Journal of Public Administration
Print ISSN: 2008-5877
Online ISSN: 2423-5342

Application of artificial intelligence for Generating Smart Behavioral Nudges

Sayed Kamal Vaezi*
*Corresponding Author, Associate Prof., Department of Leadership and Human Capital, Faculty of Public Administration and Organizational Sciences, College of Management, University of Tehran, Tehran, Iran. E-mail: vaezi_ka@ut.ac.ir

Faranak Pashaei
Ph.D. Candidate, Department of Public Policy, Kish International Campus, University of Tehran, Kish, Iran. E-mail: pashaei.far@ut.ac.ir

Abstract

Objective
The primary aim of this study is to examine the capacities and capabilities of artificial intelligence in designing and implementing intelligent, personalized behavioral nudges. These nudges, viewed as the second generation of behavioral interventions, can enhance the effectiveness of behavioral policymaking by relying on technology-driven and data-based tools. The central research question guiding this study is: How can AI-based behavioral nudges, grounded in choice architecture and nudge theory, contribute to public policymaking and influence both individual and collective behaviors? By integrating big data analytics, machine learning algorithms, and cognitive technologies, the study seeks to demonstrate how behavioral interventions can be elevated from a generalized and impersonal level to one that is precise, individualized, and adaptive.

Methods
This research follows a qualitative approach using thematic analysis. Data were collected through semi-structured interviews with 12 experts in public policy, behavioral economics, and artificial intelligence. Sampling was conducted using the snowball method until theoretical saturation was achieved. The data were analyzed through open and axial coding, and the results were synthesized into a conceptual model. To enrich the model, findings from library research and international literature were incorporated. This combination of empirical

Citation: Vaezi, Sayed Kamal & Pashaei, Faranak (2025). Application of artificial intelligence for Generating Smart Behavioral Nudges. *Journal of Public Administration*, 17(4), 938-962. (*in Persian*)

Journal of Public Administration, 2025, Vol. 17, No.4, pp. 938-962
Published by University of Tehran, Faculty of Management
https://doi.org/10.22059/JIPA.2025.395752.3700
Article Type: Research Paper

Received: May 21, 2025
Received in revised form: August 08, 2025
Accepted: October 12, 2025
Published online: December 20, 2025

and theoretical data provided the foundation for developing a comprehensive conceptual architecture of an intelligent nudge system.

Results

The theoretical exploration identified technological tools such as big data, machine learning, the Internet of Things, intelligent software agents, algorithmic methods, and cognitive technologies as the central components of an intelligent nudge system. Expert interviews led to the recognition of nine complementary tools—predictive analytics, reinforcement learning, neural networks, recommender systems, notifications, fuzzy logic, natural language processing, adaptive learning platforms, and decision-support systems—that strengthen the personalization of nudges. The proposed conceptual model is built around several key components: (1) a user profile containing descriptive data (age, gender, health, location), preferences, past behaviors, and individual capabilities, gathered explicitly (e.g., user responses) or implicitly (e.g., online behavior, wearable devices); (2) a profile learner that functions as the central processor, analyzing user data to detect behavioral patterns and design context-appropriate nudges; (3) a data collection and analysis process that uses big data and algorithms such as predictive analytics and natural language processing to transform information into actionable insights; (4) nudge design, where tailored interventions are created; and (5) evaluation of user response, where feedback and behavioral changes are measured, and new data are reintegrated into the system's learning cycle to ensure continuous refinement. The findings highlight that unlike traditional nudges—which are uniform and general—AI-based nudges can be precisely tailored to individuals and delivered at the right time and in the right context. This capacity allows policymakers to move beyond broad, often inefficient interventions toward adaptive, data-driven tools. However, risks were also identified, including privacy violations, the reproduction of human biases in AI systems, and the possibility of "dark nudges." Addressing these risks requires regulatory safeguards, algorithmic transparency, and ethical oversight.

Conclusion

The integration of artificial intelligence with behavioral sciences creates new capacities for data-driven policymaking. Intelligent nudge systems not only increase the effectiveness of interventions but also provide opportunities for continuous learning, refinement, and long-term policy improvement. By presenting a conceptual model grounded in a profile learner, this study offers policymakers, developers, and behavioral researchers a roadmap for using AI to design interventions that are more efficient, equitable, and adaptive. Ultimately, AI-based nudges represent not only tools for influencing individual behavior but also a transformative step toward reimagining public policymaking in the era of big data and intelligent decision-making.

Keywords: Artificial intelligence, Behavioral nudges, Behavioral policymaking, Personalization, Behavioral science.

مدیریت دولتی
شاپا چاپی: ۵۸۷۷-۲۰۰۸
شاپا الکترونیکی: ۵۳۴۲-۲۴۲۳

کاربرد هوش مصنوعی برای تولید تلنگرهای رفتاری هوشمند

سید کمال واعظی*

* نویسنده مسئول، دانشیار، گروه رهبری و سرمایهٔ انسانی، دانشکدهٔ مدیریت دولتی و علوم سازمانی، دانشکدگان مدیریت، دانشگاه تهران، تهران، ایران. رایانامه: vaezi_ka@ut.ac.ir

فرانک پاشایی

دانشجوی دکتری، گروه خط‌مشی‌گذاری عمومی، پردیس بین‌الملل کیش، دانشگاه تهران، کیش، ایران. رایانامه: pashaei.far@ut.ac.ir

چکیده

هدف: هدف این پژوهش بررسی ظرفیت‌های هوش مصنوعی در تولید تلنگرهای رفتاری هوشمند، در راستای بهبود فرایندهای خط‌مشی‌گذاری رفتاری است. در این راستا تلاش شد تا با استفاده از چارچوب نظری معماری سامانه تلنگر هوشمند، الگویی طراحی شود که بتواند با تکیه بر ابزارهای فناورانه و شخصی‌سازی تصمیم‌ها، اثربخشی مداخلات رفتاری را ارتقا بخشد.

روش: این پژوهش از نوع کیفی و مبتنی بر تحلیل مضمون داده‌هاست که با استفاده از مصاحبه‌های نیمه‌ساختاریافته با ۱۲ نفر از خبرگان حوزه‌های خط‌مشی‌گذاری، اقتصاد رفتاری و هوش مصنوعی اجرا شده است. نمونه‌گیری به‌روش گلولهٔ برفی انجام شد و گردآوری داده‌ها تا رسیدن به اشباع نظری ادامه یافت. داده‌های حاصل از مصاحبه‌ها با بهره‌گیری از روش کدگذاری باز و محوری تحلیل و در قالب مدل مفهومی ارائه شدند. همچنین از یافته‌های نظری و تحلیل‌های کتابخانه‌ای نیز در طراحی مدل استفاده شد.

یافته‌ها: یافته‌های نظری پژوهش نشان داد که از ابزارهای مداخله‌ای مانند کلان‌داده، یادگیری ماشین، رویکرد الگوریتمی، عامل نرم‌افزار هوشمند، اینترنت اشیا و فناوری‌های شناختی، به‌عنوان اجزای اصلی سامانهٔ تلنگر هوشمند می‌توان بهره برد. نتایج حاصل از مصاحبه‌ها حاکی است که ۹ ابزار جدید سیستم‌های توصیه‌گر، یادگیری تقویتی، شبکه‌های عصبی، تحلیل پیش‌بینی‌کننده، نوتیفیکیشن، منطق فازی، پردازش زبان طبیعی، پلتفرم‌های یادگیری سفارشی و سیستم‌های پشتیبانی تصمیم‌گیری، می‌توانند قابلیت شخصی‌سازی تلنگرها را تقویت کنند. در مدل مفهومی طراحی‌شده، پردازندهٔ مرکزی به‌عنوان یادگیرنده نمایه‌ای تعریف شد که با تحلیل داده‌های مرتبط با علایق، رفتارها و توانمندی‌های کاربران، امکان تولید تلنگرهای متناسب با موقعیت‌های خاص را فراهم می‌سازد.

نتیجه‌گیری: تلنگرهای رفتاری مبتنی بر هوش مصنوعی، برای ارتقای اثربخشی مداخلات در خط‌مشی‌گذاری رفتاری ظرفیت چشمگیری دارند و به‌جای اتکا به راه‌کارهای کلی، قادرند با پردازش داده‌های فردی، مداخلاتی شخصی‌سازی‌شده و به‌موقع ارائه دهند. استفاده از معماری سامانه‌های هوشمند و یادگیری نمایه‌ای، امکان تلفیق داده‌های چندمنبعی و تحلیل آن‌ها را برای طراحی تلنگر فراهم می‌سازد. یافته‌های این پژوهش می‌تواند به خط‌مشی‌گذاران، طراحان سیستم‌های هوشمند و محققان حوزهٔ علوم رفتاری در مسیر طراحی مداخلات اثربخش‌تر کمک کند و گامی در جهت توسعه خط‌مشی‌گذاری داده‌محور و رفتارمحور باشد.

کلیدواژه‌ها: تلنگر رفتاری، خط‌مشی‌گذاری رفتاری، شخصی‌سازی، علوم رفتاری، هوش مصنوعی.

استناد: واعظی، سیدکمال و پاشایی، فرانک (۱۴۰۴). کاربرد هوش مصنوعی برای تولید تلنگرهای رفتاری هوشمند. *مدیریت دولتی*، ۱۷(۴)، ۹۳۸- ۹۶۲.

مدیریت دولتی، ۱۴۰۴، دوره ۱۷، شماره ۴، صص. ۹۳۸- ۹۶۲
ناشر: دانشکدهٔ مدیریت دانشگاه تهران
نوع مقاله: علمی پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۲/۳۱
تاریخ ویرایش: ۱۴۰۴/۰۵/۱۷
تاریخ پذیرش: ۱۴۰۴/۰۷/۲۰
تاریخ انتشار: ۱۴۰۴/۰۹/۲۹
doi: https://doi.org/10.22059/JIPA.2025.395752.3700

مقدمه

امروزه کاربرد هوش مصنوعی[1] به‌طور فزاینده‌ای در حال افزایش است (میلز[2]، ۲۰۲۳؛ آهوجا[3]، ۲۰۲۳). در عین حال به نظر می‌رسد که بسیاری از سازمان‌ها، به‌دلیل عملکرد درخور توجه هوش مصنوعی، فعالانه در تلاش برای به‌کارگیری هوش مصنوعی در فرایندهای مختلف مدیریتی و سازمانی هستند (علی و همکاران[4]، ۲۰۲۳). فرصت‌هایی که هوش مصنوعی برای علوم رفتاری ارائه می‌دهد نیز چشمگیر است. هوش مصنوعی، به‌عنوان ابزاری برای بررسی داده‌های رفتاری، شناسایی سوگیری‌های شناختی جدید یا سوگیری‌های شناختی شناخته‌شده در کانون توجه قرار گرفته است (میلز، کاستا و سانستاین[5]، ۲۰۲۳). هوش مصنوعی می‌تواند برای تجزیه‌وتحلیل سریع و دقیق مجموعه داده‌های بزرگ استفاده شود و به محققان امکان می‌دهد که الگوها و روندهایی را که در روش‌های سنتی نادیده گرفته شده‌اند، شناسایی کنند. برخلاف روش‌های سنتی مبتنی بر آمار، هوش مصنوعی می‌تواند به‌راحتی داده‌های بزرگ را مدیریت و ارتباطات پنهان را آشکار کند (ژو و همکاران[6]، ۲۰۲۱). در واقع هوش مصنوعی دستیابی به مداخلات تغییر رفتار به‌طور خودکار را نوید می‌دهد (مک آونگوسا و میچی[7]، ۲۰۲۰). هوش مصنوعی می‌تواند برای توسعه و تنظیم دقیق تلنگرهای شخصی‌سازی شده برای هر فرد، به‌طور جداگانه به‌کار گرفته شود (روگری، بنزگا، ورا و فولکه[8]، ۲۰۲۰). الگوریتم جست‌وجوی گوگل یا الگوریتم توصیهٔ یوتیوب، سامانه‌های هوش مصنوعی هستند که برای انتخاب اطلاعاتی طراحی شده‌اند که باید به‌طور برجسته به کاربران مختلف نشان داده شوند. در صنعت مالی، به‌ویژه در شرکت‌های فناوری مالی (فین‌تک)، هوش مصنوعی برای انجام وظایف مختلفی مانند تجزیه‌وتحلیل کلان‌داده و ریسک‌های اعتباری، ارائه کمک به مشتری و کاهش پول‌شویی استفاده شده است (شاربک[9]، ۲۰۲۲).

