نظر فناوری انجام شده‌اند و بطور ضمنی فرض می‌کنند، اگر کاربران مناسب وجود داشته باشند، آنگاه توانش در عمل محقق می‌شود. درحالیکه در مبانی نظری استدلال شد که توانش‌ها و محقق‌سازی آنها باید بصورت جداگانه مورد بررسی قرار گیرند.

۳. روش‌شناسی پژوهش

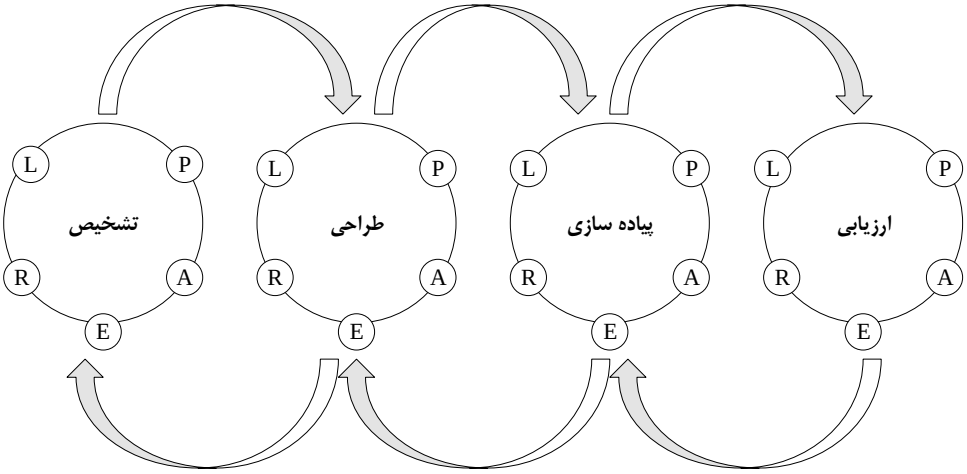
با توجه به اینکه پژوهش علاقه‌مند به بررسی پیامدهای عملی تعامل انسان– هوش مصنوعی در فرایند تنقیح قوانین و مقررات قضایی بوده لذا فلسفهٔ این پژوهش مبتنی بر عمل‌گرایی است. از این جهت، از نظر طرح پژوهش، در زمرهٔ روش پژوهش آمیخته اکتشافی است. در بخش کیفی، از روش پژوهش طراحی مبتنی بر اقدام برای توسعه و ارزیابی اصول طراحی سیستم هوش مصنوعی با قابلیت تبیین استفاده می‌شود. توجیه استفاده از این استراتژی در پژوهش، فراهم کردن سازوکار روش‌شناختی، برای مطالعهٔ ایجاد یک مصنوع فناورانه مبتنی بر نظریه در محیط‌های سازمانی است.

در این پژوهش از رویکرد پژوهش طراحی مبتنی براقدام بسط‌یافته توسط ملارکی[4] و همکاران[۲۴] پیروی می‌شود. این رویکرد، فرایند ساخت– مداخله– ارزیابی (سما) برگرفته از پژوهش بنیادین ساین[5] و همکاران[۳۲] را به چرخه‌های تشخیص، طراحی، پیاده‌سازی و ارزیابی تفکیک می‌کند. همانگونه که در شکل ۳ نمایش داده شده است، هر یک از چرخه‌های این مدل فرایندی، شامل فعالیت‌های تدوین مسأله[6]، ایجاد مصنوع[7]، ارزیابی، بازنگری[8] و یادگیری است. افزون بر این، ملارکی و همکاران[۲۴] اصل جدیدی از انتزاع را اضافه کردند که نشان می‌دهد پژوهشگران می‌توانند از طریق سطح مناسبی از انتزاع در هر یک از چرخه‌ها، مصنوعات مختلفی را معرفی کنند. برای نمونه، خروجی چرخه تشخیص، الزامات طراحی، خروجی چرخه طراحی ممکن است معماری یا مجموعه‌ای از اصول طراحی باشد. بنابراین از آنجایی که این رویکرد، نظریه را با عمل و اندیشیدن را با انجام دادن، پیوند می‌زند، برای پژوهش توسعه و ارزیابی سیستم‌های تعامل انسان– هوش مصنوعی در تصمیم‌گیری سازمانی بسیار مناسب است.

از آنجا که طراحی هوش مصنوعی با قابلیت تبیین در تعامل انسان– هوش مصنوعی، نیازمند بررسی و تشخیص قلمرو مسأله (دانش توصیفی) و ارزیابی طبقه‌ای از راه‌حل‌های فناورانه (دانش تجویزی) است، این پژوهش مطابق با چرخهٔ تشخیص (جدول ۱) است.

---

[1] Chen
[2] Castano
[3] Sil & Abhishek
[4] Mullarkey
[5] Sein
[6] Problem formulation (P)
[7] Artefact creation (A)
[8] Reflection (R)



شکل ۳ . مراحل مدل فرایندی پژوهش طراحی مبتنی بر اقدام [۲۴]

هدف این مرحله، تجزیه و تحلیل اهمیت قلمرو مسأله و تناسب طبقه‌ای از راه‌حل‌های فناورانه برای پژوهش و عمل و توافق متقابل میان تیم پژوهش و دست‌اندرکاران است. تأکید بر تشخیص در بیشتر موارد، پژوهشگران و دست‌اندرکاران را به سمت اختصاص زمان بیشتر برای مرحلهٔ آغازین پژوهش طراحی مبتنی بر اقدام هدایت می‌کند. برای انجام این کار، از رویکرد استدلال از طریق بهترین تبیین برای توسعهٔ نظریه پیروی می‌شود که بجای مسیر نظریه به داده (همانند استناج) و مسیر داده به نظریه (همانند استقرا)، یک مسیر رفت و برگشتی است که در واقع استقراء و استنتاج را با هم ترکیب می‌کند.

مصنوعاتی که در این چرخه، توسعه و ارزیابی می‌شوند، می‌توانند در قالب تعریف نیازمندی‌ها، مشخصات فنی و مفهوم‌سازی قلمرو مسأله و راه‌حل ارائه شوند [۲۴]. فعالیت‌های اصلی که در این چرخه انجام می‌شود در جدول ۱ نمایش داده شده است.

جدول ۱. فعالیت‌های چرخهٔ تشخیص [۲۷]

نام فعالیت	اقدامات
تدوین مسأله	- شناسایی فرصت‌های پژوهش - تدوین مسألهٔ پژوهش - مرور پیشینهٔ پژوهش برای درک سیستم‌های هوش مصنوعی با قابلیت تبیین (برای مثال، فضای مسأله و فضای راه حل)
ایجاد مصنوع	- شناسایی نیازمندی‌های اولیهٔ تحلیل سیستم - پالایش نیازمندی‌ها
ارزیابی	- ارزیابی نیازمندی‌ها در تطابق با خواسته‌های کاربر
بازنگری	- بازنگری چرخهٔ تشخیص برای پالایش نیازمندی‌های اولیه
یادگیری	- انتزاعی‌سازی یادگیری در مجموعه‌ای کلی از الزامات - اشتراک خروجی با دست‌اندرکاران

با توجه به اینکه در این پژوهش از چارچوب نظری توانش- محقق‌سازی، برای شناسایی الزامات سیستم‌های با قابلیت تبیین استفاده می‌شود لذا مسألهٔ اصلی پژوهش به سه سؤال فرعی تقسیم می‌شوند:

- توانش مورد نیاز کاربران در سیستم‌های با قابلیت تبیین شامل چه مؤلفه‌هایی است؟
- سیستم‌های با قابلیت تبیین از طریق چه مؤلفه‌هایی توانش مورد نیاز کاربران را محقق می‌کنند؟
- پیامد سیستم‌های با قابلیت تبیین در همکاری انسان- هوش مصنوعی چیست؟

در بخش کمّی، در گام نخست، از طریق تکنیک دلفی فازی، مؤلفه‌های بدست آمده از بخش کیفی، غربال‌گری و اعتبارسنجی شد [۱۹] و سپس بر اساس مدل‌سازی ساختاری- تفسیری، روابط بین مؤلفه‌ها با در نظرگفتن چارچوب نظری تعیین شد و همزمان سطح‌بندی و تعیین قدرت نفوذ و میزان وابستگی مؤلفه‌ها مورد بررسی قرار گرفت[۳۸].

بستر پژوهش. به منظور اجرای پژوهش طراحی مبتنی بر اقدام، ادارهٔ کل تنقیح قوانین و مقررات در معاونت حقوقی قوهٔ قضائیه زمینهٔ پژوهش انتخاب شد. این واحد سازمانی نقش بسیار مهمی را در تنقیح قوانین و مقررات ایفا می‌کند. از این جهت، مهمترین وظایف این واحد سازمانی عبارتند از: آسیب‌شناسی و ارزیابی مستمر قوانین و مقررات قضایی و اقدام لازم در خصوص تنقیح آنها؛ تهیهٔ درخت‌وارهٔ قوانین و مقررات، تبیین تعداد و دامنهٔ موضوعات تنقیحی و اصول، ضوابط و فنون ناظر بر تنقیح شکلی و ماهوی قوانین و مقررات و تدوین قوانین جامع، تهیه پیش‌نویس لوایح و مواد تنقیحی، شناسایی مواد نیازمند وضع قوانین و مقررات یا بازنگری آن و اعلام به مراجع مرتبط و تنقیح آیین‌نامه‌ها، دستورالعمل‌ها، بخشنامه‌ها و سایر مصوبات مربوط به قوهٔ قضائیه. بر همین اساس، با توجه به انبوه قوانین و مقررات، طراحی و پیاده‌سازی سیستم‌های اطلاعاتی در گردآوری، پالایش و تهیهٔ پیش‌نویس لوایح قانونی به منظور رفع ابهام، تعارض، تداخل و تکرار ضروری است.