هوش مصنوعی همچنین می‌تواند با شناسایی و تعیین کمیت سوگیری‌ها در تصمیم‌گیری‌ها و تحلیل داده‌های بزرگ، الگوهای ظریف رفتاری‌ای را تشخیص دهد که ممکن است مدیران منابع انسانی آن را نبینند. این قابلیت می‌تواند به تصمیم‌سازان کمک کند تا سوگیری‌هایی را شناسایی و به آن‌ها رسیدگی کنند که پیش از این در تصمیم‌گیری‌ها به آن‌ها توجهی نشده بود (لودویگ و مولیناتان[10]، ۲۰۲۲). هوش مصنوعی در واقع به تحلیلگران رفتاری کمک می‌کند تا به قابلیت تفکر سیستمی[11] دست یابند (هالزورث[12]، ۲۰۲۳). این فرایند ممکن است از طریق پیش‌بینی زمان و زمینهٔ بهینه

1. Artificial Intelligence (AI)
2. Mills
3. Ahuja
4. Ali et al.
5. Mills, Costa & Sunstein
6. Xu et al.
7. Mac Aonghusa & Michie
8. Ruggeri, Benzerga, Verra & Folke
9. Sharbek
10. Ludwig & Mullainathan
11. Systematic Thinking
12. Hallsworth

برای ارائهٔ مداخلات طراحی شود (میلز[۱]، ۲۰۲۲؛ یئونگ[۲]، ۲۰۱۷) یا به‌شکل بررسی رفتار به‌عنوان یک سامانهٔ پیچیده برای شناسایی نقاط اهرمی بهینه برای تأثیرگذاری بر تغییر رفتار باشد (پارک و همکاران[۳]، ۲۰۲۳؛ اشمیت و استنگر[۴]، ۲۰۲۱).

علوم رفتاری مدرن در سال‌های اخیر با انتقادهای چشمگیری مواجه شده است (چاتر و لوونشتاین[۵]، ۲۰۲۲؛ مایر، بارتوش، استنلی و واگن میکرز[۶]، ۲۰۲۲)، برخی از این نقدها نیاز به رویکردهای رفتاری را که ناهمگونی را دربرمی‌گیرد، برجسته می‌کند (میلز، ۲۰۲۲؛ سزازی و همکاران[۷]، ۲۰۲۲). به همین جهت، مدیریت این «انقلاب ناهمگونی[۸]» (برایان، تیپتون و ییگر[۹]، ۲۰۲۱) با فناوری‌های هوش مصنوعی ترویج و تسریع می‌شود (میشی و همکاران[۱۰]، ۲۰۱۷؛ روتمن[۱۱]، ۲۰۲۰). باید توجه شود که تعاملات علوم رفتاری و استراتژی‌های به‌کارگیری داده‌های بزرگ، چالش‌برانگیز است (بویالسکایا و همکاران[۱۲]، ۲۰۲۳؛ دوک‌ورث و میلک‌من[۱۳]، ۲۰۲۲) و ممکن است از آن برای خطمشی‌گذاری رفتاری به نحوی استفاده شود که به نفع عامه مردم نباشد (بار گیل، سانستین و تالگام کوهن[۱۴]، ۲۰۲۳؛ دی‌مارسلیس وارن، مارتی، تلیسون و وارین[۱۵]، ۲۰۲۲). مدل‌های رفتاری هوش مصنوعی، ممکن است هزینه‌های اجتماعی چشمگیری، از قبیل به‌خطر انداختن حریم خصوصی از طریق جمع‌آوری داده‌ها (هاگندورف[۱۶]، ۲۰۲۲؛ سترا[۱۷]، ۲۰۲۰؛ صاحب[۱۸]، ۲۰۲۲) یا تداخل در شکل‌گیری ترجیحات مصرف‌کننده (بوماسانی، کریل، کومار، جورافسکی و لیانگ[۱۹]، ۲۰۲۲) را بر تصمیم‌گیرندگان تحمیل کنند. مدل‌های رفتاری هوش مصنوعی در عین حال ممکن است دست‌کاری شوند (هکر[۲۰]، ۲۰۲۱؛ سانستاین[۲۱]، ۲۰۱۵) و سوگیری‌های جدیدی را ایجاد کنند (بار گیل و همکاران، ۲۰۲۳؛ دی مارسلیس وارن و همکاران، ۲۰۲۲؛ میلز و سانستاین، ۲۰۲۳).

این مقاله ضمن تأکید بر قابلیت‌های خطمشی‌گذاری رفتاری مبتنی بر هوش مصنوعی، تلاش می‌کند تا نشان دهد هوش مصنوعی چگونه قادر است با جمع‌آوری و تحلیل کلان‌داده، سوگیری‌های رفتاری را شناسایی و مدیریت کند. در

1. Mills
2. Yeung
3. Park et al.
4. Schmidt, Stenger
5. Chater, Loewenstein
6. Maier, Bartoš, Stanley & Wagenmakers
7. Szaszi et al.
8. Heterogeneity Revolution
9. Bryan, Tipton & Yeager
10. Michie et al.
11. Rauthmann
12. Buyalskaya et al.
13. Duckworth & Milkman
14. Bar-Gill, Sunstein & Talgam-Cohen
15. de Marcellis-Warin, Marty, Thelisson & Warin
16. Hagendorff
17. Sætra
18. Saheb
19. Bommasani, Creel, Kumar, Jurafsky & Liang
20. Hacker
21. Sunstein

این میان هدف اصلی پژوهش پاسخ به این سؤال است که مداخلات رفتاری شخصی‌سازی شده یا بیش‌تلنگرهای[1] هوشمند در خط‌مشی‌گذاری رفتاری چگونه می‌توانند مبتنی بر نظریهٔ معماری انتخاب، تلنگرهای رفتاری هوشمندتری را ارائه کنند؟

پیشینهٔ نظری پژوهش

هوش مصنوعی را می‌توان فعالیتی دانست که به هوشمندسازی ماشین‌ها اختصاص می‌یابد و یک ماشین را قادر می‌سازد تا به‌درستی و با آینده‌نگری در محیط خود عمل کند (استون[2]، ۲۰۱۶). این هوش مجموعه‌ای از شبکه‌های عصبی مصنوعی[3] به تقلید از اتصال مغز انسان به نورون‌هاست و از چهار لایهٔ زیرساخت تشکیل شده است: ۱. داده‌ها، ذخیره‌سازی و قدرت محاسباتی، الگوریتم‌های یادگیری ماشین و چارچوب هوش مصنوعی؛ ۲. لایهٔ ادراک شامل حس‌های شش‌گانه؛ ۳. لایهٔ شناختی شامل توانایی استقرا، استدلال و کسب دانش با کمک پردازش زبان طبیعی، نمودارهای دانش و یادگیری مستمر؛ ۴. لایهٔ تصمیم‌گیری شامل برنامه‌ریزی خودکار، سیستم‌های خبره و سیستم‌های پشتیبان تصمیم (ژو و همکاران، ۲۰۲۱).

امروزه قدرت محاسباتی سنگین، طراحی الگوریتم‌های کارآمد، یادگیری ماشین و اینترنت اشیا و ارتباط آن‌ها با یکدیگر، به‌تدریج کل زیست‌بوم انسان را با روش‌هایی که قبلاً تصور نمی‌شد، تغییر می‌دهند (جس[4]، ۲۰۱۸). در این میان، نقش فزایندهٔ ماشین‌های خودمختار[5]، چالش‌های منحصربه‌فردی را برای خط‌مشی‌گذاری عمومی ایجاد می‌کند؛ به‌طوری که تخمین زده می‌شود که دوسوم مشاغل در کشورهای در حال توسعه، مستعد هوشمندسازی و تغییرات اساسی هستند (بانک جهانی، ۲۰۱۶). گسترش هوش مصنوعی در حوزه‌های حیاتی زندگی انسان‌ها، مانند مراقبت‌های بهداشتی، آموزش، مراقبت‌های سالمندان و حمل‌ونقل، بایستی در زمینهٔ خط‌مشی‌گذاری عمومی موجود به‌روزرسانی شوند. بر این اساس، باید تأکید کرد که در به‌کارگیری هوش مصنوعی، ارائه توصیه‌های مبتنی بر شواهد بسیار مهم است (نیولند و آرجیریادس[6]، ۲۰۱۹).

باید پذیرفت که فرایند تصمیم‌گیری مبتنی بر هوش مصنوعی غیرشفاف است و به همین علت، ممکن است فرایند تصمیم‌گیری حتی برای سازندگان الگوریتم درک نشود و باعث ترس ناشی از حاکمیت هوش مصنوعی بر تصمیم‌های انسانی شود.

مسائل اخلاقی، مسئولیت، ممیزی تصمیم و رعایت قانون از جمله چالش‌های به‌کارگیری هوش مصنوعی در حوزهٔ علوم رفتاری هستند. خبرگان خط‌مشی‌گذاری عقیده دارند که دولت‌ها باید چنین چالش‌هایی را که با حقوق اساسی و

1. Hypernudge
2. Stone
3. Artificial Neural Network, ANN
4. Jesse
5. Autonomous
6. Chester Newland and Demetrios Argyriades

مردم در تضاد است، به‌درستی مدیریت کنند (بوتنکو و لاروش[1]، ۲۰۱۷). با درک اهمیت استراتژیک هوش مصنوعی، دولت‌های سراسر جهان، توسعهٔ هوش مصنوعی را اولویت ملی خود می‌دانند و منابع عظیمی را با هدف تبدیل شدن به رهبر جهانی در هوش مصنوعی صرف می‌کنند. دولت آمریکا از طریق «طرح ملی هوش مصنوعی»[2] سرمایه‌گذاری کلانی در تحقیق و توسعه، امنیت داده‌ها و استفاده از هوش مصنوعی در حوزه‌های دفاعی و اقتصادی کرده است. اتحادیهٔ اروپا با تمرکز بر توسعهٔ اخلاقی و قانونمند هوش مصنوعی، «لایحهٔ قانون هوش مصنوعی»[3] را با هدف مدیریت خطرهای AI و اطمینان از هم‌راستایی آن با ارزش‌های دموکراتیک اروپایی تصویب کرده است. چین نیز با راه‌اندازی «طرح توسعهٔ هوش مصنوعی نسل جدید»[4]، در تلاش است تا سال ۲۰۳۰ به رهبر جهانی در هوش مصنوعی تبدیل شود. شود. سرمایه‌گذاری‌های عظیم دولتی، حمایت از شرکت‌های فناوری و استفادهٔ گسترده از AI در حوزه‌های امنیتی، شهری و اقتصادی، از ویژگی‌های استراتژی این کشور است. در بسیاری از کشورها مانند ایران و فرانسه، ساختارهای تصمیم‌ساز در سطح عالی برای امور هوش مصنوعی تعریف شده است. نکتهٔ مهم این است که با توجه به قدرت محاسباتی در دسترس الگوریتم‌های هوشمند، امکان استفادهٔ مخرب از هوش مصنوعی دور از ذهن نیست و لازم است که مبتنی بر یک خط‌مشی نظام‌مند با ابعاد چندگانهٔ گسترش این تهدیدها مقابله شود (وندرلست و وینفیلد[5]، ۲۰۱۸).