در این پژوهش، دو واحد ادارهٔ کل تنقیح قوانین و مقررات و مرکز آمار و فناوری اطلاعات قوهٔ قضائیه، دو ذینفع اصلی این پروژه هستند. ادارهٔ کل تنقیح قوانین و مقررات، مسئولیت پیاده‌سازی فناوری‌های نوین در انجام فعالیت تنقیح را بر عهده دارد و مرکز آمار و فناوری اطلاعات، مسئولیت اجرای این پروژه را عهده‌دار است. بنابراین در انجام این پژوهش، پژوهشگران و نمایندگان این دو ذینفع اصلی، تیم پژوهش طراحی مبتنی بر اقدام را تشکیل دادند.

دست‌اندرکاران در تیم پژوهش دو نقش کلیدی را ایفا کردند. گروه نخست دست‌اندرکاران از ادارهٔ تنقیح مقررات و قوانین قضائی، فرایند گردآوری داده، پالایش و تهیهٔ پیش‌نویس لوایح تنقیحی را تشریح نمودند که به درک پژوهشگران از زمینهٔ این پژوهش کمک کرد. افزون بر این، آنها انتظارات خود از سیستم را ابراز کردند که در انجام اجرای وظایف تنقیح قوانین به آنها کمک می‌کند. گروه دوم دست‌اندرکان از مرکز آمار و فناوری اطلاعات، نقش بسیار مهمی در ارائهٔ اطلاعات در مورد سیستم‌های موجود و چگونگی تبیین آنها ارائه نمودند. افزون بر این، آنها توصیه‌هایی در مورد تقویت هوشمندی کاربران از طریق سیستم‌های موجود ارائه کردند. جدول ۲ خلاصه‌ای از مسئولیت‌های تیم پژوهش طراحی مبتنی بر اقدام در چرخهٔ تشخیص را ارائه می‌کند که در مجموع ۱۵ نفر هستند.

جدول ۲. مسئولیت‌های اعضای تیم پژوهش

اعضای تیم	تعداد	مسئولیت‌ها
ادارهٔ کل تنقیح و مقررات قضائی	۶	- ارائهٔ اطلاعات در مورد چگونگی ارزیابی مستمر قوانین و مقررات - ارائهٔ توصیه‌هایی در مورد چگونگی تهیه پیش‌نویس لوایح و مواد تنقیحی
مرکز آمار و فناوری اطلاعات	۶	- ارائه اطلاعات در مورد عملکرد سیستم‌های موجود - ارائهٔ توصیه‌هایی در مورد ساختار تصمیم‌گیری در فرایند تنقیح
اساتید راهنما و مشاور	۳	- شناسایی فرصت پژوهشی و صورت‌بندی مسأله - شناسایی الزامات هوش مصنوعی با قابلیت تبیین

به منظور طراحی سیستم هوش مصنوعی با قابلیت تبیین، درک فضای مسأله و راه‌حل‌های ممکن ضروری است. بنابراین، داده‌های این پژوهش با بکارگیری تکنیک چندسویه‌سازی[1] از منابع مختلف گردآوری شده‌اند. برای انجام این کار، در گام نخست با بررسی مسألهٔ پژوهش در قلمرو گستردهٔ تعامل انسان- هوش مصنوعی، پیکره غنی پژوهش درمورد نحوهٔ تعریف، ایجاد، انتخاب، ارزیابی و ارائهٔ تبیین شناسایی شد. در این مرحله فهرستی از مضامین تهیه شد و سؤالات کلیدی مصاحبهٔ نیمه ساختار یافته بر اساس آنها تنظیم شد (چه نوع اطلاعاتی برای مدل ذهنی کاربران مناسب‌تر است؟ وضوح و کامل بودن اطلاعات چه تأثیری بر مدل ذهنی کاربران دارد؟ و چگونه مزیت‌های ادراک شده کاربران نسبت به عامل هوشمند تغییر می‌کند؟). این گام مطابق با فعالیت ایجاد مصنوع در چرخهٔ تشخیص است. در گام دوم، ویژگی‌های مورد انتظارکاربران برای تعامل با سیستم بر اساس مصاحبه با ۱۵ نفر از ذینفعان در زمینهٔ پژوهش مفهوم‌سازی

[1] triangulation

شد. دلیل انتخاب این افراد، تجربهٔ زیستهٔ آنها در فرایند سیاستگذاری و طراحی و پیاده‌سازی سامانه‌های موجود در بستر پژوهش بوده است. در نهایت، پس از ارائهٔ نسخهٔ اولیهٔ سیستم، مشکلات آن پس از گفتگو با ذینفعان برطرف شد. در این گام، از روش مشاهدهٔ مشارکتی برای گردآوری داده‌های تکمیلی استفاده شد. این فرایند در سه مرحله و مدت سه ماه طول کشید، بطوریکه پس از پایان هر مرحله، پژوهشگران با مراجعه به پیشینهٔ پژوهش زوایای پنهان فضای مسأله را آشکار نمودند. همچنین، در این مرحله از پرسشنامه مقایسات زوجی برای غربال‌گری و اعتبار سنجی مؤلفه‌ها استفاده شده است. به عبارت دیگر، این گام مطابق با فعالیت‌های ارزیابی و بازنگری در چرخهٔ تشخیص است. در گام پایانی که مطابق با فعالیت انتزاعی‌سازی یادگیری در مجموعه‌ای کلی از الزامات است، پرسشنامه‌ای با رویکرد مدل‌سازی ساختاری–تفسیری طراحی شد و در اختیار مشارکت‌کنندگان قرار گرفت تا بر اساس آن تأثیر هریک از عوامل را بر دیگری بر اساس مقیاس ۴ درجه‌ای کاملا مؤثر است (۳)، مؤثر است (۲)، تأثیر ناچیزی دارد (۱) و بی‌تأثیر است (۰)، مشخص کنند.

تجزیه و تحلیل داده‌ها دو مرحله دارد: در مرحلهٔ اول، با کدگذاری اولیهٔ داده‌ها، مجموعه‌ای از مفاهیم گرداوری شد. این کار با حفظ انسجام مفاهیم انجام و و به کمک نرم افزار ان ویوو[1] نسخه ۸ انجام شد. با انجام این کار، خلاصه‌ای جامع از مفاهیم مرتبه اول گردآوری شد. در مرحلهٔ دوم، با بررسی شباهت‌ها و تفاوت‌ها میان مفاهیم گردآوری شده، در قالب مجموعه‌ای از مؤلفه‌ها سازماندهی گشت. درمرحلهٔ پایانی، با پالایش مستمر مفاهیم، مؤلفه‌های استخراج شده در ابعاد نظری تجمیع شدند. همچنین با ترکیب مفاهیم، مؤلفه‌ها و ابعاد، ساختار داده استخراج شد. در مرحلهٔ اول تجزیه و تحلیل داده‌ها، به دلیل اهمیت داده‌های تجربی از روش استقرایی (رویکرد جزء به کل) استفاده شد که در آن مفاهیم، روابط میان آنها و الگوها درطول گردآوری داده‌ها پدیدار شدند (شکل۵). در این پژوهش، برای فازی‌سازی دیدگاه خبرگان، از اعداد فازی مثلثی با طیف فازی ۷ درجه استفاده شد. برای تجمیع دیدگاه خبرگان از فرمول میانگین فازی و برای فازی زدایی نیز روش مینکوفسکی مورد استفاده قرار گرفت. معیار اجماع خبرگان در مورد مؤلفه‌ها، اختلاف میانگین فازی کمتر از ۰/۱ است [۱۹].

در مرحلهٔ دوم، از چارچوب نظری پژوهش برای توسعهٔ نظریه استفاده شد. بنابراین، تحلیل نظریه محور از طریق تطبیق تفسیری ساختار داده با چارچوب توانش– محقق‌سازی انجام شد و در نتیجه تناقض و تکرار میان و بین عناصر برطرف شد که منطبق با رویکرد استدلال از طریق بهترین تبیین است. در این مرحله، روش مدل‌سازی ساختاری– تفسیری نیز به منظور مدل‌سازی مجموعه‌ای کلی از الزامات استفاده شد. برای انجام این کار، روابط نظام‌مند بین مؤلفه‌ها با لحاظ نمودن نظر خبرگان تعیین شد و همزمان سطح‌بندی الزامات طراحی سیستم‌های هوش مصنوعی با قابلیت تبیین و تحلیل قدرت نفوذ–وابستگی با استفاده از نرم افزار میک–مک[2] انجام شد. در نهایت، مدل نهایی که شامل روابط معنادار عناصر هر سطح بر عناصر سطح زیرین و همچنین روابط درونی معنادار عناصر هر سطح، با ارائه میزان وابستگی و قدرت نفوذ متغیرها ارائه شد.