تلنگرهای رفتاری

نظریهٔ تلنگر روشی برای درک بهتر رفتار تصمیم‌گیران است (تالر و سانستاین[6]، ۲۰۰۸). این نظریه بر مبنای مفهوم اقتصاد رفتاری (وونگ، هو، نگوین و وونگ،[7] ۲۰۱۸)، پیامدهای عقلانیت محدود مبتنی بر مؤلفهٔ ترجیحات اجتماعی را شرح می‌دهد. بر اساس این نظری،ه محیط را می‌توان به طریقی تغییر داد تا افراد را به‌سمت انتخاب بهتر هدایت کند. به‌عنوان یک فرض اساسی، نظریهٔ تلنگر بر اساس تمایز بین دو الگوی تفکر تعقلی و تکانشی (احساسی)[8] طراحی شده است (تالر و سانستاین، ۲۰۰۸). هنگامی که افراد، منطقی عمل می‌کنند، به انگیزه‌های اقتصادی واکنش نشان می‌دهند و انتخاب‌های متفکرانه‌ای انجام می‌دهند؛ اما متفکران تکانشی، یک گزینه رضایت‌بخش و نه کاملاً بهینه را انتخاب می‌کنند (کانمن و تورسکی[9]، ۱۹۷۹). سوگیری‌ها این فرایندهای تصمیم‌سازی شهودی را تحت تأثیر خود قرار می‌دهند.

در این میان مفهوم معماری انتخاب نشان می‌دهد که در تفکر تکانشی، آنچه افراد انتخاب می‌کنند، اغلب به نحوهٔ ارائهٔ انتخاب بستگی دارد (جانسون و همکاران[10]، ۲۰۱۲: ۴۸۸). معماری انتخاب شامل هر مداخلهٔ خارجی از قبیل ساختار، طراحی و چیدمان است که می‌تواند تصمیم‌ها را هدایت کند و بر آن تأثیر بگذارد. یک معماری انتخاب، مستلزم یک

1. Butenko, Larouche
2. National AI Initiative
3. AI Act
4. New Generation AI Development Plan
5. Vanderelst, Winfield
6. Thaler and Sunstein
7. Vuong, Ho, Nguyen & Vuong
8. Thoughtful and Impulsive
9. Kahneman & Tversky
10. Johnson et al.

معمار انتخاب است که برای تغییر رفتار به‌روش پیش‌بینی‌شده‌ای، تلنگرهایی را در معماری ایجاد کند (تالر و سانستاین، ۲۰۰۸).

معمار انتخاب از مجموعهٔ گسترده‌ای از ابزارها استفاده می‌کند. تلنگرهای معمول و سنتی پیش‌فرض که به افراد گزینه‌هایی ارائه می‌کنند و با اهداف و افکار انعکاسی آنان هم‌خوانی دارند، عبارت‌اند از: تلنگر انتظار خطا که خطاهای انسانی را پیش‌بینی می‌کند؛ تلنگر بازخورد که به تصمیم‌گیری کمک می‌کند؛ تلنگرهای مشوق که برای انجام انتخاب‌های خاص مشوق‌هایی ارائه می‌کند و تلنگر ساختاربندی که انتخاب‌های پیچیده را سامان‌دهی می‌کند (جانسون و همکاران، ۲۰۱۲).

در خصوص نحوهٔ تأثیر تلنگرهای رفتاری، الگوهای مختلفی پیشنهاد شده است. هانسن و جسپرسن[1] (۲۰۱۳) طبقه‌بندی تلنگرهای نوع ۱ را که هدف آن تأثیرگذاری بر رفتارهای خودکار و تفکر خودکار است و تلنگرهای نوع ۲ که افراد را از تصمیم‌های خود آگاه می‌کند، پیشنهاد می‌کنند. سپس یک تمایز معرفتی بین تلنگرهای شفاف و غیرشفاف به‌وجود می‌آورند. تلنگرهای اطلاعاتی، ساختار تصمیم‌گیری و کمک تصمیم‌گیری را پیشنهاد می‌کنند که جنبه‌های مختلف فرایند تصمیم‌گیری را منعکس می‌کنند (مونشر، وتر و شئوارل[2]، ۲۰۱۶). براساس کارکردهای شناختی، ارادی و عاطفی، سه درجه تلنگر را می‌توان متمایز کرد. تلنگر درجهٔ ۱، شامل ارائهٔ اطلاعات ساده یا یک یادآوری؛ تلنگر درجهٔ ۲، شامل محدودیت‌های رفتاری یا ارادی و تلنگر درجهٔ ۳، شامل استراتژی‌های چارچوب، پاسخ‌های احساسی، استفاده از تکنیک‌های پنهان برای تأثیرگذاری بر تصمیمات یا شکل دادن ترجیحات است.

از سوی دیگر، هاگمن، اندرسون، واستفیال و تینگهوک[3] (۲۰۱۵) تلنگرهای خودمحور در مقابل تلنگرهای اجتماع‌محور را مطرح می‌کنند. تلنگرها با طراحی و اعمال مناسب، به یک معمار انتخاب اجازه می‌دهند که توجه کاربر را به یک جنبهٔ انتخاب خاص معطوف و اکتشاف‌های خاص و ارتباط‌های مربوطه را آغاز کند (میشالک، مران، شوارز، ایلدیز[4]، ۲۰۱۶).

باید توجه شود که تلنگرها در موقعیت‌هایی که شامل رفتار معمولی یا درگیری کم است، بهترین کارکرد را دارند (مونت، نوونن و لاهتینوخا[5]، ۲۰۱۴). بر این اساس می‌توان گفت یک تلنگر خوب، ابزارها را دست‌کاری می‌کند و نه اهداف را، از جنس مداخله شناخته می‌شود، آزادی انتخاب را در گزینه‌های ارائه‌شده در نظر می‌گیرد و در نهایت اینکه بر مفهوم پاداش یا هزینه انتخاب تأکید کمی دارد (هالپرن[6]، ۲۰۱۵؛ مله، اسپنا، کارتمو و مارزولو[7]، ۲۰۲۱).

رفتارگرایی معتقد است که رفتار انسان تابعی از محرک‌های بیرونی است و رویدادهای درونی و عصبی ممکن است بدون از دست‌دادن بینشی از رفتار سوژهٔ مدنظر نادیده گرفته شوند. شباهت‌های مختلفی بین هوش مصنوعی معاصر در

1. Hansen and Jespersen
2. Münscher, Vetter, and Scheuerle
3. Hagman, Andersson, Vastfjall, and Tinghog
4. Michalek, Meran, Schwarze, Yildiz
5. Mont, Neuvonen & Lahteenoja
6. Halpern
7. Mele, Spena, Kaartemo & Marzullo

علوم رفتاری و رفتارگرایی وجود دارد. این دقیقاً همان هدفی است که با ابزارهای تلنگری به‌دنبال آن هستیم (اسمیت و دی ویلیرز بوتا[1]، ۲۰۲۱).

اسکینر[2]، اعتقاد دارد که رفتار انسان محصول محرک‌های محیطی است و با دست‌کاری این محرک‌ها با توجه اندک به محیط درونی، می‌توان رفتار را تغییر داد. بر این اساس می‌توان گفت که رفتارگرایی نظریهٔ روان‌شناختی مبتنی بر یادگیری تقویتی و تجربهٔ ذهنی (بچتل، آبراهامسن، و گراهام[3]، ۲۰۱۷) است که می‌تواند برای توسعه‌دهندگان هوش مصنوعی بسیار مفید باشد (مک‌کارتی، مینسکی، روچستر و شانون[4]، ۱۹۵۵؛ سیمون[5]، ۱۹۹۴).

تلنگرها ابزارهای ساده و کم‌هزینه‌ای در سازمان‌ها هستند که با تغییر معماری انتخابی که در آن تصمیم گرفته می‌شود، رفتار افراد را تغییر می‌دهند. نکتهٔ مهم این است که تلنگرها هیچ چیزی را ممنوع نمی‌کنند یا انگیزه‌ها را تغییر نمی‌دهند (تالر و سانستاین، ۲۰۲۱). در محیط‌های دیجیتال، تلنگرها معماران انتخابی هستند که بر تصمیم‌گیری‌های افراد تأثیر می‌گذارند (واینمن، اشنایدر و ووم بروک[6]، ۲۰۱۶). تلنگرهای دیجیتال برای تولید تلنگرهای شخصی‌سازی شدهٔ شدهٔ هوشمند قابلیت چشمگیری دارند. بر این اساس، فرایند طراحی با توجه به ویژگی‌های منحصربه‌فرد افراد، انجام می‌شود (اشنایدر، وایـنمن و وم‌بروک[7]، ۲۰۱۸؛ وایـنمن و همکاران، ۲۰۱۶). حتی این تلنگرها می‌توانند رفتارهای فضیلت‌آمیز را تشویق کنند (یئونگ، ۲۰۱۷). این گزینه‌ها با ردیابی و تجزیه‌وتحلیل رفتار کاربر در زمان واقعی و همچنین شخصی‌سازی در رابط کاربری، می‌توانند اثربخشی تلنگرهای دیجیتال را بهینه کنند. آن‌ها اطلاعات، یادآوری‌ها یا اعلان‌های خودکار مانند پیام‌های نوار وضعیت، پنجره‌های بازشو، لرزش تلفن یا نمایشگرهای LED را ارائه می‌کنند (اوککه، سوبولو، دل، استرین[8]، ۲۰۱۸). تلنگرهای دیجیتال، علاوه‌بر اینکه نسبتاً ساده و شهودی هستند، می‌توانند از رفتارهای گروهی قابل قبول اجتماعی تغذیه کنند و به‌سرعت در بین گروه‌ها پخش شوند تا افراد را به تفکر یا عمل متفاوت ترغیب کنند (یئونگ، ۲۰۱۷). با این مزایا، سامانه‌های هوشمند امکان اصلاح و تجسم سریع محتوا را برای دستیابی به اثر تلنگر مدنظر فراهم می‌کنند. در واقع، می‌توان گفت که عامل نرم‌افزاری هوشمند با استفاده از تداعی‌های عاطفی، اثرهای اجتماعی یا سایر روش‌ها بر تصمیم‌گیری تأثیر می‌گذارد (بور، کریستیانینی و لیدیمن[9]، ۲۰۱۸).

عامل نرم‌افزاری هوشمند ویژگی‌های شخصیتی کاربران را از سیگنال‌های رفتاری آنان یاد می‌گیرد و از این اطلاعات برای تطبیق پیشنهادهای محیط با ترجیحات مصرف‌کنندگان در لحظهٔ مناسب استفاده می‌شود. در آزمایش‌های میدانی، ماتز، کوسینسکی، ناو و استیلول[10] (۲۰۱۷) دریافتند که جذابیت‌های متقاعدکننده‌ای که با نیازهای روان‌شناختی

1. Smith and de Villiers-Botha
2. Burrhus Skinner
3. Bechtel, Abrahamsen, and Graham
4. McCarthy, Minsky, Rochester & Shannon
5. Simon
6. Weinmann, Schneider & vom Brocke
7. Schneider, Weinmann, vom Brocke
8. Okeke, Sobolev, Dell, Estrin
9. Burr, Cristianini, and Ladyman
10. Matz, Kosinski, Nave, and Stillwell

مخاطبان هدف مطابقت دارد، به ۴۰درصد کلیک بیشتر و ۵۰درصد خرید بیشتر از تبلیغات ناهماهنگ یا غیرشخصی منجر می‌شود. علاوه‌براین، عامل نرم‌افزاری هوشمند می‌تواند انسان را به‌سمت تصمیم‌هایی که قبلاً نمی‌گرفت، هدایت کند. این تصمیم‌ها می‌توانند مانند مداخلات سلامتی مثبت، سودمند باشند یا مانند تصمیم‌گیری برای ادامهٔ مصرف بیش از حد سرگرمی‌های برخط مضر باشند. با این حال، تحقیقات نوظهور مرتبط با فناوری‌های شناختی و تلنگرها هنوز در مراحل ابتدایی خود هستند (مله و همکاران، ۲۰۲۱)، بنابراین به‌صورت تجربی خط‌مشی‌گذاری رفتاری مبتنی بر هوش مصنوعی که می‌تواند رفتار فردی را تغییر دهد، باید مورد مطالعه قرار گیرد. برای مثال، پوشیدنی‌های هوشمند، ابزارهای هوشمند، عوامل مکالمه/روبات‌های اجتماعی و پلتفرم‌های هوشمند چهار گروه از راه‌حل‌های مبتنی بر فناوری‌های شناختی هستند که معماری‌های انتخاب متفاوتی را برای هم‌آفرینی ارزش به وجود می‌آورند (مله و همکاران، ۲۰۲۰). باید توجه کرد که ماهیت رفتار هوشمند، آن طور که در رفتار انسانی دیده می‌شود، در رفتار ماشینی منتج از هوش مصنوعی دیده نمی‌شود (میلز، ۲۰۲۳).