اعتبارسنجی مدل. به طور معمول برای ارزیابی پژوهش طراحی مبتنی بر اقدام از اعتبار صوری استفاده می‌شود[۲۱]. بر همین اساس، چرخهٔ تشخیص از طریق مصاحبه با ذینفعان سه بار تکرار شد. بطوری که پس از هر تکرار، پژوهشگران با مراجعه به پیشینهٔ پژوهش، زوایای پنهانی از مسألهٔ پژوهش را آشکار نمودند و این فرایند تا تکرار سوم ادامه یافت و در این مرحله پژوهشگران به شناخت کاملی از فضای مسأله دست یافتند. در روش دلفی فازی، پرسشنامه در دو بازهٔ زمانی مختلف توزیع و جمع‌آوری می‌شود. از این جهت، می‌توان از نرخ ناسازگاری و اختلاف میانگین فازی‌زدایی دو مرحله برای سنجش و بررسی پایایی استفاده کرد. بدین صورت، در این رویکرد با بررسی اختلاف فازی‌زدایی و نرخ ناسازگاری، اگر این میزان از ۰/۱ کمتر باشد، می‌توان ادعا نمود که پرسشنامه از پایایی برخوردار است. با توجه به اینکه در این پژوهش، نرخ ناسازگاری و اختلاف میانگین فازی‌زدایی در دو مرحله برای پرسشنامه کمتر از ۰/۱ است، می‌توان نتیجه گرفت که پرسشنامه نخست از پایایی لازم برخوردار است. افزون بر این، برای پایایی پرسشنامه شدت اثرات نیز سعی شد پاسخ خبرگان کنترل شود تا همهٔ آنها نظرات یکپارچه‌ای در مورد تأثیر عوامل بر یکدیگر داشته باشند.

۴. تحلیل داده‌ها و یافته‌های پژوهش

با توجه به این که تنقیح قوانین و مقررات فرایندی زمینه محور است، دست‌اندرکاران باید قوانین و مقررات موجود را به منظور شناسایی و تفکیک قوانین معتبر از غیرمعتبر و همچنین رفع موارد ابهام، اجمال، تکرار، تعارض و نقص پالایش نمایند. برای انجام این کار، سامانه‌های موجود امکان یادآوری را از طریق تفکیک و طبقه‌بندی موضوعی، شناسایی و تهیهٔ منابع و استخراج و جمع‌آوری مطالب را فراهم نمودند.

[1] Nvivo 8

[2] MICMAC

با این وجود، فاقد امکان درک از طریق معرفی موارد ابهام، اجمال، تکرار، تعارض و نقص بودند. از این جهت، به منظور شناسایی الزامات طراحی سیستم‌های هوش مصنوعی با قابلیت تبیین در تعامل انسان–هوش مصنوعی، مصاحبه‌های نیمه‌ساختاریافته‌ای با اعضای تیم پژوهش ترتیب داده شد. جدول ۳، ویژگی‌های جمعیت‌شناسی مشارکت‌کنندگان پژوهش را نشان می‌دهد.

جدول ۳. توزیع داده‌های جمعیت‌شناختی مشارکت‌کنندگان

اعضای تیم	جنسیت	تحصیلات	سابقه فعالیت	سمت
اداره کل تنقیح و مقررات قضائی	مرد	دکتری حقوق عمومی	۱۰ سال	نماینده کمیسیون
	مرد	دکتری حقوق عمومی	۱۰ سال	نماینده کمیسیون
	مرد	دکتری حقوق جزا و جرم شناسی	۱۵ سال	نماینده کمیسیون
	مرد	دکتری روانشناسی بالینی	۸ سال	نماینده کمیسیون
	مرد	دکتری حقوق جزاء و جرم شناسی	۷ سال	نماینده کمیسیون
	مرد	دکتری حقوق جزاء و جرم شناسی	۸ سال	نماینده کمیسیون
مرکز آمار و فناوری اطلاعات	مرد	دکتری مهندسی کامپیوتر	۴ سال	رئیس
	مرد	کارشناس ارشد فقه و حقوق	۵ سال	قائم مقام
	مرد	کارشناس ارشد فناوری اطلاعات	۴ سال	معاون
	مرد	کارشناس ارشد فقه و اصول	۵ سال	معاون
	مرد	کارشناس ارشد فناوری اطلاعات	۳ سال	مدیر
	مرد	دکتری مهندسی کامپیوتر	سال	مدیر
اساتید راهنما و مشاور	مرد	دکتری سیستم های اطلاعات مدیریت	۱۷ سال	استاد دانشگاه
	مرد	دکتری علم اطلاعات و دانش شناسی	۱۸سال	استاد دانشگاه
	مرد	دکتری مدیریت دانش	۱۰ سال	استاد دانشگاه

با توجه به چارچوب نظری پژوهش، سؤالات مصاحبه در دو بعد توانش و محقق‌سازی طراحی شده‌اند. سپس، داده‌های گردآوری شده از طریق تحلیل مضمون کدگذاری شده‌اند. مضامین استخراج شده از فرایند مصاحبه در جدول ۴ نمایش داده شده است.

جدول ۴. مضامین استخراج شده از فرایند مصاحبه

ردیف	عناصر اطلاعاتی	فراوانی در مصاحبه
۱	قابلیت آگاهی‌بخشی برای تصمیم‌گیری	۲
۲	رسیدن به نتایج یکسان برای قوانین مشابه	۱
۳	توضیحات یکسان برای قوانین مشابه	۱
٤	کمک به تصمیم‌گیری خوب	۸
٥	کمک به تصمیم‌گیری سریع	۷
٦	قابلیت ارائه بازخورد به سیستم	۵
۷	نمایش زنجیرهٔ دقیق علی رد و پذیرش قوانین	۲
۸	بیان دلایل و توجیهات قانع کننده	۹
۹	قابلیت اشکال‌زدایی و بازرسی	۶
۱۰	ارائه دلایل نهفته در شناسایی موارد قانونی متعارض	۳
۱۱	شفافیت در فرایند پالایش	۶
۱۲	نمایش فرایند گردآوری داده	۴
۱۳	قابلیت تفکیک مواد قانونی معتبر از غیر معتبر	۵
۱٤	قابلیت تعامل و تبادل دانش	۹
۱۵	دقت زمینه‌ای در شناسایی مواد معتبر از غیرمعتبر	۱

ردیف	عناصر اطلاعاتی	فراوانی در مصاحبه
۱۶	انسجام از نظر ساختار جمله و دستور زبان	۴
۱۷	درک کافی از قوانین و مقررات	۲
۱۸	توجیه اشکالات در قوانین به شیوه و سبکی قابل فهم	۱
۱۹	قابلیت استناد به قوانین بالادستی	۶
۲۰	نمایش قلمرو و محدودهٔ بررسی قوانین	۱
۲۱	ارائه اطلاعات جامع در مورد قوانین و مقررات	۴
۲۲	نمایش جریان کار رفتار سیستم	۳
۲۳	نمایش میزان احتمال معتبر بودن و غیر معتبربودن قانون	۱
۲۴	نمایش قوت‌ها و محدودیت‌های قوانین	۱

مطابق جدول ۴، در فرایند گردآوری داده‌ها از طریق مصاحبه با مشارکت‌کنندگان در تکرارهای چرخهٔ تشخیص ۲۴ مضمون استخراج شده است. در تحلیل مرتبهٔ اول، پژوهشگران تلاش کردند تا به مضامین ارائه شده از مصاحبه‌شوندگان پایبند باشند. این مضامین، خروجی مصاحبه با ۱۵ نفر از اعضای تیم پروژه در فاز تشخیص است. پس از اجرای مصاحبه، پژوهشگران با طرح این پرسش که تا چه اندازه مضامین نوپدید به شناخت الزامات طراحی سیستم مبتنی بر همکاری انسان- هوش مصنوعی کمک کرده‌اند، عناصر اطلاعاتی را مورد بازبینی قرار دادند. سپس با اشتراک‌گذاری آن با تیم پروژه، نیازمندی‌های اولیه پالایش شد و توافق نهایی در مورد مجموعه الزامات طراحی سیستم صورت گرفت. به نوبهٔ خود، این مرحله منطبق با اشباع نظری در پژوهش‌های کیفی است. در ادامه با بررسی شباهت‌ها و تفاوت‌های میان عناصر اطلاعاتی، عناصر مشابه در مقوله‌های مرتبهٔ دوم بررسی شدند. این مقوله‌ها عبارتند از: قابلیت درک (عناصر ۱ ،۱۶و ۱۸)، دقت توصیفی (عناصر ۲، ۳، ۷ و ۲۳)، سودمندی (عناصر ۴، ۵ ،۶، ۹ و ۱۴)، قابلیت اقناع (عنصر۸ ،۱۰ و ۲۴)، شفافیت (۱۱، ۱۲، ۲۰، ۲۱ و ۲۲ و ۲۳)، قابلیت حکمرانی (۱۳، ۱۵و ۱۹).