سوگیری‌های رفتاری را می‌توان به‌عنوان الگوها یا خطاهای پیش‌بینی‌شدنی در رفتار انسان تلقی کرد (کانمن[1]، ۲۰۱۱؛ تالر و سانستاین، ۲۰۰۸). مطالعات نشان می‌دهد که قابلیت‌های تشخیص الگوی هوش مصنوعی برای شناسایی سوگیری‌های ناشی از داده‌های رفتاری مناسب است (کلینبرگ، لودویگ، مولیناتان و اوبرمایر[2]، ۲۰۱۵؛ لودویگ و مولیناتان، ۲۰۲۱). هوش مصنوعی می‌تواند سوگیری‌هایی را شناسایی کند که قبلاً هرگز شناسایی نشده‌اند (لودویگ و مولیناتان، ۲۰۲۱). در عین حال هوش مصنوعی می‌تواند نویزهای رفتاری را شناسایی و کمّی کند.

بر این اساس، به نظر می‌رسد که هوش مصنوعی برای پیشرفت قابلیت مداخلات رفتاری فرصتی منحصربه‌فرد ارائه می‌دهد (هالزورث، ۲۰۲۳). این نوآوری‌ها در مورد اثربخشی مداخلات رفتاری (مایر و همکاران، ۲۰۲۲) و با توجه به توقع تأثیرگذاری‌های محدود (بشیرز و کوزوسکی[3]، ۲۰۲۰؛ دلاویگنا و کوزوسکی[4]، ۲۰۲۲) مطرح شده است. هوش مصنوعی در واقع در بستر علوم رفتاری، پیچیدگی ذاتی رفتار انسان را پذیرفته و به درک رفتار به‌عنوان بخشی از یک سامانه انطباقی پیچیده اشاره می‌کند (هالزورث، ۲۰۲۳).

بر این اساس می‌توان مدعی شد که رویکردهای سایبرنتیک، تصمیم‌ساز را تشویق می‌کند تا رفتار را به‌عنوان بخشی از یک سامانهٔ پیچیده و مجموعه‌ای از نقاط اهرمی متغیرهای رفتاری درک کند (بیر[5]، ۱۹۹۳؛ فورستر[6]، ۱۹۷۱؛ آبسون و همکاران[7]، ۲۰۱۷؛ لونتون، آبسون و لانگ[8]، ۲۰۲۱؛ اشمیت و استنگر، ۲۰۲۱). در یک سامانهٔ پیچیدهٔ رفتاری، این متغیرها تأثیر بزرگی بر سامانهٔ مداخلات رفتاری خواهند داشت (آبسون و همکاران، ۲۰۱۷؛ هالزورث، ۲۰۲۳؛ وست،

1. Kahneman
2. Kleinberg, Ludwig, Mullainathan & Obermeyer
3. Beshears & Kosowsky
4. DellaVigna & Linos
5. Beer
6. Forrester
7. Abson et al.
8. Leventon, Abson & Lang

میچی، چادویک، آتکینز و لورنکاتو[1]، ۲۰۲۰؛ کوماکی، کداک، ناکامورا، اونو و کوهتاکه[2]، ۲۰۲۱؛ میدوز[3]، ۱۹۹۷؛ سیمون، ۱۹۸۱). شفافیت ماهیت نقاط اهرمی باعث شده است که مفهوم «حرکات موزون رفتاری[4]» مطرح شود (میدوز، ۲۰۰۱). هوش مصنوعی برای نقشه‌برداری سامانه‌های رفتاری و شناسایی نقاط اهرمی رویکردی امیدوارکننده نشان می‌دهد (نگ[5]، ۲۰۱۶) که به سهم خود ممکن است اثربخشی مداخلات رفتاری را افزایش دهد (هالزورث، ۲۰۲۳؛ اشمیت و استنگر، ۲۰۲۱). میلز (۲۰۲۲) یک چارچوب دو جزئی را برای شخصی‌سازی تلنگرها پیشنهاد می‌کند که شامل انتخاب نوع تلنگر استفاده شده و شخصی‌سازی آن براساس داده‌های ناهمگون است. میلز استدلال می‌کند که معماران انتخاب از بین راهبردهای تلنگر و گزینه‌های قابل تلنگر دست به انتخاب می‌زنند. این تلنگرهای شخصی‌سازی‌شده را می‌توان به‌عنوان تلنگرهای شخصی‌شده «خام»، با بیش‌تلنگرهای «بسیار پیچیده» توصیف‌شده توسط یئونگ (۲۰۱۷) مقایسه کرد. بر این اساس، می‌توان گفت هوش مصنوعی می‌تواند فرایند شخصی‌سازی تلنگرها را تسریع کند. این فرایند بر اساس دسترسی به داده‌های ناهمگون و احتمالاً داده‌های بزرگ صورت می‌پذیرد. این فرایند شخصی‌سازی در چهار مرحلهٔ شخصی‌سازی انتخاب، شخصی‌سازی تحویل، شخصی‌سازی خام و شخصی‌سازی پیچیده صورت می‌پذیرد (میلز، ۲۰۲۳). در این میان باید توجه شود که تلنگرهای شخصی‌سازی شده توسط هوش مصنوعی، به‌شدت از فرایندهای شناختی تأثیر می‌پذیرد که توسط عامل انسانی در اختیار هوش مصنوعی قرار گرفته است؛ از این رو ضرورت دارد که قبل از هرگونه تکیه بر تلنگرهای تولیدشده توسط هوش مصنوعی، از عدم سرایت سوگیری‌های انسانی در نحوهٔ تعریف این فرایندهای شناختی اطمینان حاصل کرد (اشمادر، کارپوس، مول، بهرامی و دروی[6]، ۲۰۲۳).

پیشینهٔ تجربی

باید پذیرفت که کاربرد هوش مصنوعی در حوزهٔ علوم رفتاری و تلنگرهای هوشمند، در سال‌های اخیر توجه فراوانی را در مجامع علمی و خط‌مشی‌گذاری به خود جلب کرده است. در جدول ۱ به بخشی از این مطالعات اشاره شده است.

جدول ۱. بعضی از مطالعات جدید در حوزه به‌کارگیری هوش مصنوعی در حوزه مداخلات رفتاری

نویسنده و سال نشر	موضوع	خلاصه نتایج پژوهش
میلز (۲۰۲۳)	هوش مصنوعی برای علوم رفتاری	اهمیت هوش مصنوعی برای تأثیرگذاری بر رفتارهای فردی از طریق تغییر زمینه‌های تصمیم‌گیری
میلز، کوستا و سانستین[7] (۲۰۲۳)	هوش مصنوعی، علوم رفتاری و رفاه مصرف‌کننده	بررسی تأثیر هوش مصنوعی بر رفتار مصرف‌کنندگان از منظر علوم رفتاری و چگونگی تصمیم‌گیری با استفاده از الگوریتم‌ها.

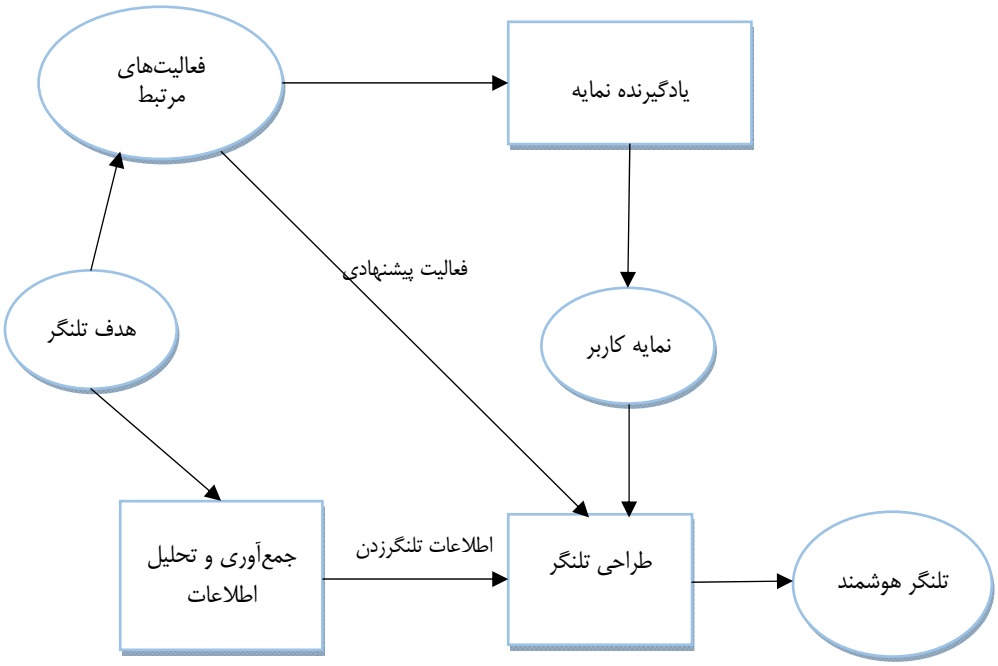
1. West, Michie, Chadwick, Atkins & Lorencatto
2. Komaki, Kodaka, Nakamura, Ohno & Kohtake
3. Meadows
4. Choreography
5. Ng
6. Schmauder, Karpus, Moll, Bahrami & Deroy
7. Mills, Costa & Sunstein