برای غربال‌گری و اعتبارسنجی مقوله‌های مرتبهٔ دوم، از تحلیل فازی مبتنی بر دیدگاه ۱۵ مشارکت‌کننده در پژوهش استفاده شد. جداول ۵ و ۶ نتایج میانگین فازی و مقدار فازی‌زدایی شده مؤلفه‌ها در مرحلهٔ اول و دوم دلفی فازی را نمایش داده است.

جدول۵. نتایج اجرای مرحلهٔ نخست دلفی فازی

مؤلفه‌ها	میانگین فازی (m,a,B)	فازی‌زدایی شده
قابلیت درک	(۰/۹۶۴، ۰/۸۳۹، ۰/۵۸۹)	۰/۶۲۰
دقت توصیفی	(۰/۹۸۲، ۰/۹۴۶، ۰/۶۹۶)	۰/۷۰۵
قابلیت اقناع	(۰/۹۶۴، ۰/۸۵۷، ۰/۶۰۷)	۰/۶۳۳
شفافیت	(۰/۹۸۲، ۰/۹۱۰، ۰/۶۶۰)	۰/۶۷۸
قابلیت حکمرانی	(۰/۹۶۴، ۰/۸۵۷، ۰/۶۰۷)	۰/۶۳۳
سودمندی	(۰/۹۸۲، ۰/۹۲۸، ۰/۶۷۸)	۰/۶۹۱

با پایان مرحلهٔ نخست، ضرورت دارد که مرحله دوم نیز انجام شود تا نتایج هر دو مرحله با هم مقایسه و ارائه شود. در جدول ۶ نتایج نظرسنجی مرحلهٔ دوم دلفی فازی نمایش داده شده است.

با پایان نظرسنجی در مرحلهٔ دوم، ضرورت دارد که اختلاف میانگین فازی‌زدایی شده الزامات طراحی سیستم‌های با قابلیت تبیین در تعامل انسان- هوش مصنوعی مورد تحلیل قرار گیرد.

جدول۶. نتایج اجرای مرحلهٔ دوم دلفی فازی

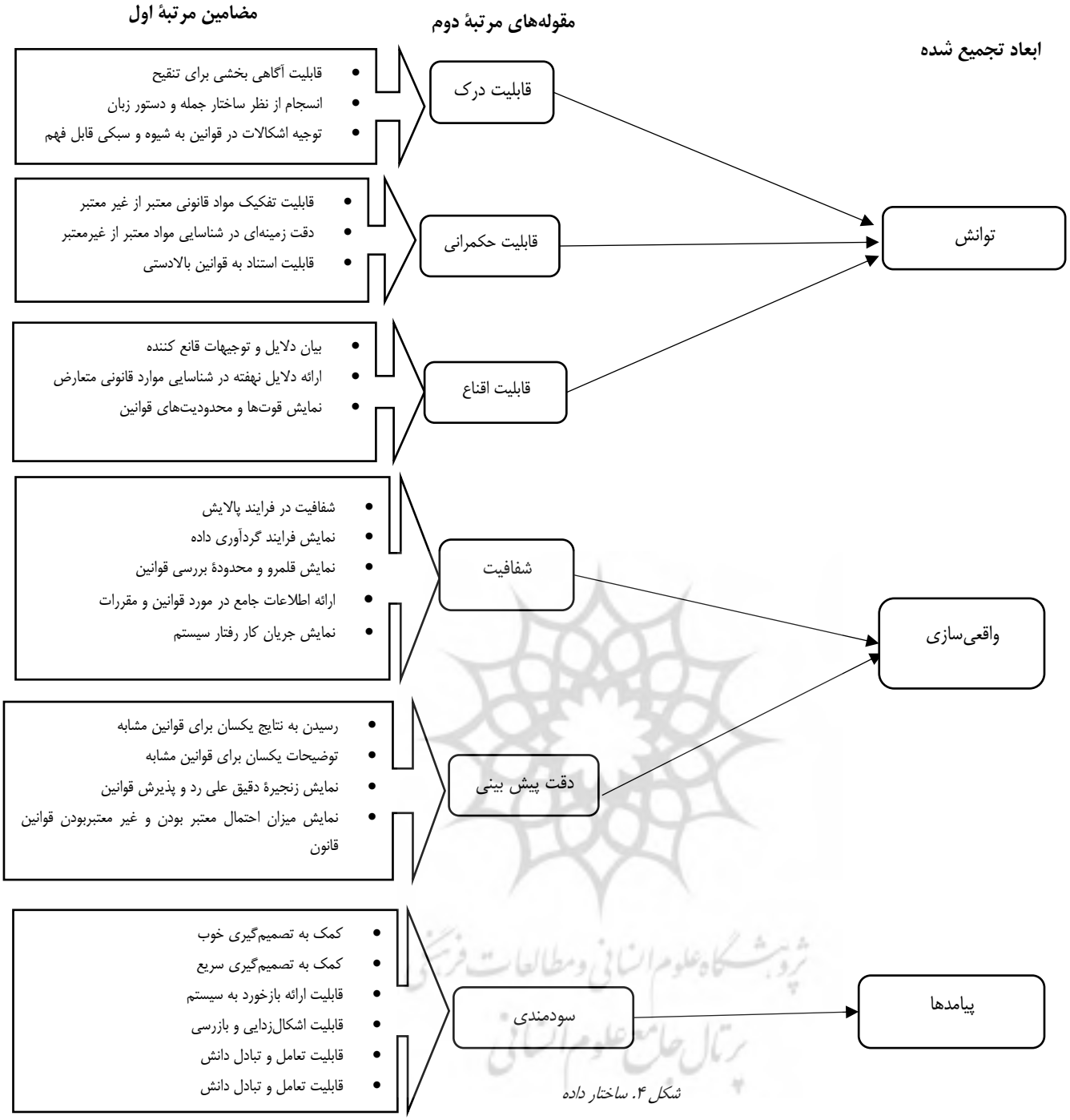
مؤلفه‌ها	میانگین فازی (m,a,B)	فازی‌زدایی شده
قابلیت درک	(۰/۹۶۴، ۰/۸۵۷، ۰/۶۰۷)	۰/۶۳۳
دقت توصیفی	(۰/۹۸۲، ۰/۹۶۴، ۰/۷۱۴)	۰/۷۱۸
قابلیت اقناع	(۰/۹۶۴، ۰/۸۷۵، ۰/۶۲۵)	۰/۶۴۷
شفافیت	(۰/۹۸۲، ۰/۹۱۰، ۰/۶۶۰)	۰/۶۷۸
قابلیت حکمرانی	(۰/۹۴۶، ۰/۸۳۹، ۰/۵۸۹)	۰/۶۱۶
سودمندی	(۰/۹۶۴، ۰/۹۱۰، ۰/۶۶۰)	۰/۶۷۴

جدول ۷ بررسی اختلاف میانگین فازی‌زدایی شده مرحلهٔ اول و دوم نمایش داده شده است.

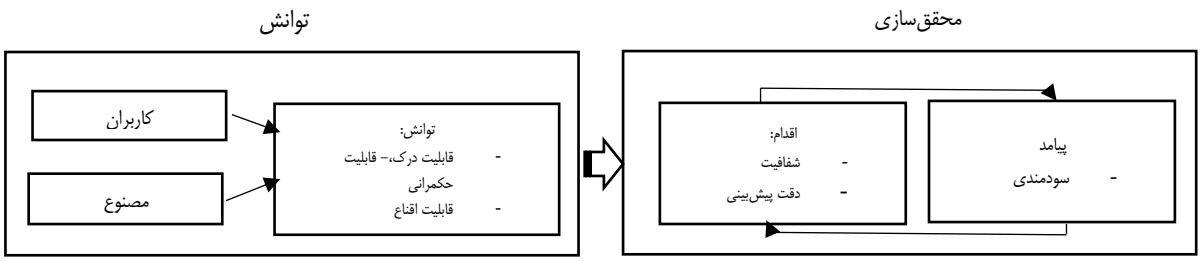
جدول۷. میانگین دیدگاه خبرگان حاصل از نظرسنجی مرحله اول و دم

مؤلفه‌ها	مقدار فازی‌زدایی شده مرحله اول	مقدار فازی‌زدایی شده مرحله دوم	اختلاف میانگین
قابلیت درک	۰/۶۲۰	۰/۶۳۳	۰/۰۱۳
دقت توصیفی	۰/۷۰۵	۰/۷۱۸	۰/۰۱۳
قابلیت اقناع	۰/۶۳۳	۰/۶۴۷	۰/۰۱۴
شفافیت	۰/۶۷۸	۰/۶۷۸	۰
قابلیت حکمرانی	۰/۶۳۳	۰/۶۱۶	۰/۰۱۷
سودمندی	۰/۶۹۱	۰/۶۷۴	۰/۰۱۷

با توجه به نتایج جدول ۷ اختلاف میانگین فازی‌زدایی شده در مرحله اول و دوم دلفی فازی کمتر از ۰/۱ است. از این جهت، می‌توان نتیجه گرفت که مشارکت‌کنندگان در مورد الزامات طراحی سیستم‌های با قابلیت تبیین به اجماع رسیده‌اند و نظرسنجی در این مرحله متوقف می‌شود. در نهایت هر یک از این مؤلفه‌ها در دو بُعد اصلی توانش و محقق‌سازی خلاصه شدند. شکل ۴ ساختار داده را نمایش می‌دهد.
هدف از مرحلهٔ پایانی پژوهش طراحی مبتنی بر اقدام، تعمیم خروجی و قرار دادن آن در دامنهٔ گسترده‌تر مسائل است. بنابراین تبدیل پدیدهٔ انضمامی (وابسته به زمان و مکان) به یک پدیدهٔ انتزاعی بخش مهمی از پژوهش طراحی مبتنی بر اقدام است [۲۴]. با تجزیه و تحلیل داده‌های کیفی، مقوله‌های نوپدیدی توسعه یافت که می‌توان پیوند مفهومی آنها را با مبانی نظری برقرار نمود. برای انجام این کار، با وام گرفتن از نظریهٔ توانش، مدل مفهومی‌پژوهش استخراج و روابط میان یافته‌ها ترسیم شد. بر اساس این نظریه، پیش‌نیاز محقق‌سازی یک رویداد، وجود توانش است. همانطور که در شکل ۵ نمایش داده شده است، توانش در تقاطع مصنوع و کاربران قرار دارد. در این پژوهش سه دسته توانش در سیستم‌های هوش مصنوعی با قابلیت تبیین شناسایی شده است که عبارتند از : قابلیت درک، قابلیت حکمرانی و قابلیت اقناع. مرحلهٔ محقق‌سازی این توانش‌ها از رهگذر شفافیت و دقت پیش‌بینی سیستم محقق می‌شود تا تعامل انسان– هوش مصنوعی منجر به برآورده شدن هدف‌های سازمانی شود که پیامد آن سودمندی این همکاری است.



شکل ۴. ساختار داده



شکل۵. مدل مفهومی پژوهش (یافته‌های پژوهش)

برای اعتبار سنجی مدل مفهومی پژوهش از روش مدل‌سازی ساختاری-تفسیری استفاده شد. سازه‌های مورد بررسی شامل: قابلیت درک(C01)، دقت پیش‌بینی (C02)، قابلیت اقناع (C03)، شفافیت (C04)، قابلیت حکمرانی (C05) و سودمندی(C06) هستند. در روش مدل‌سازی ساختاری- تفسیری، نخستین گام، تشکیل ماتریس خودتعاملی ساختاری است. این ماتریس نشان می‌دهد، یک متغیر بر کدام متغیرها تأثیر دارد و از کدام متغیرها تأثیر می‌پذیرد. جدول ۸ ماتریس دستیابی را نمایش می‌دهد که از تبدیل ماتریس خودتعاملی ساختاری به یک ماتریس دو ارزشی صفر و یک بدست آمده است. در این ماتریس، درایه‌های قطر اصلی برابر یک قرار می‌گیرند. برای ایجاد ماتریس دسترسی نهایی لازم است که روابط ثانویه کنترل و در صورت لزوم اطلاعات ماتریس اصلاح شود. به عبارت دیگر، اگر A منجر به B و B منجر به C شود، در این صورت، A نیز منجر به C می‌شود.

جدول۸. ماتریس دستیابی نهایی متغیرهای پژوهش

	C01	C02	C03	C04	C05	C06
C01	۱	۱	۱	۱	۱	۱
C02	۰	۱	۰	۱	۰	۱
C03	۱	۱	۱	۱	۱	۱
C04	۰	۱	۰	۱	۰	۱
C05	۱	۱	۱	۱	۱	۱
C06	۰	۰	۰	۰	۰	۱

برای تعیین روابط و سطح‌بندی معیارها باید مجموعه دستیابی و مجموعه پیش‌نیاز برای هر معیار از ماتریس دستیابی نهایی استخراج شود: مجموعه دستیابی (عناصر سطر، اثرگذاری‌ها): برای متغیر Ci مجموعه دستیابی شامل متغیرهایی است که از طریق متغیرCi می توان به آنها رسید. مجموعه پیش‌نیاز (عناصر ستون، اثرپذیری‌ها): مجموعه پیش‌نیاز شامل متغیرهایی است که از طریق آنها می‌توان به متغیر Ci رسید.

پس از تعیین مجموعه دستیابی و مجموعه پیش‌نیاز، اشتراک دو مجموعه محاسبه می‌شود. بر اساس جدول ۹ اولین متغیری که اشتراک دو مجموعه برابر با مجموعه قابل دستیابی (خروجی) باشد سطح اول در نظر گرفته می‌شود. از این جهت عناصر سطح اول، تأثیرگذارترین عناصر در مدل هستند. در هر مرحله، پس از تعیین سطح، معیاری که سطح آن معلوم شده است، از تمامی مجموعه حذف و دوباره مجموعه ورودی‌ها و خروجی‌ها تشکیل می‌شود و بدین ترتیب، سطح‌بندی متغیرهای دیگر بدست می‌آید.

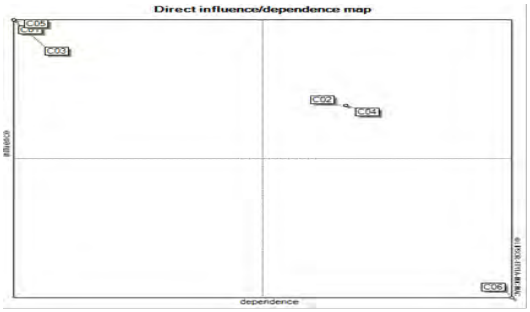
جدول ۹. تعیین سطح الزامات طراحی سیستم‌های با قابلیت تبیین

متغیر	مجموعه دستیابی (اثرگذاری)	مجموعه پیش‌نیاز (اثرپذیری)	اشتراک
C01	C01,C02,C03,C04,C05,C06	C01,C03,C05	C01,C03,C05
C02	C02,C04,C06	C01,C02,C03,C04,C05	C02,C04
C03	C01,C02,C03,C04,C05,C06	C01,C03,C05	C01,C03,C05
C04	C02,C04,C06	C01,C02,C03,C04,C05	C02,C04
C05	C01,C02,C03,C04,C05,C06	C01,C03,C05	C01,C03,C05
C06	Co6	C01,C02,C03,C04,C05,C06	C06

بر اساس نتایج مدل‌سازی ساختاری تفسیری ارائه شده در جدول ۹، متغیر سودمندی (C06) در سطح نخست قرار گرفت. متغیرهای دقت پیش‌بینی (C02) و شفافیت (C04) در سطح دوم و متغیرهای قابلیت درک (C01)، قابلیت اقناع (C03) و قابلیت حکمرانی (C05) در سطح سوم قرار گرفتند. پس از تعیین سطوح الزامات طراحی سیستم‌های هوش مصنوعی با قابلیت تبیین و تحلیل قدرت نفوذ- وابستگی با استفاده از نرم افزار میک مک انجام شد. میزان وابستگی و قدرت نفوذ متغیرها در شکل ۶ نمایش داده شده است.

بر اساس مدل ساختاری- تفسیری، قدرت نفوذ و میزان وابستگی متغیرها در چهار زیرگروه مستقل، خود مختار، پیوندی (رابط) و مستقل قابل دسته‌بندی و تحلیل هستند. بر اساس مدل پیشنهادی، متغیرهای پیشران قابلیت درک (C01)، قابلیت اقناع (C03) و قابلیت حکمرانی (C05) دارای قدرت نفوذ بالا و میزان وابستگی پایین هستند و در گروه متغیرهای مستقل قرار می‌گیرند. متغیرهای دقت پیش‌بینی (C02) و شفافیت (C04) دارای قدرت نفوذ و میزان وابستگی بالایی هستند. بنابراین در گروه متغیرهای پیوندی قرار می‌گیرند. میزان تأثیرگذاری

و تأثیرپذیری متغیرهای پیوندی بسیار بالا است و هر تغییر کوچکی بر روی این متغیرها باعث تغییرات اساسی در سیستم می‌شود. متغیر سودمندی (C06) نیز از وابستگی بالا اما نفوذ اندکی برخوردار است. بنابراین متغیر وابسته محسوب می‌شود. در بین متغیرهای پژوهش، هیچ متغیری از دسته متغیر خودمختار مورد شناسایی قرار نگرفت. با توجه به این نتایج گزاره‌های پژوهش طراحی شده است.



شکل ۶. میزان وابستگی و قدرت نفوذ متغیرها

توانش. توانش‌های مورد نیاز در تعامل انسان با سیستم‌های هوش مصنوعی با قابلیت تبیین در بخش دولتی منتهی به گزاره‌های۱، ۲ و ۳ پژوهش به شرح زیر می‌شود.