نویسنده و سال نشر	موضوع	خلاصه نتایج پژوهش
شین و احمد[1] (۲۰۲۳)	تلنگر الگوریتمی: رویکردی برای طراحی هوش مصنوعی مولد انسان‌محور	تغییرات اساسی در زیست‌بوم رسانه‌های اجتماعی برای یافتن مداخلات مؤثر
آدوماویسیوس و یانگ[2] (۲۰۲۲)	ادغام بینش‌های رفتاری، اقتصادی و فنی برای درک و رسیدگی به سوگیری الگوریتمی: دیدگاه انسان‌محور	ارائه راه‌کارهایی برای بهبود عدالت و شفافیت الگوریتم‌ها با ادغام بینش‌های رفتاری، اقتصادی و فنی.
سورا، ریبیرو سوریانو و زگارا سالدانیا[3] (۲۰۲۲)	ارزیابی مسائل مربوط به حریم خصوصی علم داده رفتاری در استقرار هوش مصنوعی دولتی	چالش‌های استقرار هوش مصنوعی در بخش دولتی و مخاطرات اخلاقی، شفافیت و حفاظت از داده‌های شهروندان.
میلز و سترا[4] (۲۰۲۲)	معمار انتخاب خودمختار	هدایت انتخاب‌های کاربران بدون محدودکردن آزادی فردی و تعامل میان خودمختاری، طراحی انتخاب و تأثیرات اخلاقی
میلز (۲۰۲۲)	تلنگرهای خودمختار و معماران انتخاب هوش مصنوعی ـ مسئولیت تصمیم‌گیری با واسطه رایانه کجاست؟	لنگرهای خودمختار و نقش هوش مصنوعی در معماری انتخاب، چالش‌های اخلاقی و حاکمیتی مرتبط با واسطه‌های رایانه‌ای در تصمیم‌گیری انسان‌ها
بروکز و همکاران[5] (۲۰۲۲)	استفاده شرکت‌های غذا و نوشیدنی فراملیتی از هوش مصنوعی برای فعال کردن تلنگرهای تاریک	استفاده از هوش مصنوعی برای ایجاد تلنگرهای تاریک و هدایت رفتار مصرف‌کنندگان به‌طور نامحسوس به سمت انتخاب‌های ناسالم و پیامدهای آن را برای سلامت عمومی و سیاست‌گذاری.
بالاسوبرامانیان[6] (۲۰۲۱)	وقتی هوش مصنوعی با اقتصاد رفتاری روبه‌رو می‌شود	کاربردهای هوش مصنوعی در بهینه‌سازی تصمیم‌گیری، شخصی‌سازی انتخاب‌ها و طراحی تلنگرهای دیجیتال و چالش‌های اخلاقی و سیاست‌گذاری مرتبط با آن.
پترسون، بورگین، آگراوال، رایشمن و گریفیتس[7] (۲۰۲۱)	استفاده از آزمایش‌های مقیاس بزرگ و یادگیری ماشین برای کشف نظریه‌های تصمیم‌گیری انسانی	توسعه نظریه‌های تصمیم‌گیری انسانی. ترکیب روش‌های تجربی و الگوریتم‌های یادگیری ماشین برای شناسایی الگوهای رفتاری
مک آونگوسا و میچی (۲۰۲۰)	هوش مصنوعی و علوم رفتاری از طریق لنز: چالش‌هایی برای کاربرد در دنیای واقعی	چالش‌های استفاده از هوش مصنوعی در علوم رفتاری و کاربرد آن در دنیای واقعی از جمله مسائل اخلاقی، پیچیدگی رفتار انسانی و مشکلات پیاده‌سازی در مقیاس وسیع.
مله و همکاران (۲۰۲۰)	تلنگر هوشمند: چگونه فناوری‌های شناختی، معماری‌های انتخابی را برای خلق مشترک ارزش امکان‌پذیر می‌سازند.	نقش فناوری‌های شناختی در بهبود معماری‌های انتخاب و ایجاد ارزش مشترک در فرایندهای همکاری.
روبیلا و روبیلا[8] (۲۰۱۹)	کاربرد روش‌های هوش مصنوعی در علوم رفتاری و اجتماعی	کاربرد روش‌های هوش مصنوعی در تحلیل رفتارهای انسانی و اجتماعی در پیش‌بینی و مدل‌سازی رفتارهای فردی و جمعی.

1. Shin & Ahmad
2. Adomavicius & Yang
3. Saura, Ribeiro-Soriano & Zegarra Saldaña
4. Mills & Sætra
5. Brooks et al.
6. Balasubramanian
7. Peterson, Bourgin, Agrawal, Reichman & Griffiths
8. Robila & Robila

مدل مفهومی

مدل مفهومی این پژوهش با اقتباس از مدل کارلسن و اندرسون[1] (۲۰۱۹) طراحی شده است. بر این اساس و با نقد این نگاه که تمایلات گذشته، اغلب شاخص‌های خوبی برای انتخاب‌های آینده هستند (آگاروال[2]، ۲۰۱۶)، به کاربران توصیه می‌شود که مسیرهای بهتر و اهداف تازه‌تری در تصمیم‌گیری‌های خود اعمال کنند. چارچوب نظری این پژوهش برمبنای مدل عملیاتی معماری یک سامانهٔ تلنگر هوشمند، برای شکل‌گیری و ارزیابی تلنگرهای مبتنی بر هوش مصنوعی تعریف می‌شود.



شکل ۱. معماری یک سامانهٔ تلنگر هوشمند

منبع: کارلسن و اندرسن (۲۰۱۹)

همان طور که در شکل ۱ مشاهده می‌شود، برای داشتن تلنگر هوشمند که در آن تلنگرها براساس نیازها و موقعیت یک کاربر خاص شخصی‌سازی می‌شوند، نمایهٔ کاربر مهم است. نمایه را یادگیرندهٔ نمایه ایجاد می‌کند و اطلاعات توصیفی کاربر (مانند سن، جنسیت، علایق و رفتار) را جمع‌آوری می‌کند. یادگیرندهٔ نمایه همچنین بر فعالیت‌های انتخاب‌شده قبلی برای تعیین انتخاب(های) ترجیحی کاربر و قابلیت‌های کاربر نظارت می‌کند.

بر اساس مدل تلنگر هوشمند کارلسن و اندرسون، هدف از طراحی یک الگوی هوشمند، شناسایی و درک نحوهٔ تصمیم‌سازی یک رفتارکننده، بر اساس رصد فعالیت‌های پیشین و فعلی اوست. در این راستا، هوش مصنوعی با ایجاد ابزاری تحت عنوان یادگیرندهٔ نمایه‌ای و بر اساس چهار مؤلفه داده‌های شخصی، علائق، رفتار و قابلیت‌ها، یک فعالیت

1. Karlsen & Andersen
2. Aggarwal

انتخابی را به کاربر توصیه می‌کند. در صورت واکنش مثبت کاربر به این فعالیت، طراحی تلنگر هوشمند و شخصی‌سازی شده صورت می‌گیرد. به عبارت دیگر، تلنگر هوشمند از تلفیق ویژگی‌های یک تلنگر عمومی با ویژگی‌های شاخص کاربر مورد نظر صورت می‌پذیرد. برای نمونه در خصوص حمل‌ونقل و رفتارهای ترافیکی، به‌جای ایجاد تلنگرهای عمومی برای همهٔ کاربران، یک تلنگر هوشمند توسط یادگیرنده نمایه‌ای پیشنهاد و به‌کار بسته می‌شود. خروجی این سیستم می‌تواند تلنگری باشد که یک کاربر مشخص را در یک روز مشخص، به استفاده از دوچرخه ترغیب کند؛ در حالی که کاربر دیگر را به پیاده‌روی توصیه کند. یا در خصوص تردید در حضور یا غیبت کارمندی که از یک حملهٔ عصبی در یک روز خاص رنج می‌برد یک تلنگر شخصی‌سازی شده می‌تواند راه‌گشا باشد. این یادگیرنده نمایه‌ای، در واقع همواره در حال رصد مسیرهای ترافیکی یا سلامت روانی کارکنان و تلفیق آن با ویژگی‌های فردی کاربر خاص است.

مؤلفهٔ طراحی تلنگر، حرکات شخصی‌سازی شده‌ای را ایجاد می‌کند که با هدف مرتبط است و برای کاربر خاص طراحی شده است. این مؤلفه اطلاعاتی را ارائه می‌دهد که برای انتخاب فعالیت پیشنهادی مفید است. تلنگر همچنین باید به‌موقع و مرتبط با تصمیماتی باشد که کاربر باید اتخاذ کند. نتیجهٔ طراحی تلنگر، تلنگر هوشمندی است که به کاربر ارائه می‌شود. ارزیابی تلنگر شامل نظارت بر واکنش کاربر به تلنگر و شناسایی فعالیت انتخاب‌شده توسط کاربر است. اثربخشی تلنگر، به‌عنوان ورودی برای یادگیرندهٔ نمایه استفاده می‌شود و در نمایهٔ کاربر منعکس می‌شود. خود تلنگر و نتیجه ذخیره می‌شوند و به‌عنوان یک پایگاه دانش برای تلنگر بیشتر استفاده می‌شوند.

روش‌شناسی پژوهش

در این پژوهش از روش کیفی مصاحبه با خبرگان استفاده شده است. مصاحبه یکی از ابزارهای جمع‌آوری دادهٔ کیفی محسوب می‌شود. با کمک این ابزار، ارتباط مستقیم با مصاحبه‌شونده برقرار می‌شود؛ همچنین می‌توان به ارزیابی عمیق‌تر ادراک‌ها، نگرش‌ها و علایق آزمودنی‌ها پرداخت. مصاحبه ابزاری است که امکان بررسی موضوع‌های پیچیده، پیگیری پاسخ‌ها یا پیداکردن علل آن و اطمینان یافتن از درک سؤال از سوی آزمودنی را فراهم می‌کند. جامعهٔ آماری پژوهش از صاحب‌نظران و خبرگان دانشگاهی در حوزهٔ خط‌مشی‌گذاری و آشنا به اقتصاد رفتاری و هوش مصنوعی تشکیل شده است که با روش گلولهٔ برفی و به تعداد ۱۲ نفر انتخاب شدند.

گردآوری داده‌ها از طریق فرایند اشباع نظری تا زمان وصول به نقطهٔ اشباع و اطمینان از تکاملی و تعاقبی بودن انتخاب نمونه‌ها ادامه یافت و پس از اطمینان از کفایت تعداد مصاحبه‌شوندگان، ابزارهای مداخله‌ای هوش مصنوعی برای ساخت تلنگرهای هوشمند دراختیار آنان قرار گرفت. در جدول ۲ یافته‌های کتابخانه‌ای و در جدول ۳ یافته‌های حاصل از مصاحبه ارائه شده است.

یافته‌های پژوهش

همان طور که در جدول ۲ مشاهده می‌شود، ۶ ابزار شامل کلان‌داده، یادگیری ماشین، رویکرد الگوریتمی، عامل نرم‌افزاری، اینترنت اشیا و فناوری‌های شناختی به‌عنوان ابزارهای هوش مصنوعی برای ساخت تلنگرهای هوشمند

شناسایی شد. در فاز مطالعات میدانی، از مصاحبه‌شوندگان درخواست شد تا ضمن ارزیابی قابلیت این ۶ ابزار مداخله‌ای هوش مصنوعی، ابزارهای مکمل و جدید را نیز شناسایی کنند که نتایج آن در جدول ۳ آورده شده است.

جدول ۲. یافته‌های نظری ابزارهای مداخله‌ای هوش مصنوعی برای ساخت تلنگرهای هوشمند

ابزارهای مداخله‌ای هوش مصنوعی برای ساخت تلنگرهای هوشمند	کد اختصاری	نحوه مداخله	استناد
کلان‌داده[1]	BD	کلان‌داده در خصوص ابعاد مختلف زندگی یک فرد، امکان ساخت تلنگرهای کاملاً شخصی‌سازی شده و با هدف یکتا برای تغییر رفتار فرد مورد نظر را فراهم می‌کند.	پیترسون و همکاران، ۲۰۲۱؛ روگری و همکاران، ۲۰۲۰؛ مله و همکاران، ۲۰۲۰؛ جس، ۲۰۱۸
یادگیری ماشین[2]	ML	روش‌های یادگیری ماشین، این قابلیت را ایجاد می‌کند که الگوی رفتاری فرد در ورودی داده شناسایی و تلنگرهای شخصی‌سازی شده مؤثرتری ایجاد کرد.	ژو و همکاران، ۲۰۲۱؛ روگری و همکاران، ۲۰۲۰
رویکرد الگوریتمی به مسئله[3]	AL	الگوریتم‌ها با دریافت کلان‌داده و تحلیل آن می‌توانند الگوهای رفتاری را شناسایی و نیز پیش‌بینی کنند. همچنین می‌توانند مداخلاتی را بیابند که بر رفتار کاربر در بلندمدت تأثیر بگذارد.	شین و احمد، ۲۰۲۳؛ روبیلا و روبیلا ۲۰۱۹
عامل نرم‌افزار هوشمند[4]	IA	مجموعه‌ای از ابزارها و الگوریتم‌های آماری که رایانه‌ها را قادر می‌سازد تا عناصر رفتار انسانی مانند یادگیری، استدلال و طبقه‌بندی را شبیه‌سازی کنند.	ماتز؛ کوسینسکی؛ ناو و استیلول، ۲۰۱۷، بور، کریستیانینی و لیدیمن، ۲۰۱۸
اینترنت اشیا[5]	IoT	اینترنت اشیا می‌تواند با تمرکز بر یادگیری ماشین انتقال و دریافت داده‌های مربوط به شخص را انجام می‌دهد.	جس، ۲۰۱۸
فناوری‌های شناختی[6]	CT	پوشیدنی‌های هوشمند، ابزارهای هوشمند، روبات‌های اجتماعی و پلتفرم‌های هوشمند چهار گروه از راه‌حل‌های مبتنی بر فناوری‌های شناختی هستند که معماری‌های انتخاب متفاوتی را به وجود می‌آورند.	مله و همکاران، ۲۰۲۱