توانش۱: قابلیت درک. برای شناسایی و انتخاب اقدامات مناسب در مورد زمینه‌ای که اقدام در آن رخ می‌دهد، اطلاعاتی باید گردآوری شود. یکی از این مسیرها درک محیط پیرامون است. این توانش در پژوهش با سه مضمون " قابلیت آگاهی بخشی"، " انسجام مفهومی" و " ارائه مطالب به شیوهٔ قابل فهم " تعریف می‌شود. با بررسی پیشینه موضوع مشخص شد، قابلیت درک یکی از معیارهای ارزیابی تبیین معرفی شده است [۳۱]،[۱]. در این پژوهش‌ها، قابلیت درک به توانایی یک سیستم در عرضهٔ دانش خود به شیوه‌ای قابل درک اطلاق می‌شود که با مضامین " انسجام مفهومی" و "ارائهٔ مطالب به شیوه‌ای قابل فهم" منطبق است. برخی از پژوهش‌ها نیز قابلیت درک تبیین را از جنبهٔ شهود مورد بررسی قرار داده‌اند [۱۰]،[۱۵]. آنها استدلال کردند که تجربهٔ کاربر، میزان درک شناختی آنها را تعیین می‌کند. میزان درک شناختی در زمینهٔ سیستم‌های اطلاعاتی مدیریت، بیانگر تجربهٔ کاربر در نحوهٔ سازماندهی عناصر مجزای اطلاعات در یک مجموعه است. بنابراین، سیستم‌های با قابلیت تبیین، باید متناسب با سطح پیچیدگی تخصص کاربران باشند. این پیشینه مطابق با مضمون "قابلیت آگاهی" در پژوهش حاضر است. بنابراین، در این پژوهش قابلیت آگاهی‌بخشی به سطح بینش‌افزایی مورد انتظار سامانه اشاره می‌کند. با توجه به موارد اشاره شده، گزارهٔ اول پژوهش به این صورت تدوین می شود:

گزارهٔ ۱: قابلیت درک عامل هوشمند در تعامل انسان– هوش مصنوعی منجر به تقویت هوشمندی می‌شود.

قابلیت حکمرانی. پیشینهٔ موضوع نشان می‌دهد که کاربرد سیستم‌های هوش مصنوعی در تصمیم‌گیری‌های بخش دولتی در حال گسترش است. یکی از الزامات کاربست این نوع سیستم‌ها، منطبق بودن گزینه‌های تصمیم پیشنهادی با قوانین و مقررات حاکم بر کشور است[۷]. یافته‌های این پژوهش در قالب مضامین "تفکیک مواد قانونی معتبر از غیر معتبر"، "دقت زمینه‌ای" و "استناد به قوانین و مقررات بالادستی" این موضوع را تصدیق می‌کنند. مضامین "تفکیک مواد قانونی معتبر از غیر معتبر" و "استناد به قوانین و مقررات بالادستی " را می‌توان بر اساس بعد تفسیرپذیری جهانی در برابر محلی مدل‌های پیش‌بینی نیز تفسیر نمود. در تفسیرپذیری جهانی، زمانی نتایج مدل قابل فهم است که بتوان منطق کلی سیستم را درک نمود، بدین معنا که همهٔ پیش‌بینی‌های مدل تصمیم‌گیری منطبق بر یک مدل تحلیلی بوده و خروجی سیستم بر پایه این چارچوب تحلیلی است. در حالیکه تفسیرپذیری محلی وضعیتی را نشان می‌دهد که تنها می‌توان به دلایل یک تصمیم خاص پی برد [۱۳]. همچنین برخی از پژوهشگران دریافتند زمانی قابلیت درک تقویت می‌شود که مصنوع سطح بالایی از جزئیات راهبردی را فراهم نماید [۱۲] که بیانگر اهمیت مضمون "دقت زمینه‌ای" در طراحی سیستم‌های مبتنی بر تعامل انسان– هوش مصنوعی است. از این جهت می‌توان استنباط نمود که عامل هوشمند می‌بایست قابلیت حکمرانی را در قانون‌گذاری شبیه‌سازی نماید. با توجه به این بحث گزارهٔ دوم پژوهش مطرح می‌شود:

گزارهٔ ۲: قابلیت حکمرانی عامل هوشمند در تعامل انسان– هوش مصنوعی منجر به تقویت هوشمندی می‌شود.

قابلیت اقناع. یکی از معیارهای ارزیابی تبیین خوب، میزان قانع‌کنندگی آن است. مشارکت‌کنندگان از سه مضمون " دلایل و توجیهات قانع کننده"، "استدلال‌های نهفته در شناسایی موارد قانونی متعارض" و "نمایش قوت‌ها ومحدودیت‌های قوانین" برای ارزیابی قانع‌کنندگی سیستم استفاده کرده‌اند. این یافته‌ها، مطابق با پیشینه موجود در حوزهٔ سیستم‌های با قابلیت تبیین است که بر ضرورت قانع‌کنندگی سیستم برای کاربران تأکید دارند [۳۹]،[۱۸]. برخی از پژوهش‌ها نیز نشان دادند که قابلیت پایین اقناع‌کنندگی یک عامل هوشمند، نه‌تنها می‌تواند به دلیل روش تبیین ضعیف باشد، بلکه دقت پایین مدل تبیین و روش استدلال متفاوت از درک انسان نیز در ضعف قانع‌کنندگی نقش دارند. بنابراین، ضروری است مدل استدلال کاربران و عامل هوشمند شبیه هم باشد [۳۹]. گرچه در پیشینه موضوع، عوامل مختلفی برای ارزیابی درونی سیستم‌های با قابلیت تبیین ارائه شده است، اما یک توافق عمومی میان پژوهشگران و دست‌اندرکان هوش مصنوعی با قابلیت تبیین وجود دارد که مطرح می‌کند، چرایی و چگونگی استفاده از تبیین‌ها توسط کاربران در ارزیابی سیستم‌های با قابل تبیین حائز اهمیت است. از این جهت، در بستر این پژوهش، ویژگی قانع‌کنندگی یک توانش مورد نیاز برای تقویت هوشمندی شناسایی شده است. بر همین اساس، گزارهٔ سوم پژوهش به این صورت تدوین می شود.

گزارهٔ ۳: قابلیت اقناع عامل هوشمند در تعامل انسان– هوش مصنوعی منجر به تقویت هوشمندی می‌شود.

محقق‌سازی. مرحلهٔ محقق‌سازی، زمانی رخ می‌دهد که همهٔ توانش‌های مورد نیاز کاربران برای رسیدن به هدف‌های سازمانی از طریق ویژگی‌های دقت پیش‌بینی و شفافیت محقق شود و این سودمندی عامل هوشمند را در پی دارد که منجر به تقویت هوشمندی کاربران می‌شود. در ادامه، هر یک از این ویژگی‌ها مورد بحث قرار می‌گیرند که منجر به گزاره‌های ۴ ، ۵ و ۶ می‌شود.

شفافیت. در پیشینه موضوع سیستم‌های هوش مصنوعی با قابلیت تبیین، شفافیت به خودبسندگی مدل در بیان روابط میان اجزا تبیین می‌شود. همچنین از طریق رویکردهای انسان‌محور و مشارکت کاربران در طراحی سیستم، می‌توان شفافیت سیستم‌های هوش مصنوعی باقابلیت تبیین را افزایش‌داد. با این وجود، روش‌های تبیین موجود در بیشتر موارد به مسألهٔ شفافیت مدل‌های جعبه‌سیاه پرداخته‌اند [۱۲،۳۱]. اما در این پژوهش، از رویکرد وابسته به زمینه طراحی مبتنی بر اقدام، موضوع شفافیت مورد بحث قرار گرفت و مضامین"شفافیت در فرایند پالایش"، "نمایش قلمرو و محدودهٔ بررسی قوانین"، " ارائهٔ اطلاعات جامع در مورد قوانین و مقررات" و " نمایش جریان کار رفتار سیستم"، بیانگر شفافیت مورد انتظار کاربران از عامل هوشمند در زمینهٔ مورد مطالعه است. افزون بر این، تأکید شده است درک عملی برای اتکاء به توصیه‌های عامل هوشمند، نه‌تنها مستلزم دانستن نحوهٔ رفتار عامل هوشمند است، بلکه آگاهی از قلمرو استدلال عامل هوشمند نیز بسیار حائز اهمیت است [۱۸]. این مبانی بیانگر این است که هریک از مضامین استخراج شده بطور ضمنی در راستای پیشینه موضوع هستند. از این جهت گزارهٔ چهارم پژوهش به این صورت تدوین می شود:

گزارهٔ ٤: *شفافیت عامل هوشمند در تعامل انسان–هوش مصنوعی منجر به تقویت هوشمندی می‌شود.*

دقت پیش‌بینی. این مقوله جزء وظیفه‌ای عامل هوشمند محسوب می‌شود و از مضامین" نتایج یکسان برای قوانین مشابه"، "توضیحات یکسان برای قوانین مشابه"، "نمایش زنجیرهٔ دقیق علی رد و پذیرش قوانین" و "میزان احتمال معتبر بودن و غیر معتبر بودن قوانین" تشکیل شده است. با بررسی پیشینه موضوع می‌توان استنباط کرد که هر یک از این مضامین بر دقت پیش‌بینی عامل هوشمند می‌افزایند. به نوبهٔ خود، دقت پیش‌بینی بیانگر توانایی حل مسأله از طریق تبیین تعریف شده است [۸]. یکی از ویژگی‌های مورد انتظار عامل هوشمند این است که کاربران را قادر می‌سازد درک خود از سیستم را بهبود بخشند. به عبارت دیگر، اطلاعات باید نمایش دقیقی از عملکرد سیستم را القاء کند. بدین ترتیب، رابط‌های تبیین‌گر، باید عملکرد سیستم را به دقت منعکس نمایند [۳۱] با توجه به این موارد گزارهٔ پنجم به شرح زیر توسعه می‌یابد:

گزارهٔ ۵: دقت پیش‌بینی عامل هوشمند در تعامل انسان– هوش مصنوعی منجر به تقویت هوشمندی می‌شود.