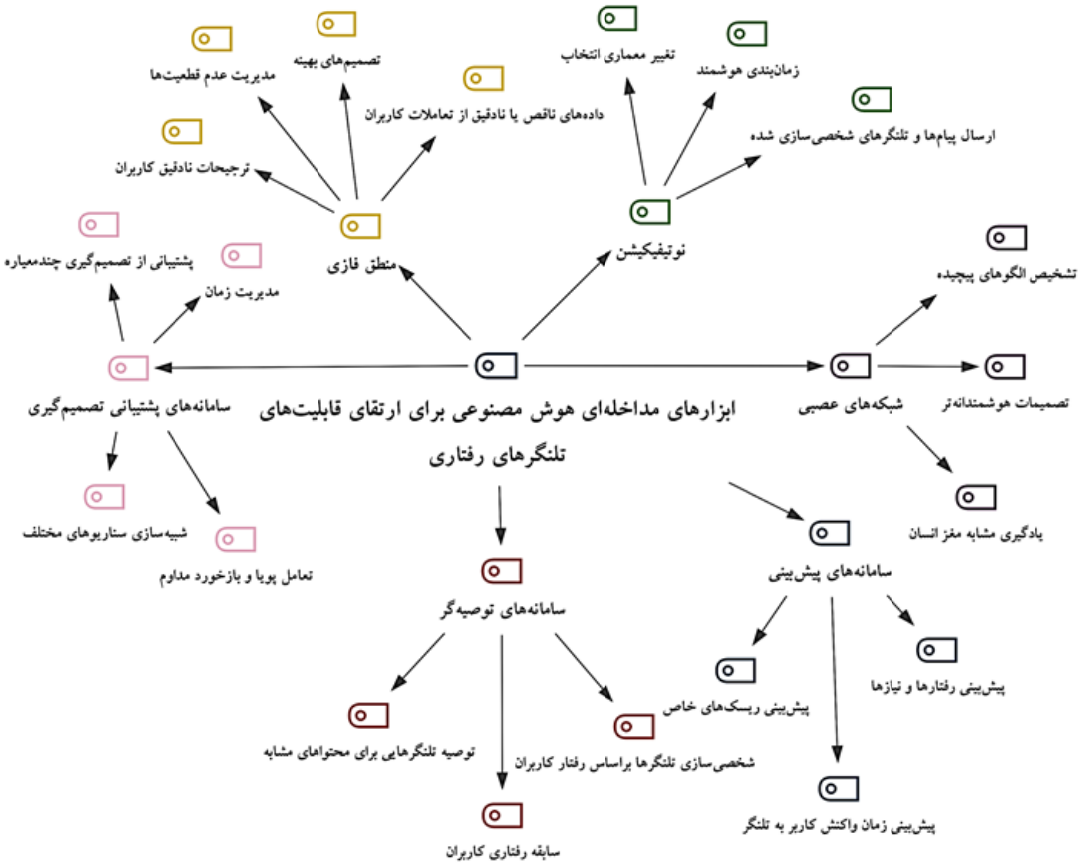
مطالعات میدانی نشان داد که از نظر خبرگان و با توجه به توسعهٔ فزاینده هوش مصنوعی، ۹ ابزار مداخله‌ای جدید می‌تواند در تولید تلنگرهای هوشمند شخصی‌سازی شده نقش ایفا کند. قابلیت ابزارهای جدید، بیشتر در حوزهٔ شخصی‌سازی تلنگرها با استفاده از داده‌های بزرگ است.

1. Big Data
2. Machine Learning
3. Algorithmic Approach
4. Software Agent
5. Internet of Things (IoT)
6. Cognitive Gadgets

جدول ۳. ابزارهای مداخله‌ای هوش مصنوعی برای ساخت تلنگرهای هوشمند حاصل مصاحبه با خبرگان و خروجی نرم‌افزار مکس‌کیودا

تم‌ها(۲۶ تم)	کد اختصاری	مضامین فرعی (۱۸ مضمون)	مضامین اصلی (۹ مضمون)
پیش‌بینی رفتارها و نیازها، پیش‌بینی زمان واکنش کاربر به تلنگرها، پیش‌بینی ریسک‌های خاص	PA	پیش‌بینی زمان واکنش ریسک	تحلیل پیش‌بینی‌کننده[1]
یادگیری از تجربه، تعادل بین تعداد و نوع تلنگرها و بازخوردهای دریافتی	RL	یادگیری بهینه‌سازی رفتار بهینه سازی تصمیم	یادگیری تقویتی[2]
یادگیری، تشخیص الگوهای پیچیده‌ای، تصمیمات هوشمندانه‌تر	NN	شبیه‌سازی عملکرد مغز انسان الگوشناسی مغز انسان	شبکه‌های عصبی[3]
شخصی‌سازی تلنگرها براساس رفتار کاربران، توصیه تلنگرهایی بر اساس سابقه رفتاری کاربران	RS	تحلیل سابقه کاربر ارائه توصیه‌ها مبتنی بر نیاز کاربر	سیستم‌های توصیه‌گر[4]
ارسال پیام‌ها و تلنگرهای شخصی‌سازی شده، زمان‌بندی هوشمند	NT	ارسال اعلان پیام کوتاه ارسال تصویر در زمان مناسب تغییر معماری انتخاب	نوتیفیکیشن[5]
ترجیحات نادقیق کاربران، داده‌های ناقص یا نادقیق از تعاملات کاربران	FL	مدیریت عدم قطعیت‌ها شناخت تصمیم‌های بهینه	منطق فازی[6]
پیشنهادات مبتنی بر متن، خلاصه‌سازی متن، ارائه اطلاعات کلیدی	NLP	تحلیل متون و گفتار تحلیل احساسات	پردازش زبان طبیعی[7]
پشتیبانی از تصمیم‌گیری چندمعیاره، شبیه‌سازی سناریوهای مختلف	DSS	مدیریت زمان تعامل پویا بازخورد مداوم	سیستم‌های پشتیبانی تصمیم‌گیری[8]
تعیین میزان پیشرفت، نقاط قوت و ضعف، الگوهای یادگیری	ALP	بازخورد فوری و هدفمند تلنگرهایی برای یادآوری	پلتفرم‌های یادگیری سفارشی[9]

1. Predictive Analytics
2. Reinforcement Learning
3. Neural Networks
4. Recommendation Systems
5. Notifications
6. Fuzzy Logic
7. Neuro-linguistic programming
8. Decision Support Systems
9. Adaptive Learning Platforms



شکل ۲. شبکهٔ مضامین برای مضمون ابزارهای مداخله‌ای هوش مصنوعی برای ارتقای قابلیت‌های تلنگرهای رفتاری حاصل از مصاحبه با خبرگان (خروجی نرم‌افزار مکس‌کیودا)

نتیجه‌گیری و پیشنهادها

این مقاله امکان به‌کارگیری قابلیت‌های هوش مصنوعی برای تولید تلنگرهای رفتاری هوشمند را بررسی کرده است. به نظر می‌رسد که با توسعهٔ روند نظری و عملیاتی کاربرد تلنگرها در هدایت رفتار بشری، اهمیت طراحی و به‌کارگیری تلنگرهای هوشمند به‌شدت افزایش پیدا کرده است. مطالعات نظری این پژوهش، بر اهمیت به‌کارگیری فرصت‌های هوش مصنوعی در حوزهٔ علوم رفتاری تأکید دارد و نشان می‌دهد که هوش مصنوعی می‌تواند به‌طور فزاینده‌ای بر مهم‌ترین چالش تصمیم‌گیری بشر، یعنی سوگیری‌های شناختی غلبه کند. به نظر می‌رسد که هوش مصنوعی با تأکید بر چهار لایهٔ زیرساخت، ادراک، شناخت و تصمیم‌گیری، می‌تواند درک عمیق‌تر و منحصربه‌فردتری از تحلیل رفتارهای انسانی با هدف باز تولید تلنگرهای هوشمند شخصی‌سازی شده از خود نشان دهد.

تلنگرهای عمومی تاکنون به‌علت کم‌هزینه و آسان‌بودن مورد استقبال قرار گرفته‌اند و به‌کارگیری آن نتایج چشمگیری را به همراه داشته است. این در حالی است که به مقولهٔ تلنگرهای هوشمند و شخصی‌سازی‌شده، به‌عنوان موج دوم تلنگرها به‌طور جدی توجه شده است. چنین تلنگرهایی مبتنی بر شناخت دقیق محیط پیرامونی و شناخت دقیق

از رفتارهای پسین و فعلی و آینده افرادی خواهد بود که در این محیطها به کنش می‌پردازند. برای مثال، در حالی که تلنگرهای سنتی همانند تلنگر پیش فرض، مقایسه، ساختارساز یا یادآور برای کاهش مصرف انرژی ممکن است شامل پیام‌های عمومی برای بخش عمده‌ای از جمعیت هدف باشد، تلنگر هوشمند می‌تواند براساس الگوی مصرف هر خانوار، پیشنهادهایی اختصاصی برای همان خانواده یا حتی تک‌تک اعضای آن ارائه کند. در محیط سازمانی نیز می‌توان برای مقابله با پدیدهٔ غیبت کارکنان از تلنگرهای اختصاصی هوشمند و شخصی‌سازی شده استفاده کرد. بر این اساس، می‌توان ادعا کرد که تلنگرهای هوشمند در غالب مواقع برداشتی هوشمند و شخصی‌سازی‌شده از تلنگرهای سنتی است. هوش مصنوعی در این فرایند، امکان اصلاح و تجمیع سریع محتوا را برای دستیابی به تلنگرهای بهتر فراهم می‌کند. در واقع، در این روش به‌دنبال آن هستیم تا فرصتی فراهم کنیم که عامل نرم‌افزاری هوشمند ویژگی‌های شخصیتی کاربران را از سیگنال‌های رفتاری آنان یاد بگیرد و تلنگرهای هوشمند و شخصی‌سازی شده‌ای را برای لحظه‌های طلایی بسازد. همان طور که میلز (۲۰۲۳) می‌گوید، با تلفیق هوش مصنوعی و معماری، انتخاب یک هم‌آفرینی ارزشی برای تولید و دوام رفاه ایجاد خواهد شد. این تلنگرهای هوشمند ما را قادر می‌سازد با شناخت بهتر سوگیری‌ها، ادغام ناهمگونی‌های شناختی و مدیریت پیچیدگی رفتار، بیش‌تلنگرها و تلنگرهای هوشمندی‌ای را بسازیم که مانند داروهای هدف‌محور دقیقاً به هدف برخورد کند. تأکید هوش مصنوعی در این میان بر مفهوم شخصی‌سازی است و فرایندی از مفهوم شخصی‌سازی انتخاب تا شخصی‌سازی پیچیده را طی می‌کند.

باید توجه شود که فرایند طراحی و پیاده‌سازی تلنگرهای رفتاری مبتنی بر هوش مصنوعی و سازگار کردن آن با نمایهٔ رفتاری کاربران، قبل از مرحلهٔ اجرا با تحلیل داده‌های پیشین کاربر از قبیل رفتار گذشته، ترجیحات و شبیه‌سازی واکنش‌های احتمالی به تلنگرهای مختلف بر اساس مدل‌های یادگیری تقویتی و ارزیابی خطرها و پیامدهای ناخواسته مانند واکنش منفی، بار شناختی بیش از حد یا اثرهای متضاد صورت می‌گیرد.

در مرحلهٔ بعد از اجرا، تمرکز بر تحلیل الگوهای رفتاری فردی در گذر زمان و سنجش میزان یادگیری سامانه از تعاملات قبلی با کاربر است. در ارزیابی تلنگرهای هوشمند، برخلاف رویکردهای سنتی، معیارهای ارزیابی جمعی نمی‌تواند نشانگر دقیقی از اثربخشی باشند؛ زیرا هر کاربر با محتوایی متفاوت روبه‌رو شده است. بدیهی است تحلیل تغییرات رفتاری نسبی در سطح فردی، سنجش اثربخشی نسبی تلنگرها، ارزیابی حس کاربران از تلنگرها و توان سامانه در اصلاح رفتار یادگیرنده نمایه باید در کانون توجه قرار گیرد.

نتایج مطالعات نظری این پژوهش با الهام از مدل معماری یک سامانه تلنگر هوشمند (کارلسن و اندرسون، ۲۰۱۹)، مبتنی بر محوریت یادگیرنده نمایه‌ای، طراحی شده است. بر این اساس یادگیری نمایه‌ای به‌عنوان پردازنده مرکزی هوش مصنوعی، چهار مؤلفه داده‌ها، علاقه‌مندی‌ها، رفتار و توانمندی‌های فردی را که باید برای آن تلنگر ساخته شود، بررسی می‌کند. ورودی این پردازنده مجموعه‌ای از فعالیت‌های پیشین و پیشنهاد شده است که در نهایت فعالیت انتخابی فرد را تحت تأثیر قرار خواهد داد. طراحی تلنگر هوشمند و فرایند شخصی‌سازی در این مرحله صورت می‌گیرد.

مطالعات نظری پژوهش، شش ابزار مداخله‌ای کلان‌داده‌ها، یادگیری ماشینی، رویکرد الگوریتمی به مسئله، عامل

نرم‌افزاری هوشمند، اینترنت اشیا و فناوری‌های شناختی را شناسایی و در قالب جدول ۲ ارائه کرد. یادگیری نمایه‌ای با توجه به این شش قابلیت، فرایند تولید تلنگرهای هوشمند را مدیریت می‌کند.

برای شناخت ابزارهای مداخله‌ای بیشتر، مطالعات میدانی پژوهش انجام پذیرفت. مطالعات میدانی ابزارهای مداخله‌ای جدیدی را شناسایی کرد. این ابزارها می‌توانند در راستای بازتولید و ارتقای تلنگرهای هوشمند مؤثر باشند. این ابزارهای مداخله‌ای جدید، در قالب تحلیل‌های پیش‌بینی‌کننده، یادگیری تقویتی، شبکه‌های عصبی، سامانه‌های توصیه‌گر تصمیم، فرایندهای اعلام‌کننده، قابلیت‌های تصمیم مبتنی بر منطق فازی و پردازش زبان طبیعی، سامانه‌های پشتیبان تصمیم‌گیری و در نهایت پلتفرم‌های یادگیری سفارشی، می‌توانند ابزارهای شش‌گانه شناسایی شده در مطالعات نظری را تقویت کنند.

این مقاله تأکید می‌کند که با توجه به رفتار پیچیدهٔ انسانی، استفاده از هوش مصنوعی یک ضرورت است و این امر می‌تواند اثربخشی مطالعات رفتاری را به نحو فزاینده‌ای افزایش دهد. این یافته با نتایج مطالعات هالزورث (۲۰۲۳) و اشمیت و استنگر (۲۰۲۱) هم‌سو است. در عین حال، به نظر می‌رسد که نتایج این پژوهش در واقع تکمیل‌کنندهٔ مطالعات میلر (۲۰۲۴) در خصوص ویژگی‌های بیش‌تلنگرها و تلنگرهای هوشمند و همچنین، مطالعات کارلسن و اندرسن (۲۰۱۹) در خصوص معماری یک سامانه تلنگر هوشمند است. این پژوهش هم‌راستا با نتایج دو مقالهٔ مذکور، بر ضرورت حرکت به‌سمت شخصی‌سازی تلنگرها در راستای ساخت تلنگرهای هوشمند مطابقت دارد.

به سیاست‌گذاران رفتاری پیشنهاد می‌شود که هم‌زمان با توسعهٔ استفاده از هوش مصنوعی در خط‌مشی‌گذاری رفتاری، به الزامات حقوقی واخلاقی و حریم خصوصی کاربران، تضمین رضایت آگاهانهٔ کاربران در جمع‌آوری داده‌ها و مقابله با سوگیری‌های الگوریتمی و متعاقب آن قابلیت ممیزی الگوریتمی و تلنگرهای دیجیتال توجه خاص داشته باشند.

بدیهی است که در خصوص طراحی سامانه‌های تلنگری هوشمند، تنظیم‌پذیری سطح مداخله، استفاده از الگوریتم‌های هوشمند و طراحی سازوکارهایی برای مقابله با سرایت سوگیری انسانی در سامانه‌های هوش مصنوعی از اهمیت ویژه ای برخوردار بوده و باید مورد توجه توسعه‌دهندگان سامانه‌های هوشمند قرار گیرد. حضور توأم پژوهشگران علم داده و علوم رفتاری، می‌تواند درک صحیح‌تری از تعامل بین مفهوم معماری انتخاب، تصمیم‌سازی ماشینی و پذیرش اجتماعی و روان‌شناختی تلنگرهای هوشمند را فراهم کند.

باید تأکید کرد تحقق ظرفیت‌های هوش مصنوعی در مداخلات رفتاری تلنگری با چالش‌های جدی فنی و زیرساختی از قبیل کیفیت و در دسترس‌بودن داده‌های دقیق، متنوع و به‌روز درباره رفتار، ترجیحات و زمینه‌های تصمیم‌گیری کاربران، وضعیت زیرساخت‌های دیجیتال قدرتمند، تعداد پایگاه‌های داده بزرگ، سیستم‌های توصیه‌گر بلادرنگ و پردازش ابری روبه‌روست. ریسک‌های الگوریتمی و خطاهای ناشی از داده‌های مغرضانه، تنظیمات نادرست یا محدودیت‌های مدل، می‌تواند تلنگرهای نامناسب یا آسیب‌زا تولید کنند و کاربر را به‌سمتی سوق دهد که با اهداف اخلاقی یا اجتماعی در تضاد باشد. فقدان شفافیت در فرایند تصمیم‌گیری الگوریتمی می‌تواند مانع اعتماد کاربران و ناظران به این سامانه‌ها شود.

بی‌توجهی به این چالش‌های فنی ممکن است به شکست سامانه‌های تلنگر هوشمند در مرحله اجرا شود. از این‌رو توصیه می‌شود که در کنار ملاحظات رفتاری و سیاستی، ظرفیت‌سازی فنی و نهادی نیز به‌عنوان یک پیش‌نیاز جدی برای توسعه این سامانه‌ها در نظر گرفته شود.

ذکر این نکته ضروری است که به‌دلیل ماهیت اکتشافی و کیفی این پژوهش و مصاحبه‌های نیمه‌ساختاریافتهٔ محدود، یافته‌ها به جوامع بزرگ‌تر با وسواس بیشتری تعمیم یابد. در عین حال، به‌دلیل تمرکز پژوهش بر تحلیل مفهومی و مدل‌سازی نظری بدون آزمون تجربی در محیط‌های واقعی یا شبیه‌سازی، میزان اثربخشی عملی ابزارهای پیشنهادی تنها در سطح نظری قابل استنتاج است. بدیهی است که مطالعات تکمیلی میدانی شامل بررسی چارچوب‌های ارزیابی دقیق و قابل تکرار برای اثربخشی تلنگرهای هوشمند، ساخت سناریوهای کاربردی و انجام مطالعاتی با روش‌های ترکیبی پیشنهاد می‌شود.

References

Abson, D. J., Fischer, J., Leventon, J., Newig, J., Schomerus, T., Vilsmaier, U., …, & Lang, D. J. (2017). Leverage points for sustainable transformation. *Ambio, 46*, 30–39. https://doi.org/10.1007/s13280-016-0800-y

Adomavicius, G. & Yang, M. (2022). Integrating behavioral, economic, and technical insights to understand and address algorithmic bias: A human-centric perspective. *ACM Transactions on Management Information Systems*, *13*(3), 1-27.

Aggarwal, C. C. (2016). *Recommender systems*. Cham: Springer International Publishing.

Ahuja, A. (2023, March 1). Generative AI is sowing the seeds of doubt in serious science. *The Financial Times*. https://www.ft.com/content/e34c24f6-1159-4b88-8d92-a4bda685a73c

Alasubramanian, G. (2020). When artificial intelligence meets behavioural economics. *NHRD Network Journal*, 14(2), 216-277. https://doi.org/10.1177/2631454120974810

Ali, S., Abuhmed, T., El-Sappagh, S., Khan Muhammad, J. M., Alonso-Moral, R., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N. & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion, 99*, 101805. https://doi.org/10.1016/j.inffus.2023.101805

Aonghusa, P. M. & Michie, S. (2020). Artificial intelligence and behavioral science through the looking glass: Challenges for real-world application. *Annals of Behavioural Medicine, 54*, 942–947. https:// doi.org/10.1093/abm/kaaa095

Balasubramanian, G. (2021). When artificial intelligence meets behavioural economics. *NHRD Network Journal*, *14*(2), 216-277.

Bar-Gill, O., Sunstein, C. R. & Talgam-Cohen, I. (2023). *Algorithmic harm in consumer markets.* SSRN at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4321763. Accessed 23 June 2023.

Bechtel, W., Abrahamsen, A. & Graham, G. (2017). The life of cognitive science. In *The Blackwell Companion to Cognitive Science*. https://doi.org/10.1002/9781405164535.part1.

Beer, S. (1993). *Designing freedom.* Anansi: Canada.

Beshears, J. & Kosowsky, H. (2020). Nudging: Progress to date and future directions. *Organizational Behavior and Human Decision Processes, 161*(Suppl.), 3–19. https://doi.org/10.1016/j.obhdp.2020.09.001

Bommasani, R., Creel, K. A., Kumar, A., Jurafsky, D. & Liang, P. (2022). Picking on the same person: Does algorithmic monoculture lead to outcome homogenization? *Advances in Neural Information Processing Systems, 35*, 3663-3678. https://arxiv.org/abs/2211.13972

Brooks, R., Nguyen, D., Bhatti, A., Allender, S., Johnstone, M., Lim, C. P. & Backholer, K. (2022). Use of artificial intelligence to enable dark nudges by transnational food and beverage companies: Analysis of company documents. *Public Health Nutrition, 25*(5), 1–23. https://doi.org/10.1017/S1368980022000490

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature human behaviour*, *5*(8), 980-989.

Burr, C., Cristianini, N. & Ladyman, J. (2018). An analysis of the interaction between intelligent software agents and human users. *Minds and Machines, 28*(4), 735–774. https://doi.org/10.1007/s11023-018-9479-0

Butenko, A. & Larouche, P. (2017). Regulation for innovativeness or regulation of innovation? *TILEC Discussion Paper No. 2015-007*. https://ssrn.com/abstract=2584863

Buyalskaya, A., Ho, H., Milkman, K. L., Li, X., Duckworth, A. L., & Camerer, C. (2023). What can machine learning teach us about habit formation? Evidence from exercise and hygiene. *Proceedings of the National Academy of Sciences*, *120*(17), e2216115120.

Chater, N. & Loewenstein, G. (2022). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioural public policy astray. *Behavioral and Brain Sciences*, 46, e147. https://doi.org/10.1017/S0140525X22002023

De Marcellis-Warn, N., Marty, F., Thelisson, E. & Warin, T. (2022). Artificial intelligence and consumer manipulations: From consumer's counter algorithms to firm's self-regulation tools. *AI and Ethics, 2*, 239–268. https://doi.org/10.1007/s43681-022-00149-5

DellaVigna, S. & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econo- metrica, 90*(1), 81–116. https://doi.org/10.3982/ECTA18709

Duckworth, A. L. & Milkman, K. L. (2022). A guide to megastudies. *PNAS Nexus, 1*(5), 1–5. https://doi. org/10.1093/pnasnexus/pgac214

Forrester, J. W. (1971). Counterintuitive behavior of social systems. *Technological Forecasting and Social Change, 3*, 109–140. https://doi.org/10.1016/S0040-1625(71)80001-X

Hacker, P. (2021). Manipulation by algorithmics: Exploring the triangle of unfair commercial practice, data protection, and privacy law. *European Law Journal*, 1–34. Advanced online publication. https://doi.org/10.1111/eulj.12389

Hagendorff, T. (2022). A virtue-based framework to support putting AI ethics into practice. *Philosophy & Technology, 35*(3), 55.Hagman, W., Andersson, D., Västfjäll, D. & Tinghög, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, 6(3), 439-453. https://doi.org/10.1007/s13164-015-0263-2

Hallsworth, M. (2023). A manifesto for applying behavioural science. *Nature Human Behaviour, 7*, 310–323. https://doi.org/10.1038/s41562-023-01555-3

Halpern, D. (2015). *'Inside the Nudge Unit'* W. H. Allen.

Hansen, P.G., Jespersen, A.M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, 31 (4), 1-28, 10.1017/S1867299X00002762

Jesse, N. (2018). Internet of things and big data: The disruption of the value chain and the rise of new software ecosystems. *AI & Society, 33*(2), 229-239. https://doi.org/10.1007/s00146-018-0807-y

Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., ... & Weber, E. U. (2012). Beyond nudges: Tools of a choice architecture. *Marketing letters, 23*(2), 487-504.

Kahneman, D. & Tversky, A. (1979). 'Prospect Theory: An Analysis of Decision Under Risk'. *Econometrica,* 47(2), 263–291.

Kahneman, D. (2011). *Thinking, fast and slow.* Penguin Books: UK.

Karlsen, R. & Andersen, A. (2019). Recommendations with a nudge. *Technologies, 7*(2), 45. https://doi.org/10.3390/technologies7020045

Kleinberg, J., Ludwig, J., Mullainathan, S. & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review, 105*(5), 491–495. https://doi.org/10.1257/aer.p20151023

Komaki, A., Kodaka, A., Nakamura, E., Ohno, Y. & Kohtake, N. (2021). System design canvas for identify- ing leverage points in complex systems: A case study of the agricultural system models, Cambodia. *Proceedings of the Design Society, 1*, 2901–2910. https://doi.org/10.1017/pds.2021.551

Leventon, J., Abson, D. J. & Lang, D. J. (2021). Leverage points for sustainability transformations: Nine guiding questions for sustainability science and practice. *Sustainability Science, 16*, 721–726. https:// doi.org/10.1007/s11625-021-00961-8

Ludwig, J. & Mullainathan, S. (2021). Fragile algorithms and fallible decision-makers: Lessons from the justice system. *Journal of Economic Perspectives, 35*(4), 71–96. https://doi.org/10.1257/jep.35.4.71

Mac Aonghusa, P. & Michie, S. (2020). Artificial intelligence and behavioral science through the looking glass: Challenges for real-world application. *Annals of Behavioral Medicine, 54*(12), 942–947. https://doi.org/10.1093/abm/kaaa095

Maier, M., Bartoš, F., Stanley, T. D. & Wagenmakers, E. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Science, 119*(31), 2200300119. https:// doi.org/10.1073/pnas.2200300119

Matz, S., Kosinski, M., Nave, G. & Stillwell, D. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences, 114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence [Online] [Date accessed: 23/03/2021]: http://www-formal.stanford.edu/jmc/history/dartmouth /dartmouth.html

Meadows, D. (1997). Leverage points: Places to intervene in a system. *Whole Earth, 91*(1), 78–84.

Meadows, D. (2001). Dancing with systems. *Whole Earth*, *106*(3), 58–63.

Mele, C., Spena, T. R., Kaartemo, V. & Marzullo, M. L. (2021). Smart nudging: How cognitive technologies enable choice architectures for value co-creation. *Journal of Business Research, 129*, 949–960. https://doi.org/10.1016/j.jbusres.2020.09.004

Michalek, G., Meran, G., Schwarze, R. & Yildiz, Ö. (2016). Nudging as a new 'soft' tool in environmental policy – An analysis based on insights from cognitive and social psychology. *Zeitschrift für Umweltpolitik & Umweltrecht, 39*, 169–207.

Michie, S., Thomas, J., Johnston, M., Aonghusa, P. M., Shawe-Taylor, J…, & West, R. (2017). The Human Behaviour-Change Project: Harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implementation Science, 12*(121). https://doi.org/10.1186/ s13012-017-0641-5

Mills, S. & Sætra, H. S. (2022). The autonomous choice architect. *AI and Society*, 1–13. https://doi.org/10.1007/s00146-022-01352-0

Mills, S. (2022, August 2). Autonomous nudges and AI choice architects – Where does responsibility lie in computer mediated decision making? *Impact of Social Sciences Blog*.

Mills, S. (2023). *AI for behavioural science*. London: Taylor & Francis Group.

Mills, S., Costa, S. & Sunstein, C. R. (2023). AI, behavioural science, and consumer welfare. *Journal of Consumer Policy*, *46*(3), 387–400. https://doi.org/10.1007/s10603-023-09526-0

Mills, S., Costa, S., & Sunstein, C. R. (2023). AI, behavioural science, and consumer welfare. *Journal of Consumer Policy*, *46*(3), 387-400.

Mont, O., Neuvonen, A. & Lahteenoja, S. (2014). Sustainable lifestyles 2050: Stakeholder visions, emerging practices and future research. *Journal of Cleaner Production*, 63(2), 24–32. https://doi.org/10.1016/j.jclepro.2013.09.007.

Newland, C., & Argyriades, D. (2019). Reclaiming Public Space: Drawing Lessons from the Past as We Confront the Future: Sustainable Development Goal 16. In *Public service excellence in the 21st century* (pp. 1-30). Singapore: Springer Singapore.

Ng, C. F. (2016). Behavioral mapping and tracking. In R. Gifford (Ed.), *Research methods for environmental psychology* (pp. xx–xx). https://doi.org/10.1002/9781119162124.ch3

Okeke, F., Sobolev, M., Dell, N., and Estrin. D. (2018). Good vibrations: can a digital nudge reduce digital overload? In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18). Association

for Computing Machinery, New York, NY, USA, Article 4, 1–12. https://doi.org/10.1145/3229434.3229463

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P. & Bernstein, M. S. (2023). *Genera- tive agents: Interactive simulacra of human behavior.* ArXiv at https://arxiv.org/pdf/2304.03442. pdf. Accessed 20 Apr 2023.

Peterson, J. C., Bourgin, D. D., Agrawal, M., Reichman, D. & Griffiths, T. L. (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science, 372*, 1209–1214. https://doi.org/10.1126/science.abe2629

Rauthmann, J. F. (2020). A (more) behavioural science of personality in the age of multi-modal sensing, big data, machine learning, and artificial intelligence. *European Journal of Personality, 34*, 593–598.

Robila, M. & Robila, S. (2020). Applications of artificial intelligence methodologies to behavioral and social sciences. *Journal of Child and Family Studies, 29*, 1–13. https://doi.org/10.1007/s10826-019-01689-x

Ruggeri, K., Benzerga, A., Verra, S., Folke, T. (2020). A behavioral approach to personalizing public health. Behav Public Policy. https://doi. org/10.1017/bpp.2020.31

Sætra, H. S. (2020). Privacy as an aggregate public good. *Technology in Society*, *63*, 101422. https://doi. org/10.1016/j.techsoc.2020.101422

Sætra, H. S. (2022). *AI for the sustainable development goals*. CRC Press.

Saheb, T. (2022). *Ethically contentious aspects of artificial intelligence surveillance: A social science perspective*. Advance online publication. https://doi.org/10.1007/s43681-022-00196-y

Saura, J. R., Ribeiro-Soriano, D. & Zegarra Saldaña, P. (2022). Exploring the challenges of remote work on Twitter users' sentiments: From digital technology development to a post-pandemic era. *Journal of Business Research, 142*, 242–254.

Schmauder, C., Karpus, J., Moll, M., Bahrami, B., & Deroy, O. (2023). Algorithmic nudging: The need for an interdisciplinary oversight. *Topoi*, *42*(3), 799-807. https://doi.org/10.1007/s11245-023-09907-4

Schmidt, R., & Stenger, K. (2021). (Dis) Embodied Rationality and 'Choice Posture': Addressing Behavioral Science's Mind-Body Problem. *Available at SSRN 4185086*.

Sharbek, N. (2022, August). How traditional financial institutions have adapted to artificial intelligence, machine learning and FinTech? In *Proceedings of the International Conference on Business Excellence*, 16(1), 837-848.

Shin, D. & Ahmad, N. (2023). Algorithmic nudge: An approach to designing human-centered generative artificial intelligence. *ZU Scholars All Works, 5980*. https://zuscholars.zu.ac.ae/works/5980

Simon, H. A. (1981). *The sciences of the artificial* (2nd ed.). MIT Press.

Smith, J. & de Villiers-Botha, T. (2021). Hey, Google, leave those kids alone: Against hypernudging children in the age of big data. *AI and Society*, 38(4), 1639-1649.

Stone, P. (2016). *Artificial intelligence and life in 2030* (Report No. 52). Stanford University.

Sunstein, C. R. (2015). *The ethics of influence.* Cambridge University Press: USA.

Szaszi, B., Higney, A., Charlton, A., Gelman, A., Ziano, I., Aczél, B., Goldstein, D. G., Yeager, D. S. & Tipton, E. (2022). No reason to expect large and consistent effects of nudge interventions. *Proceedings of the National Academy of Sciences, 119*(31), e2200732119. https://doi.org/10.1073/pnas.2200732119

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.

Vanderelst, D. & Winfield, A. (2018). An architecture for ethical robots inspired by the simulation theory of cognition. *Cognitive Systems Research, 48*, 56–66. https://doi.org/10.1016/j.cogsys.2017.04.002

Vuong, Q. H., Ho, T. M., Nguyen, H. K. & Vuong, T. T. (2018). Healthcare consumers' sensitivity to costs: A reflection on behavioural economics from an emerging market. *Palgrave Communications*, 4(1), 1–10. https://doi.org/10.1057/s41599-018-0127-3.

Weinmann, M., Schneider, C. & vom Brocke, J. (2016). Digital nudging. *Business & Information Systems Engineering*, 58(6), 433–436. https://doi.org/10.2139/ ssrn.2708250.

West, R., Michie, S., Chadwick, P., Atkins, L. & Lorencatto, F. (2020). Achieving behaviour change: A guide for national government. *Public Health England*. https://assets.publishing.service.gov.uk/gover nment/uploads/system/uploads/attachment_data/file/933328/UFG_National_Guide_v04.0 0_1_1_.pdf. Accessed 24 Apr 2023.

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., ... & Zhang, J. (2021). Artificial intelligence: A powerful paradigm for scientific research. *The Innovation, 2*(4), 100179. https://doi.org/10.1016/j.xinn.2021.100179

Yeung, K. (2017). Hypernudge: Big data as a mode of regulation by design. *Information Communication and Society, 1*, 118–136. https://doi.org/10.1080/1369118X.2016.1186713