سودمندی. از دیدگاه مشارکت‌کنندگان، زمانی عامل هوشمند در زمینهٔ قانون‌گذاری سودمند است که به"تصمیم‌گیری خوب و سریع" کمک نماید؛ " قابلیت بازخورد " داشته باشد؛ "قابلیت اشکال‌زدایی و بازرسی" داشته باشد و در نهایت "امکان تبادل و تعامل دانش" را فراهم نماید. با این وجود، بررسی پیشینه موضوع نشان می‌دهد که سودمندی سیستم‌های با قابلیت تبیین به ماهیت پیش‌بینی و شفافیت آن بستگی دارد [۱] که به نوبهٔ خود تصدیق‌کنندهٔ مضامین استخراج شده است. بنابراین، می‌توان استنباط نمود که عامل هوشمند از طریق دقت پیش‌بینی و شفافیت، توانش موردنیاز کاربران را فراهم می‌کند. با توجه به این موارد گزارهٔ ششم پژوهش طراحی می‌شود:

گزارهٔ ٦. سودمندی عامل هوشمند، در تعامل انسان هوش مصنوعی منجر به تقویت هوشمندی می‌شود.

۵. نتیجه‌گیری و پیشنهاد

هدف این پژوهش، طراحی یک مدل نظری و کاربردی بود که درآن الزامات طراحی سیستم‌های هوش مصنوعی با قابلیت تبیین در تعامل انسان- هوش مصنوعی شناسایی شد. این مدل، در واقع مصنوع چرخهٔ تشخیص در پژوهش طراحی مبتنی بر اقدام معرفی شد. از این جهت یافته‌های پژوهش نشان داد یک سیستم مبتنی بر تعامل انسان- هوش مصنوعی برای تصمیم‌گیری در سازمان‌ها می‌بایست پنج ویژگی قابلیت درک، قابلیت حکمرانی، قابلیت اقناع، دقت توصیفی، شفافیت و سودمندی را به همراه داشته باشد. این قابلیت‌ها در دو بُعد طبقه‌بندی شدند. بُعد توانش شامل قابلیت درک، قابلیت حکمرانی و قابلیت اقناع است. از دیدگاه پژوهشگران، طراحان این طبقه از سیستم‌ها، با در نظر گرفتن این قابلیت‌ها، می‌توانند زمینهٔ تقویت هوشمندی را در تعامل انسان-هوش مصنوعی فراهم کنند. بنابراین، در طراحی سیستم‌ها در طبقهٔ تعامل انسان- هوش مصنوعی، نقش کاربر و سطح پیچیدگی تخصص آنها به چند دلیل اهمیت دارد: نخست، میزان درک شناختی کاربران، مدل تبیین سیستم را مشخص می‌کند؛ دوم، دقت زمینه‌ای و سطح بالای جزئیات استراتژیک سطح اعتماد را در کاربران افزایش می‌دهد و این درک را القاء می‌کند که عامل هوشمند، رفتار قانون‌گذار را شبیه‌سازی کرده است؛ و در نهایت، شیوهٔ ارزیابی عملکرد سیستم را مشخص می‌کند. بُعد محقق‌سازی نیز شامل دقت پیش‌بینی، شفافیت و سودمندی است که ویژگی‌های درونی سیستم شناخته می‌شوند. افزون بر این، یافته‌ها نشان داد که شفافیت، پدیده‌ای وابسته به زمینه است و انتظارات کاربران، سطح پاسخگویی، انصاف و شفافیت را مشخص می‌کند. همچنین تناسب مدل‌های پیش‌بینی با زمینه، میزان دقت پیش‌بینی را افزایش می‌دهد که به نوبهٔ خود منجر به سودمندی ادراک شده توسط کاربران می‌شود. از این جهت طراحان باید اهمیت رابطهای تبیین‌کننده را که منعکس کنندهٔ عملکرد دقیق سیستم است مورد توجه قرار دهند. در نهایت، یافته‌های این پژوهش منطبق بر مطالعهٔ میلر[1] [۲۳] بوده که در آن بر نقش سوءگیری‌های شناختی و انتظارات اجتماعی کاربران در پیکربندی ویژگی‌های تبیین خوب تأکید شده است. همچنین این پژوهش گام مهمی را در راستای مطالعهٔ گیدوتی[2] و همکاران[۱۵] برداشته و نظریهٔ توانش را یک ساختار مناسب برای تعریف تبیین معرفی می‌کند.

بر همین اساس، با توجه به یافته‌های پژوهش، موارد زیر را می‌توان سهم نظری این پژوهش برشمرد.

نخست، این پژوهش از طریق توسعهٔ یک مدل نظری و کاربردی، سازوکار طراحی سیستم‌های مبتنی بر تعامل انسان- هوش مصنوعی را بنا نهاد. با بررسی یافته‌های این پژوهش و پیشینه موضوع مشخص می‌شود که در پژوهش سیستم‌های اطلاعاتی مدیریت، الزامات طراحی چنین سیستم‌هایی مورد مطالعه قرار نگرفته است و دانش توصیفی در مورد شناسایی فضای مسألهٔ طراحی سیستم‌های مبتنی بر تعامل انسان- هوش مصنوعی در تصمیم‌گیری وجود ندارد. بنابراین درک کامل این فضا ضروری به نظر می‌رسید. از این جهت، این پژوهش با انتزاعی‌کردن چرخهٔ تشخیص، یک مدل را ارائه نمود که الزامات تعامل انسان- هوش مصنوعی را در دو بُعد توانش و محقق‌سازی صورت‌بندی کرده است. برای این کار از لنز نظریهٔ توانش نیز استفاده شد. افزون بر این، پژوهش، با یک دیدگاه انسان محور، ویژگی‌هایی که کاربر با استفاده از مصنوع برای تقویت هوشمندی نیاز دارد را شناسایی کرد که عبارتند از: قابلیت درک، قابلیت حکمرانی، قابلیت اقناع.

دومین سهم نظری پژوهش، ارائهٔ دانش تجویزی در مورد فضای راه‌حل است. در این پژوهش مطرح شد عامل هوشمند از طریق دقت پیش‌بینی و شفافیت، توانش مورد نیاز کاربران را محقق می‌کند. این ویژگی‌های سیستم به نوبهٔ خود، پیش‌نیاز اصول طراحی برای سیستم-های مبتنی بر تعامل انسان- هوش مصنوعی است که می‌بایست در چرخهٔ طراحی لحاظ شود.

سومین سهم نظری پژوهش، مجموعهٔ الزامات پیاده‌سازی سیستم‌های مبتنی بر تعامل انسان- هوش مصنوعی است. این مجموعه الزامات در قالب یک مدل نظری، راهنمای توسعهٔ اصول طراحی سیستم‌های هوش مصنوعی در سازمان‌ها را تشکیل می‌دهد که دغدغهٔ اصلی پژوهش‌ها بوده است.

پیشنهادهای این پژوهش برای طراحان و پژوهشگران به شرح زیر است:

طراحان. این پژوهش، رهنمودهای کاربردی را برای طراحان و توسعه‌دهندگان سیستم‌های مبتنی بر تعامل انسان- هوش مصنوعی در تصمیم‌گیری فراهم می‌کند. مصنوع این پژوهش که در قالب مجموعه الزامات طراحی سیستم‌های با قابلیت تبیین در تعامل انسان- هوش مصنوعی است، در زمینه‌های عملی گسترده‌تر نیز قابل استفاده هستند؛ زیرا قانون‌گذاری مستلزم دیدگاه فرابخشی است. افزون بر

[1] Miller

[2] Guidotti

این، مصنوعات این پژوهش در بستر قوهٔ قضائیه توسعه یافته است و نتایج این پژوهش را می‌توان از طریق اجرای آن در سایر دستگاه‌های اجرایی تعمیم داد. از این رو، به طراحان پیشنهادهای زیر ارائه می‌شود:

۱- طراحی سیستم‌های مبتنی بر تعامل انسان- هوش مصنوعی متناسب سطح تخصص کاربران
۲- انتخاب مدل‌های تصمیم بر اساس دقت زمینه‌ای
۳- ارزیابی عملکرد سیستم بر اساس چرایی و چگونگی استفادهٔ کاربران
۴- مشخص کردن قلمرو استدلال مدل پیش‌بینی
۵- طراحی رابط‌های تبیین‌کننده متناسب با سطح کاربران

پژوهشگران. مجموعه الزامات این پژوهش، در قالب گزاره‌هایی طراحی شده است. با این وجود محدودیت‌هایی در انجام این پژوهش وجود دارد که به پژوهش‌های آتی کمک می‌کند از مسیرهای جدید به مسأله پژوهش وارد شوند. از این جهت پیشنهادهای زیر مطرح می‌شود:

۱- *استفاده از رویکرد کمّی پژوهش برای ارزیابی اعتبار گزاره‌های پژوهش*؛ با انجام پژوهش در یک زمینه نمی‌توان ادعای تعمیم‌پذیری مدل پیشنهادی را مطرح کرد. بنابراین قبل از طرح چنین ادعاهایی نیاز به پژوهش‌های بیشتر وجود دارد. بنابراین آزمون گزاره‌های این پژوهش در سازمان‌هایی که اهداف مشابهی دارند بسیار مؤثر است. انجام این کار نیازمند بکارگیری روش‌های پیمایشی است که قابلیت تعمیم نتایج را فراهم می‌کند.
۲- *بسط چارچوب پیشنهادی از طریق مشارکت ذینفعان خارجی در پژوهش*؛ گرچه این پژوهش، بطور کامل ذینفعان داخلی سیستم را مورد مطالعه قرار داده است. با این وجود، پژوهشگران معتقدند تصویر کاملی از سیستم ارائه نشده است، زیرا ذینفعان خارجی سیستم مورد مطالعه قرار نگرفته‌اند. از این رو در در نظر گرفتن بخش‌های مرتبط در معاونت قوانین مجلس شورای اسلامی و معاونت حقوقی نهاد ریاست جمهوری از جمله ذینفعان خارجی سیستم تصویر کاملی را ارائه می‌کند.
۳- *بررسی پیشایندهای سیستم‌های هوش مصنوعی با قابلیت تبیین؛ در این پژوهش پیشایندهای مؤثر در کاربرد اثربخش* سامانه، همانند مهارت‌های کاربر بررسی نشده است. از اینرو، فرصت‌هایی برای پژوهشگران وجود دارد تا تأثیر آموزش را بر جنبه‌های مختلف تعامل انسان- هوش مصنوعی، مورد مطالعه قرار دهند.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, *6*, 52138-52160.,doi: 10.1109/ACCESS.2018.2870052.
2.Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: "automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, *33*(1), 153-169, https://doi.org/10.1093/jopart/muac007.
3.Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society, 35*(3), 611-633, https://doi.org/10.1007/s00146-019-00931-w.
4. Burton-Jones, A., & Volkoff, O. (2017). How Can We Develop Contextualized Theories of Effective Use? A Demonstration in the Context of Community-Care Electronic Health Records. *Information Systems Research, 28*(3), 451-679, https://doi.org/10.1287/isre.2017.0702.
5.Castano, S., Falduti, M., Ferrara, A., & Montanelli, S. (2022). A knowledge-centered framework for exploration and retrieval of legal documents. *Information Systems, 106*, 101842, https://doi.org/10.1016/j.is.2021.101842.
6.Chen, H., Wu , L., Chen, J., Lu, W., & Ding, J. (2022). A comparative study of automated legal text classification using random forests and deep learning. *Information Processing and Management, 59*(2), 102798, https://doi.org/10.1016/j.ipm.2021.102798.
7.Cobbe, J. (2019). Administrative law and the machines of government: judicial review of automated public-sectordecisionmaking. *Legal Studies, 39*(4), 1-20, https://doi.org/10.2139/ssrn.3226913.
8.De Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & Society, 35*(4), 917-926, https://doi.org/10.1007/s00146-020-00960-w.
9.Di Vaio, A., Hassan, R., & Alavoine, C. (2022). Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting & Social Change, 174*, 121201, https://doi.org/10.1016/j.techfore.2021.121201
10.Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
11.Du, W., Pan, S. L., Leidner, D. E., & Ying, W. (2019). Affordances, experimentation and actualization of FinTech:A blockchain implementation study. *Journal of Strategic Information Systems, 28*(1), 50-65, https://doi.org/10.1016/j.jsis.2018.10.002.
12.Ehsan, U., & O. Riedl, M. (2020). Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach. *International Conference on Human-Computer Interaction* (pp. 449-466). Springer.
13. Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research, 33*(2), 678-696, https://doi.org/10.1287/isre.2021.1079.
14.Gilpin, L. H., Bau, D., Z, B. Y., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *IEEE 5th International Conference on data science and advanced analytics (DSAA)* (pp. 80-89). IEEE.
15.Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). A Survey Of Methods For Explaining Black Box Models. *ACM computing surveys (CSUR), 51*(5), 1-42, https://doi.org/10.1145/3236009.
16.Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior, 119*, 106730. https://doi.org/10.1016/j.chb.2021.106730
17. Jacovi, A., Shalom, O. S., & Goldberg, Y. (2018). Understanding convolutional neural networks for text classification. *arXiv preprint arXiv:1809.08037*.
18.Jarrahi, M. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577-586, https://doi.org/10.1016/j.bushor.2018.03.007
19. Kuo, Y.-F., & Chen, P.-C. (2008). Constructing performance appraisal indicators for mobility of the service industries using Fuzzy Delphi Method. *Expert systems with applications, 35*(4), 1930-1939, https://doi.org/10.1016/j.eswa.2007.08.068.

20.Kulesza, T., Burnett, M., Wong, W.-K., & Stumpf, S. (2015). Principles of Explanatory Debugging to Personalize Interactive Machine Learning. *Proceedings of the 20th international conference on intelligent user interfaces* (pp. 126-137). IUI.
21.Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790*.
22. Maier, J. R., & Fadel, G. M. (2009). Affordance based design: a relational theory for design. *Research in Engineering Design, 20*(1), 13-27, https://doi.org/10.1007/s00163-008-0060-3.
23.Miller, T. (2019). Explanation in Artificial Intelligence:Insights from the Social Sciences. *Artificial intelligence, 267*, 1-38.
24. Mullarkey, M. T., Hevner, A. R., & Ågerfalk, P. (2019). An elaborated action design research process model. *European Journal of Information Systems, 28*(1), 6-20, https://doi.org/10.1080/0960085X.2018.1451811.
25.Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management, 51*(6), 801-809, https://doi.org/10.1016/j.im.2014.01.002.
26.Ogunbiyi , N., Basukoski, A., & Chaussalet, T. (2021). An Exploration of Ethical Decision Making with Intelligence Augmentation. *social sciences, 10*(2), 57, https://doi.org/10.3390/socsci10020057.
27.Pan, S. L., Li, M., Pee, L. G., & Sandeep, M. S. (2020). Sustainability Design Principles for a Wildlife Management Analytics System: An Action Design Research. *European Journal of Information Systems, 30*(4), 1-22, https://doi.org/10.1080/0960085X.2020.1811786.
28.Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity, 25*(4), 507-522, https://doi.org/10.3233/IP-200253.
29.Rahbari, E., & Shabanpoor, A. (2023). The Challenges in Employing of AI Judge in Civil Proceedings. *Legal Research Quarterly, 25*, 419-444, 10.52547/JLR.2022.228967.2335 [In Persian]
30.Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies*, 1(1), 33-36, https://doi.org/10.1002/hbe2.117.
31.Schoonderwoerd, T. A., Jorritsma, W., Neerincx, M. A., & van den Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human - Computer Studies, 154*, 102684.
32.Sein, M. K., Henfridsson, O., Purao, S., & Rossi, M. (2011). Action design research. *MIS quarterly*, 37-56, DOI:10.2307/23043488.
33.Sil, R., & Abhishek, R. (2021). Machine learning approach for automated legal text classification. *International Journal of Computer Information Systems and Industrial Management, 13*, 242-251.
34.Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66-83, https://doi.org/10.1177/0008125619862257.
35.Sowa, K., Przegalinska, A., & Ciechanowski, L. (2021). Cobots in knowledge work Human– AI collaboration in managerial professions. *Journal of Business Research, 125*, 135-142, https://doi.org/10.1016/j.jbusres.2020.11.038.
36.Tim, Y., Pan, S. L., Bahri, S., & Fauzi, A. (2017). Digitally enabled affordances for community driven environmental movement in rural Malaysia. *Information Systems Journal, 28*(1), 48-75, https://doi.org/10.1111/isj.12140.
37.Vincent, V. U. (2021). Integrating intuition and artificial intelligence in organizational decision making. *Business Horizons, 64*(4), 425-438, https://doi.org/10.1016/j.bushor.2021.02.008.
38.Watson, R. W. (1978). Interpretive structural modeling—A useful tool for technology assessment? *Technological Forecasting and Social Change, 11*(2), 165-185, https://doi.org/10.1016/0040-1625(78)90028-8.
39. Wiegreffe, S., & Marasović, A. (2021). Teach me to explain: A review of datasets for explainable natural language processing. *arXiv preprint arXiv:2102.12060*.
40. Wolf, C. T. (2019). Explainability scenarios: towards scenario-based XAI design. *24th International Conference on Intelligent User Interfaces*, (pp. 252-257).
41.Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? *Philosophy & Technology, 32*(4), 661-683, https://doi.org/10.1007/s13347-018-0330-6

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی