

Enjoining Good and Forbidding Evil as an Ethical Algorithm; Establishing a Quran-Based Paradigm for the Ethical Governance of Artificial Intelligence

Abdollah Alizadeh¹, Esmail Danesh², Ahmad Javid Anvari³

¹ Level Four and Researcher, Al-Mustafa International University, Qom, Iran (**Corresponding author**).
alizadeh135612@gmail.com

² PhD., Department of Philosophy and Theology, University of Tehran, Tehran, Iran.
asmaaeldanish@gmail.com

³ PhD., Student, Prophet's Biography, Al-Mustafa International University, Qom, Iran.
javed.qom@gmail.com

Abstract

The emergence of Artificial General Intelligence and advanced models such as large language models and decision-support systems has brought about profound transformations in contemporary human life. These technologies, with their capabilities in processing massive datasets, machine learning, automated decision-making, and behavioral prediction, are being deployed across various domains including medicine, education, transportation, communications, and public administration. While they have significantly increased the speed, accuracy, and accessibility of services, they have also introduced new ethical and social challenges. Algorithmic discrimination, privacy violations, misuse of personal data, dissemination of misinformation, hate speech propagation, and erosion of public trust are among the most critical of these challenges. Given the extensive and profound impact of these systems on both individual and collective life, the development of a robust, sustainable, and proactive ethical framework for their governance has become an inescapable necessity. This article begins by highlighting the limitations of conventional Western approaches—such as general principles based on liberal democracy, human rights, and utilitarianism—and demonstrates that these models are often plagued by vagueness, reactivity, lack of solid metaphysical foundations, and an inability to provide coherent implementation mechanisms. In contrast, the present study seeks to offer an ethical and governance-oriented model for artificial intelligence by turning to a comprehensive religious source: the Holy Quran. The central research problem is how the Islamic duty of *enjoining good and forbidding evil* can be elevated from an individual ethical obligation to a systemic ethical algorithm capable of guiding artificial intelligence. The primary

Cite this article: Alizadeh, A., Danesh, E. & Javid Anvari, A. (2026). Enjoining Good and Forbidding Evil as an Ethical Algorithm; Establishing a Quran-Based Paradigm for the Ethical Governance of Artificial Intelligence. *Studies of Qur'anic Sciences*, 8(1), p. 31-54. <https://doi.org/10.22081/jqss.2026.73405.1415>

Received: 2025-10-12 ; **Revised:** 2025-12-19 ; **Accepted:** 2026-01-13 ; **Published online:** 2026-03-30

Type of article: Research Article **Publisher:** Islamic Sciences and Culture Academy

©2026/authors retain the copyright and full publishing rights



objective of the article is to formulate a Quran-based paradigm for the ethical governance of artificial intelligence. The authors aim to demonstrate that *enjoining good and forbidding evil*, when reinterpreted in a modern and systematic framework, can function as an ethical algorithm that directs intelligent technologies toward justice, human dignity, social responsibility, and continuous reform. A secondary goal is to transform technology from a merely passive tool susceptible to domination and exploitation into an uplifting and ethics-driven instrument—one that not only prevents harm but actively promotes good, reduces corruption, and strengthens public trust. Ultimately, the article seeks to show that the teachings of the Holy Quran, particularly the duty of *enjoining good and forbidding evil*, possess the potential to serve as the foundation for a practical operational system for AI ethics. The core research question is: How can a framework for the ethical governance of artificial intelligence be established based on Quranic teachings, especially the principle of *enjoining good and forbidding evil*? More precisely, the article examines whether this religious duty can transcend its traditional, primarily individual scope and be reconstructed as a dynamic, systemic, and implementable algorithm for the ethical management of intelligent technologies. It also explores the theoretical foundations of this ethical algorithm, its stages, and how it can practically contribute to reducing the social and ethical harms caused by artificial intelligence. The research method is descriptive-analytical. Drawing upon Quranic sources, Islamic interpretations, and related texts, the article first clarifies the concepts of *ma'ruf* (good) and *munkar* (evil), then analyzes this teaching in relation to emerging issues in the ethics of technology and AI governance. Within this framework, foundational principles such as justice, human dignity, rationality, social responsibility, gradualism, proportionality, and balance between rationality and spirituality are introduced as the pillars of the proposed Quran-based paradigm. Structurally, the article examines *enjoining good and forbidding evil* as an ethical algorithm with distinct levels and operational stages—including monitoring and input, detection and classification, and other practical phases formulated in the article as a cyclical and hierarchical process. The study concludes that the doctrine of *enjoining good and forbidding evil*, when reread in light of the needs of the technological age, has the capacity to become an effective and practical ethical algorithm for AI governance. By relying on the Holy Quran, this model not only addresses the shortcomings of purely legalistic or utilitarian Western approaches but also offers a comprehensive, meaning-centered, and operational framework for guiding technology toward human and divine objectives. In this paradigm, artificial intelligence is not viewed as an external threat but as a trainable and ethically directable tool that can serve justice, dignity, and social reform. In other words, *enjoining good and forbidding evil*, as an ethical algorithm, has the potential to elevate the technological ecosystem beyond mere technical functionality and transform it into a domain for realizing responsibility, human excellence, and ethics-based governance.

Keywords: Enjoining Good and Forbidding Evil, AI Ethics, AI Governance, Ethical Algorithm, Quran-Based Paradigm.

الأمر بالمعروف والنهي عن المنكر بوصفه خوارزمية أخلاقية؛ تأسيس باراداييم (نموذج معرفي) قرآني للحوكمة الأخلاقية للذكاء الاصطناعي

عبدالله عزيزاده^١، إسماعيل دانش^٢، أحمد جاويد أنوري^٣

^١ خريج المستوى الرابع في الحوزة العلمية بقم، قم، إيران (الباحث المسؤول)، alizadeh135612@gmail.com
^٢ دكتوراه، الفلسفة وعلم الكلام، جامعة طهران، طهران، إيران، asmaealdanish@gmail.com
^٣ طالب دكتوراه، سيرة النبي (ص)، جامعة المصطفى (ص) العالمية، قم، إيران، javed.qom@gmail.com

الملخص

واجه تطوير تقنيات الذكاء الاصطناعي تحديات أخلاقية متعددة، من أبرزها التمييز الخوارزمي، وانتهاك الخصوصية، والآثار الاجتماعية السلبية. وفي هذا السياق، تبرز الحاجة الملحة لإيجاد إطار متمحور حول الأخلاق لتوجيه هذه التقنيات نحو تحقيق العدالة والكرامة الإنسانية. تتناول هذه المقالة، بالاعتماد على المنهج الوصفي التحليلي، فريضة "الأمر بالمعروف والنهي عن المنكر" بوصفها خوارزمية أخلاقية، وتقدم نظرية بعنوان "الباراداييم القرآني للحوكمة الأخلاقية للذكاء الاصطناعي". يستند هذا الباراداييم إلى الأصول الجوهرية للشريعة الإسلامية، مثل: العدالة، والكرامة الإنسانية، والمسؤولية الاجتماعية، والعقلانية، والتدرج والتناسب، والتوازن بين العقلانية والمعنوية. ولا تقتصر هذه الأسس على تقديم إطار نظري للحوكمة الأخلاقية في عصر الذكاء الاصطناعي فحسب، بل تُعدّ استراتيجيات تشغيلية لتصميم أنظمة ذكية متمحورة حول الأخلاق. وتُظهر نتائج هذا البحث أن الأمر بالمعروف والنهي عن المنكر، بوصفه أحد الواجبات الإلهية، يوفر نظاماً هرمياً وتدرجياً لإصلاح السلوكيات الفردية والاجتماعية. ومن خلال الجمع بين العقل والوحي، يقدم هذا الباراداييم نموذجاً أخلاقياً وعملياً لتصميم الذكاء الاصطناعي وتطويره واستخدامه؛ بغية تحقيق العدالة والكرامة الإنسانية والمسؤولية الاجتماعية. ومن خلال تحليل الأسس المعرفية والبنية والأبعاد المتعلقة بهذا الباراداييم، قدم البحث استراتيجية أخلاقية وحضارية للحوكمة التقنية، قادرة على التكيف مع الاحتياجات المتغيرة للمجتمع العالمي.

الكلمات المفتاحية: الأمر بالمعروف والنهي عن المنكر، أخلاقيات الذكاء الاصطناعي، حوكمة الذكاء الاصطناعي، الخوارزمية الأخلاقية، الباراداييم القرآني.

استناداً إلى هذه المقالة: عزيزاده، عبدالله؛ دانش، إسماعيل؛ جاويد أنوري، أحمد (٢٠٢٦). الأمر بالمعروف والنهي عن المنكر بوصفه خوارزمية أخلاقية؛ تأسيس باراداييم (نموذج معرفي) قرآني للحوكمة الأخلاقية للذكاء الاصطناعي. *دراسات علوم القرآن*، ١٨(١)، ص ٣١-٥٤.

<https://doi.org/10.22081/jqss.2026.73405.1415>

تاريخ الإستلام: ٢٠٢٥/١٠/١٢؛ تاريخ التعديل: ٢٠٢٥/١٢/١٩؛ تاريخ القبول: ٢٠٢٥/١١/١٣؛ تاريخ الإصدار: ٢٠٢٦/٠٣/٣٠

الناشر: المعهد العالي للعلوم والثقافة الإسلامية

نوع المقالة: بحثية



امربه معروف و نهی از منکر به مثابه الگوریتم اخلاقی؛ تأسیس پارادایم قرآن بنیان برای حکمرانی اخلاقی هوش مصنوعی

عبدالله علیزاده^۱، اسماعیل دانش^۲، احمد جاوید انوری^۳

^۱ سطح چهار حوزه علمیه قم، قم، ایران (نویسنده مسئول)، alizadeh135612@gmail.com
^۲ دکتری، گروه فلسفه و کلام، دانشگاه تهران، تهران، ایران. asmaaelandanih@gmail.com
^۳ دانشجوی دکتری، سیره پیامبر(ص)، جامعه المصطفی العالمیه، قم، ایران. javed.qom@gmail.com

چکیده

توسعه فناوری‌های هوش مصنوعی با چالش‌های اخلاقی متعددی مواجه شده است که از جمله مهم‌ترین آن‌ها می‌توان به تبعیض الگوریتمی، نقض حریم خصوصی و تأثیرات منفی اجتماعی اشاره کرد. در این زمینه، نیاز به ایجاد یک چارچوب اخلاق محور برای هدایت این فناوری‌ها به سمت تحقق عدالت و کرامت انسانی، بیش از پیش احساس می‌شود. این مقاله با روش توصیفی-تحلیلی، فریضه امر به معروف و نهی از منکر را به عنوان الگوریتمی اخلاقی معرفی کرده و نظریه‌ای تحت عنوان «پارادایم قرآن بنیان حکمرانی اخلاقی هوش مصنوعی» ارائه می‌دهد. پارادایم قرآن بنیان بر اصول بنیادین شریعت اسلامی نظیر عدالت، کرامت انسانی، مسئولیت‌پذیری اجتماعی، عقلانیت، تدریج و تناسب، و توازن میان عقلانیت و معنویت استوار است. این مبانی نه تنها چارچوب نظری برای حکمرانی اخلاقی در دوره هوش مصنوعی ارائه می‌دهند، بلکه راهبردهای عملیاتی برای طراحی سیستم‌های هوشمند اخلاق محور نیز محسوب می‌شوند. یافته‌های این پژوهش نشان می‌دهند که امر به معروف و نهی از منکر، به عنوان یکی از واجبات الهی، یک نظام سلسله‌مراتبی و تدریجی برای اصلاح رفتارهای فردی و اجتماعی فراهم می‌آورد. این پارادایم با ترکیب عقلانیت و وحی، الگویی اخلاقی و عملیاتی برای طراحی، توسعه و بهره‌برداری از هوش مصنوعی به منظور تحقق عدالت، کرامت انسانی و مسئولیت‌پذیری اجتماعی ارائه می‌دهد. این پژوهش با تجزیه و تحلیل مبانی معرفتی، ساختار و ابعاد مربوط به این پارادایم، یک استراتژی اخلاقی و تمدن ساز برای حکمرانی فناورانه معرفی کرده است که قادر است با نیازهای متغیر جامعه جهانی انطباق یابد.

کلیدواژه‌ها: امر به معروف و نهی از منکر، اخلاق هوش مصنوعی، حکمرانی هوش مصنوعی، الگوریتم اخلاقی، پارادایم قرآن بنیان.

استاد به این مقاله: علیزاده، دانش، اسماعیل؛ جاوید انوری، احمد (۱۴۰۵). امر به معروف و نهی از منکر به مثابه الگوریتم اخلاقی؛ تأسیس پارادایم قرآن بنیان برای حکمرانی اخلاقی هوش مصنوعی. *مطالعات علوم قرآن*، ۸(۱)، ص ۳۱-۵۴. <https://doi.org/10.22081/jqss.2026.73405.1415>

تاریخ دریافت: ۱۴۰۴/۰۷/۲۰؛ تاریخ اصلاح: ۱۴۰۴/۰۹/۲۸؛ تاریخ پذیرش: ۱۴۰۴/۱۰/۲۳؛ تاریخ انتشار آنلاین: ۱۴۰۵/۰۱/۱۰

نوع مقاله: پژوهشی
ناشر: پژوهشگاه علوم و فرهنگ اسلامی

© ۱۴۰۵ نویسندگان دارنده حق مؤلف مقاله خود بدون محدودیت هستند.



۱. مقدمه

ظهور هوش مصنوعی عمومی به عنوان یکی از پیشرفته‌ترین فناوری‌های امروزی، تحولات عمیقی را در تمام جنبه‌های زندگی بشر به وجود آورده است. این فناوری با توانایی‌های منحصر به فرد خود، از جمله پردازش کلان‌داده‌ها، یادگیری ماشینی، خودکارسازی تصمیم‌گیری و پیش‌بینی رفتارها، به سرعت به یکی از عناصر کلیدی در زندگی مدرن تبدیل شده است. هوش مصنوعی در زمینه‌های مختلف مانند پزشکی، حمل و نقل، آموزش، ارتباطات و حتی مدیریت دولتی مورد استفاده قرار گرفته و به تسهیل زندگی و سرعت بخشی به دسترسی به خدمات کمک کرده است. سیستم‌های هوش مصنوعی عمومی، نظیر GPT-4 و مدل‌های بنیادی، مرزهای جدیدی را در زمینه اخلاق فناوری گشوده‌اند. با این حال، این فناوری در کنار فرصت‌های متعدد، چالش‌های اخلاقی و اجتماعی بسیاری نیز به وجود آورده است. از جمله این چالش‌ها می‌توان به تبعیض در تصمیم‌گیری‌های الگوریتمی، نقض حریم خصوصی، سوءاستفاده از داده‌های شخصی، انتشار اطلاعات نادرست و نفرت‌پراکنی اشاره کرد. این مسائل نه تنها تأثیر منفی بر ارزش‌های انسانی و حقوق بنیادین افراد دارند، بلکه ممکن است به گسترش فساد و انحرافات در جوامع منجر شوند؛ همچنین بی‌اعتمادی عمومی نسبت به فناوری‌های نوین را افزایش داده و موجب کاهش بهره‌وری و پیشرفت اجتماعی می‌گردند. پیچیدگی ذاتی این سیستم‌ها، عدم شفافیت در تصمیم‌گیری و اثرگذاری گسترده بر جوامع انسانی، ضرورت تدوین چارچوبی محکم و جامع برای اخلاق و حکمرانی هوش مصنوعی را به یکی از اولویت‌های جهانی تبدیل کرده است. چندین رویکرد برای پاسخ به این نیازها مطرح شده است؛ از جمله اصول اخلاقی لیبرال‌دموکراسی غرب (مانند اصول OECD و اتحادیه اروپا) و رویکردهای مبتنی بر حقوق بشر و فایده‌گرایی. هرچند این تلاش‌ها ارزشمندند، اما با چالش‌های اساسی مواجه هستند؛ از جمله کلی‌گویی، نبود مبانی متافیزیکی قوی و عدم وجود مکانیزم‌های عملیاتی مؤثر. این رویکردها غالباً واکنش‌گرا هستند و پس از بروز بحران‌ها شکل می‌گیرند؛ همچنین در دام نسبیّت فرهنگی گرفتار می‌شوند و فاقد معیارهای جهان‌شمول و ثابت اخلاقی هستند. این مشکلات منجر به آن می‌شود که مسائل چندبعدی اخلاقی به مجموعه‌ای از قواعد ایستا و پیش‌تعریف‌شده تقلیل یابند که انعطاف‌پذیری لازم برای انطباق با موقعیت‌های پیچیده و پویا را ندارند. در این شرایط، اهمیت مراجعه به منابع دینی و به‌ویژه «دین خاتم» با کتاب آسمانی جاودانش، قرآن کریم، نمایان می‌شود. تدوین یک الگوی جامع، نظام‌مند و پیش‌دستانه که از عمق نظری برخوردار بوده و قابلیت اجرایی داشته باشد، برای هدایت هوش مصنوعی به سمت اهداف عالی انسانی و الهی ضروری به نظر می‌رسد. اسلام به عنوان دینی جامع و پایدار، برای تمامی جنبه‌های زندگی انسان، از جمله مسائل نوظهوری مانند هوش مصنوعی، برنامه‌ها و راهکارهایی دارد. قرآن، به عنوان منبع اصلی تعالیم اسلامی، اصولی جامع برای هدایت اخلاقی انسان‌ها ارائه داده است. یکی از مهم‌ترین این اصول، آموزه

«امر به معروف و نهی از منکر» است که بر مسئولیت فردی و اجتماعی انسان‌ها در ترویج خیر و جلوگیری از فساد تأکید دارد. مسئله اصلی این پژوهش، شناسایی و تحلیل چگونگی استفاده از آموزه‌های قرآنی، به‌ویژه اصل امر به معروف و نهی از منکر، برای تدوین الگوریتم اخلاقی هوش مصنوعی است. پرسش‌هایی که این پژوهش در صدد پاسخ‌گویی به آن‌هاست، عبارت‌اند از: چگونه می‌توان اصل امر به معروف و نهی از منکر را به‌عنوان الگوریتمی اخلاقی برای هدایت هوش مصنوعی به‌کار بُرد؟ چه اصول اخلاقی قرآنی در حوزه عدالت، کرامت انسانی، مقابله با فساد و ترویج خیر می‌توانند در طراحی هوش مصنوعی لحاظ شوند؟ چگونه می‌توان «پارادایم قرآن‌بنیان» را به‌عنوان چارچوبی برای حکمرانی اخلاقی هوش مصنوعی تأسیس کرد؟ این مقاله با الهام از آموزه امر به معروف و نهی از منکر، در پی آن است که با عبور از خوانش‌های صرفاً فردی و ایستا از این فریضه الهی، آن را به‌عنوان یک «الگوریتم اخلاقی» پویا و سیستمی بازتعریف کند.

۲. مفهوم‌شناسی و چارچوب نظری بحث

۲-۱. تعریف معروف و منکر

طبق تعریف راغب اصفهانی، «معروف» به هر عملی اطلاق می‌شود که به‌راحتی می‌توان حسن آن را با عقل یا شرع شناسایی کرد؛ در مقابل، «منکر» به عملی اشاره دارد که از نظر عقل یا شریعت، ناپسند شناخته می‌شود (راغب اصفهانی، ۱۴۱۲ق، ص ۵۶۱). این تعریف، ارتباط میان عقل و شریعت را در تبیین ماهیت کنش‌های اخلاقی روشن می‌سازد. در چارچوب معارف دینی، «معروف» به‌معنای اطاعت از اوامر الهی و «منکر» به‌معنای نافرمانی از دستورات پروردگار است. طبرسی نیز در تفسیر خود، «معروف» را شامل هرآنچه که خدا و رسولش به آن دستور داده‌اند و «منکر» را شامل هرآنچه که از آن نهی شده‌اند، می‌داند (طبرسی، بی‌تا، ج ۲، ص ۳۵۷). در فقه اسلامی، «معروف» به هر گفتار و عملی اطلاق می‌شود که علاوه بر حسن ذاتی، از نظر شرعی واجب یا مستحب باشد؛ در مقابل، «منکر» به عملی اشاره دارد که انجام آن حرام است (طوسی، ۱۴۰۹ق، ج ۲، ص ۵۴۹). برخی فقها قلمرو «منکر» را فراتر از افعال حرام در نظر گرفته و آن را شامل اعمال مکروه نیز می‌دانند (ابن حمزه، ۱۴۰۸ق، ص ۲۰۷). این دیدگاه نشان‌دهنده انعطاف‌پذیری مفهومی «منکر» در نظام فقه اسلامی است. در پارادایم «قرآن‌بنیان»، «معروف» به هر ویژگی، حالت یا سیستمی اطلاق می‌شود که با شریعت اسلام و ارزش‌های الهی همخوانی داشته باشد. به‌اعتقاد نگارنده، این تعریف، «معروف» را از یک مفهوم صرفاً عرفی به سطحی «نهاده‌شده» و مبتنی بر وحی ارتقا می‌دهد. در عرصه هوش مصنوعی، «معروف» می‌تواند شامل خصیصه‌هایی چون عدالت (عدم تبعیض)، شفافیت، حفاظت از حریم خصوصی، خدمت به بشریت، صداقت و راست‌گویی باشد. به‌عنوان نمونه، طراحی الگوریتمی که بر اساس داده‌های

غیرتبعیض آمیز عمل کند، مصداقی از «معروف» به حساب می آید. در مقابل، «منکر» به هر چیزی اشاره دارد که شریعت و عقل آن را ناپسند می دانند. در حوزه هوش مصنوعی، «منکر» می تواند شامل رفتارها و پیامدهای مختلفی باشد؛ از تبعیض نژادی یا جنسیتی در سیستم های استخدام هوشمند، تا تولید محتوای مخرب، گفتار نفرت آمیز، نقض حریم خصوصی و طراحی سیستم های نظارتی استبدادی.

۲-۲. حکمرانی اخلاقی هوش مصنوعی

از دیدگاه نگارندگان، «حکمرانی اخلاقی هوش مصنوعی» به مجموعه ای از اصول، چارچوب ها، سیاست ها و فرآیندهای نظارتی اطلاق می شود که هدف آن هدایت طراحی، توسعه و بهره برداری از فناوری های مبتنی بر هوش مصنوعی در راستای رعایت ارزش های اخلاقی، حفظ حقوق بشر، ارتقای عدالت اجتماعی و حفاظت از کرامت انسانی است. این رویکرد ریشه در بینش اسلامی دارد که کرامت انسان را امری ذاتی می داند: «وَلَقَدْ كَرَّمْنَا بَنِي آدَمَ؛ ما فرزندان آدم را گرامی داشتیم» (اسراء، ۷۰). از منظر اسلامی، حکمرانی اخلاقی با حکمت و عدالت عجین است؛ همان گونه که قرآن هدف از بعثت انبیا را برپایی عدالت معرفی می کند: «لَقَدْ أَرْسَلْنَا رُسُلَنَا بِالْبَيِّنَاتِ وَأَنْزَلْنَا مَعَهُمُ الْكِتَابَ وَالْمِيزَانَ لِيَقُومَ النَّاسُ بِالْقِسْطِ وَأَنْزَلْنَا الْحَدِيدَ فِيهِ بَأْسٌ شَدِيدٌ وَمَنَافِعُ لِلنَّاسِ وَلِيَعْلَمَ اللَّهُ مَن يَنْصُرُهُ وَرُسُلَهُ بِالْغَيْبِ إِنَّ اللَّهَ قَوِيٌّ عَزِيزٌ؛ ما رسولان خود را با دلایل روشن فرستادیم و با آن ها کتاب [آسمانی] و میزان [شناسایی حق از باطل و قوانین عادلانه] نازل کردیم تا مردم قیام به عدالت کنند؛ و آهن را نازل کردیم که در آن نیروی شدید و منافی برای مردم است، تا خداوند بداند چه کسی او و رسولانش را یاری می کند بی آنکه او را ببیند؛ خداوند قوی و شکست ناپذیر است!» (حدید: ۲۵). این نوع حکمرانی، با ایجاد نظامی چندبُعدی از اصول و سیاست ها، فناوری های هوش مصنوعی را از چالش های اخلاقی مصون داشته و به ابزاری برای تعالی انسانی و تحقق عدالت تبدیل می کند (مطهری، ۱۳۹۰، ج ۱، ص ۱۹۵). حکمرانی اخلاقی هوش مصنوعی با رویکردی پویا و انطباق پذیر، بستر لازم برای بهره برداری از فناوری در مسیر کمال را فراهم می آورد. علامه طباطبایی در تفسیر آیه «وَمَا خَلَقْتُ الْجِنَّ وَالْإِنْسَ إِلَّا لِيَعْبُدُونِ» (ذاریات: ۵۶)، هدف نهایی آفرینش انسان را حرکت به سوی کمال و شناخت حقیقت می داند و «عبادت» را به معنای شناختی تفسیر می کند که انسان را به مقام قرب الهی می رساند (طباطبایی، ۱۳۷۴، ج ۱۸، ص ۵۸۳). بنابراین، فناوری نیز باید به عنوان ابزاری در این مسیر تعالی بخش به کار گرفته شود.

۳-۲. الگوریتم اخلاقی

الگوریتم ها به عنوان مجموعه ای از قواعد منطقی و مراحل ساختاریافته، از ابزارهای کلیدی برای حل مسائل پیچیده به شمار می روند. در همین راستا، فریضه امر به معروف و نهی از منکر نیز فرآیندی ساختارمند و سلسله مراتبی به منظور شناسایی، تذکر، اصلاح و مقابله با انحرافات اجتماعی است. بر اساس دیدگاه

نویسندگان، «الگوریتم اخلاقی امر به معروف و نهی از منکر» یک نظام تصمیم‌گیرنده و کنش‌محور است که با الهام از آموزه‌های وحیانی قرآن کریم، سیره نبوی و اصول عقلانی اسلام، رفتارهای انسانی را در جهت ترویج معروف (اعمال پسندیده) و جلوگیری از منکر (اعمال ناپسند) هدایت می‌کند (آل‌عمران، ۱۰۴؛ لقمان، ۱۷). این الگوریتم بر پایه اصول بنیادین عدالت، کرامت انسانی، مسئولیت‌پذیری، تدریج و تناسب استوار بوده و با تحلیل اخلاقی و اجتماعی، تصمیمات اصلاحی متناسب با شدت منکر و ظرفیت پذیرش فرد یا جامعه را در سه سطح قلبی، زبانی و عملی اتخاذ می‌کند. یکی از ویژگی‌های برجسته این الگوریتم، اصل «تدریج و تناسب» در برخورد با منکر است که آن را از افراط و تفریط محافظت کرده و کرامت انسانی را در تمام مراحل تصمیم‌گیری حفظ می‌نماید. این الگوریتم، علاوه بر اصلاح رفتارهای فردی و اجتماعی، به‌عنوان یک نظام عملیاتی برای حکمرانی اخلاقی در حوزه فناوری‌های پیشرفته مانند هوش مصنوعی عمل می‌کند. با تأمین شفافیت، مسئولیت‌پذیری و انطباق‌پذیری، این الگوریتم بستر مدیریت اخلاقی در تعاملات انسانی و فناورانه را فراهم می‌سازد. در زمینه هوش مصنوعی، الگوریتم اخلاقی به‌عنوان مجموعه‌ای از قواعد تصمیم‌گیری عمل می‌کند که بر اساس داده‌ها، ارزش‌ها و اهداف تعیین شده طراحی شده است. هنگامی که این منطق با آموزه‌های قرآنی هم‌راستا می‌شود، امر به معروف و نهی از منکر نقش «تصمیم‌سازی اخلاقی» را ایفا می‌کند؛ یعنی هدایت رفتارها بر اساس خیر و شر حقیقی. این الگوریتم نه تنها به‌عنوان یک سیستم نظارتی و اصلاحی برای جامعه مؤمنان عمل می‌کند، بلکه الگوی مؤثری برای هدایت رفتارهای سیستم‌های هوشمند نیز فراهم می‌آورد (آل‌عمران، ۱۰۴).

۲-۴. پارادایم قرآن‌بنیان حکمرانی

از دیدگاه نویسندگان، «پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی» چارچوبی جامع و نظام‌مند است که بر اساس آموزه‌های وحیانی قرآن کریم، اصول اخلاقی اسلامی و عقلانیت بنا شده است. این پارادایم الگویی نظری و عملی برای هدایت طراحی، توسعه و بهره‌برداری از فناوری‌های هوش مصنوعی به‌منظور تحقق عدالت، کرامت انسانی و مسئولیت‌پذیری اجتماعی ارائه می‌دهد. قرآن کریم، هدف از بعثت انبیا را تأسیس عدالت می‌داند (حدید: ۲۵) و کرامت ذاتی انسان را پایه هر حق و رفتاری می‌شمارد (اسراء، ۷۰). این پارادایم با ترکیب عقل و وحی، فناوری را از ابزاری برای قدرت و سلطه، به وسیله‌ای برای ارتقا و کمال انسان و تقویت ارزش‌های اخلاقی تبدیل می‌کند. نویسندگان بر این باورند که هوش مصنوعی، به‌عنوان یکی از ارکان تمدن بشری، باید از خطاهای اخلاقی و انحرافات ارزشی در امان باشد و همواره با اهداف نهایی انسانیت و مقاصد شریعت اسلامی همخوانی داشته باشد. این پارادایم با استفاده از اصول بنیادی «امر به معروف و نهی از منکر»، فناوری را به ابزاری برای اصلاح رفتارها، پیشگیری از انحرافات اجتماعی و ترویج خیر عمومی بدل کرده و از طریق رویکردی پویا و انعطاف‌پذیر، قابلیت

انطباق با تحولات فناورانه و نیازهای متغیر جوامع انسانی را حفظ می‌کند. قرآن نیز این اصل را مایه برکت و خیر برای امت معرفی می‌کند (آل عمران، ۱۱۰).

«الگوریتم» در تعریف گسترده، ابزاری برای حل مسئله است. در این پژوهش، از استعاره «الگوریتم اخلاقی» استفاده شده است تا بر جنبه‌های زیر تأکید شود:

۱) **روش مندی:** امر به معروف و نهی از منکر دارای مراحل و شروط مشخص و تعریف شده‌ای است. امام علی (ع) در روایتی، مراتب سه‌گانه آن (انکار قلبی، زبانی و عملی) را با ترتیب و اولویت بیان فرموده‌اند (نهج البلاغه، حکمت ۳۷۴).

۲) **کاربردی بودن:** این الگوریتم قابلیت ترجمه به پروتکل‌های عملیاتی در حوزه فناوری را دارد.

۳) **هدفمندی:** هدف نهایی آن، بهینه‌سازی فضای اجتماعی، نزدیک کردن آن به وضعیت «مطلوب» و دور کردن از وضعیت «نامطلوب» است.

۴) **خودتنظیمی:** این الگوریتم دارای مکانیزم‌های درونی برای جلوگیری از سوءاستفاده و بهینه‌سازی عملکرد است. مفهوم «تأثیرگذاری» و پرهیز از «ایجاد فتنه» که در روایات آمده است (حرّعاملی، ۱۴۱۶ق، ج ۱۶، ص ۱۳۸)، نمونه‌ای از مکانیزم خودتنظیمی برای جلوگیری از افراط و تفریط است.

۳. اصول بنیادین الگوریتم اخلاقی امر به معروف و نهی از منکر

امر به معروف و نهی از منکر، به‌عنوان یک الگوریتم اخلاقی، بر پایه اصول بنیادین اخلاق اسلامی و آموزه‌های وحیانی قرآن کریم استوار است. این اصول نه تنها چارچوبی نظری و دقیق برای هدایت رفتارهای فردی و اجتماعی فراهم می‌آورند، بلکه به‌عنوان مبانی عملی برای حکمرانی اخلاقی در حوزه فناوری‌های پیشرفته، از جمله هوش مصنوعی، نیز عمل می‌کنند.

۳-۱. اصل عدالت

«عدالت» در شریعت اسلامی، یکی از اصول بنیادین اخلاقی و معنوی است که در ساحت‌های فردی و اجتماعی، محور اصلاح رفتارها و تصمیم‌گیری‌ها قرار می‌گیرد. قرآن کریم به عدالت به‌عنوان یک فریضه الهی اشاره دارد: «إِنَّ اللَّهَ يَأْمُرُ بِالْعَدْلِ وَالْإِحْسَانِ» (نحل، ۹۰). در فریضه «امر به معروف و نهی از منکر» نیز عدالت تضمین می‌کند که اقدامات اصلاحی، متناسب با شرایط و مقتضیات فردی و اجتماعی باشند و از هرگونه افراط یا تفریط پرهیز شود؛ چنان‌که امام علی (ع) می‌فرماید: «الْعَدْلُ سَائِسٌ عَامٌّ؛ عدل، نگاهدارنده همگان است» (نهج البلاغه، حکمت ۴۳۷).

در «پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی»، اصل عدالت به طراحی و توسعه الگوریتم‌هایی اشاره دارد که از تبعیض، بی‌عدالتی و جانبداری در فرآیند تصمیم‌گیری سیستم‌های هوش مصنوعی جلوگیری می‌کنند. افزون بر این، عدالت ایجاب می‌کند که فناوری‌های هوشمند به‌گونه‌ای

طراحی شوند که دسترسی برابر به منابع، امکانات و فرصت‌ها را برای همگان فراهم آورند. بدین ترتیب، این اصل تضمین می‌کند که هوش مصنوعی همواره در خدمت برقراری عدالت اجتماعی و کاهش نابرابری‌ها باشد.

۳-۲. اصل کرامت انسانی

«کرامت انسانی» به‌عنوان یکی از اصول بنیادین اخلاق اسلامی، ارزش ذاتی انسان را فارغ از جنسیت، قومیت، دین یا موقعیت اجتماعی، به رسمیت می‌شناسد (اسراء: ۷۰). فریضه «امر به معروف و نهی از منکر» همواره بر حفظ کرامت انسانی تأکید دارد و اقدامات اصلاحی در این چارچوب باید با احترام کامل به ارزش‌ها و حقوق انسانی صورت پذیرد. این اصل تضمین می‌کند که تمامی مراحل اصلاح رفتار - اعم از تغییر درونی، گفتار و کنش عملی - با رعایت شأن و منزلت افراد همراه باشد. در «پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی»، اصل کرامت انسانی مستلزم طراحی فناوری‌هایی است که در تمامی مراحل توسعه و بهره‌برداری، شأن و حقوق بنیادین انسان را حفظ کنند. این اصل ایجاب می‌کند که سیستم‌های هوش مصنوعی از هرگونه آسیب به حریم خصوصی، تبعیض و یا تخریب شأن انسانی پرهیز کنند. فناوری‌های هوشمند باید به‌گونه‌ای طراحی و پیاده‌سازی شوند که نه تنها حقوق انسانی را رعایت کنند، بلکه در جهت تعالی و ارتقای کرامت انسانی نیز گام بردارند.

۳-۳. اصل مسئولیت‌پذیری اجتماعی

«مسئولیت‌پذیری اجتماعی» در اسلام، از اصول بنیادین اخلاقی است که بر ضرورت شفافیت، پاسخ‌گویی و پابندی به حقوق دیگران تأکید دارد (انعام، ۱۶۴). این اصل، مسئولیت اخلاقی و اجتماعی را یکی از ارکان رفتار انسانی دانسته و فرد را ملزم می‌سازد که در برابر اعمال و تصمیم‌های خود پاسخ‌گو باشد. در فریضه «امر به معروف و نهی از منکر» نیز مسئولیت‌پذیری اجتماعی جایگاهی محوری دارد؛ به‌گونه‌ای که اقدامات اصلاحی اخلاقی، همواره با در نظر گرفتن پیامدهای اجتماعی و پذیرش مسئولیت کامل نسبت به نتایج کنش‌ها همراه است. در «پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی»، مسئولیت‌پذیری اجتماعی مستلزم طراحی سیستم‌هایی است که شفافیت و پاسخ‌گویی را در تمامی مراحل توسعه و بهره‌برداری تضمین کنند. این اصل، فناوری‌های هوشمند را ملزم می‌کند که در فرآیند تصمیم‌گیری و کنش‌های خود، مسئولیت اخلاقی و اجتماعی را به‌طور کامل رعایت کرده و پیامد تصمیم‌های خود را به‌صورت شفاف و قابل‌فهم برای کاربران و ذی‌نفعان ارائه دهند.

۳-۴. اصل عقلانیت

«عقلانیت» در اسلام، یکی از ابزارهای بنیادین برای شناخت معروف و منکر و هدایت رفتارهاست. قرآن

کریم تعقل را از نشانه‌های بارز خردورزی انسانی برمی‌شمارد: «إِنَّ فِي ذَلِكَ لآيَاتٍ لِّقَوْمٍ يَعْقِلُونَ» (رعد، ۴). در شریعت اسلامی، عقلانیت نه تنها ابزاری برای درک اصول اخلاقی و اتخاذ تصمیم‌های صحیح است، بلکه یکی از ارکان اصلی حکمرانی اخلاقی نیز محسوب می‌شود. فریضه «امر به معروف و نهی از منکر» نیز بر پایه عقلانیت استوار است و تضمین می‌کند که اقدامات اصلاحی، مبتنی بر تحلیل دقیق و درک درست از شرایط و مقتضیات باشد. این اصل، «الگوریتم اخلاقی امر به معروف و نهی از منکر» را به نظامی عقلانی بدل می‌سازد که قادر است در مواجهه با چالش‌های پیچیده اجتماعی و اخلاقی، تصمیم‌های سنجیده ارائه دهد. در «پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی»، عقلانیت مستلزم به‌کارگیری تحلیل‌های منطقی و داده‌های دقیق در طراحی و توسعه الگوریتم‌های هوش مصنوعی است. این اصل تضمین می‌کند که فناوری‌های هوشمند، ضمن شناسایی و اصلاح رفتارهای غیراخلاقی بر پایه تحلیل‌های داده‌محور، از اتخاذ تصمیم‌های غیرمنطقی یا ناعادلانه پرهیز کنند.

۳-۵. اصل تدریج و تناسب

اصل «تدریج و تناسب» از آموزه‌های کلیدی در شریعت اسلامی است که بر ضرورت رعایت ترتیب و تناسب در فرآیند اصلاح رفتارها تأکید دارد. این اصل ریشه در آموزه‌های مرتبط با فریضه «امر به معروف و نهی از منکر» و سیره نبوی دارد. نزول تدریجی احکام (مانند احکام مرتبط با تحریم شراب) و آیه شریفه «ادْعُ إِلَى سَبِيلِ رَبِّكَ بِالْحُكْمِ وَالْمَوْعِظَةِ الْحَسَنَةِ وَجَادِلْهُمْ بِالَّتِي هِيَ أَحْسَنُ» (نحل، ۱۲۵)، به روشنی بر اهمیت حکمت، موعظه حسنه و رعایت تناسب با شرایط مخاطب دلالت دارد. این اصل تضمین می‌کند که فرآیند اصلاح رفتار با رویکردی سلسله‌مراتبی - که از سطح قلبی آغاز شده و به مراحل زبانی و عملی می‌رسد - انجام پذیرد و از هرگونه افراط یا تفریط در مواجهه با رفتارهای ناهنجار پرهیز شود. در «پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی»، اصل تدریج و تناسب مستلزم طراحی سیستم‌هایی است که توانایی تطبیق‌پذیری با شرایط و نیازهای متغیر کاربران و محیط را داشته باشند. این اصل تضمین می‌کند که فناوری‌های هوشمند، کنش‌های اصلاحی خود را به‌صورت مرحله‌بندی‌شده، متناسب با زمینه‌های فرهنگی و اجتماعی، و با رعایت اعتدال انجام دهند تا ضمن پیشگیری از انحرافات، از هرگونه تصمیم‌گیری افراطی یا واکنش‌های نامتناسب جلوگیری شود.

۳-۶. اصل توازن میان عقلانیت و معنویت

اسلام بر یگانگی «عقلانیت» و «معنویت» تأکید ویژه‌ای دارد: «وَكَذَلِكَ جَعَلْنَاكُمْ أُمَّةً وَسَطًا» (بقره، ۱۴۳). این آیه، اصل «وسطیت» و توازن را به‌عنوان هویت امت اسلامی ترسیم می‌کند (طبرسی، ۱۴۱۲ق، ج ۱، ص ۴۳۲). در اندیشه اسلامی، این اصل متضمن برقراری توازن میان ابعاد عقلانی و معنوی زیست انسانی است؛ چنان‌که در فریضه «امر به معروف و نهی از منکر» نیز اقدامات اصلاحی نه

صرفاً بر پایه تحلیل‌های عقلانی، بلکه با اتکا به آموزه‌های معنوی و اخلاقی صورت می‌پذیرد (مطهری، ۱۳۷۲، ج ۲، ص ۱۹۶). در «پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی»، توازن میان عقلانیت و معنویت، مستلزم طراحی فناوری‌هایی است که فراتر از تحلیل‌های منطقی، ارزش‌های متعالی و اخلاقی را نیز مدنظر قرار دهند. این اصل تضمین می‌کند که فناوری‌های هوشمند در راستای تحقق اهداف غایی خلقت - که همانا عبودیت و قرب الهی است - گام بردارند (ذاریات، ۵۶). این نظام اخلاقی با ایجاد تعادل میان عقل و وحی، الگویی اخلاق‌محور برای حکمرانی فناورانه ارائه می‌دهد. در این چارچوب، فناوری‌های هوشمند نه صرفاً به‌عنوان ابزارهای تولید و کاربست، بلکه به‌مثابه بخشی از تمدن اخلاقی اسلام، در خدمت تعالی انسان قرار می‌گیرند. این رویکرد، همسو با هدف بعثت انبیاست که پیامبر اکرم (ص) آن را چنین تبیین فرموده‌اند: «إِنَّمَا بُعِثْتُ لِأَتَمِّمَ مَكَارِمَ الْأَخْلَاقِ» (مجلسی، ۱۴۰۳، ج ۶۸، ص ۳۸۵). «پارادایم قرآن‌بنیان» با بهره‌گیری از اصول «تدریج و تناسب»، بر اصلاح رفتارها و حکمرانی اخلاقی در قلمرو فناوری، به‌ویژه هوش مصنوعی، متمرکز است و آینده‌ای اخلاق‌محور را با تلفیق عقلانیت و وحی ترسیم می‌کند. قرآن کریم بر این اصل تأکید دارد: «ادْعُ إِلَى سَبِيلِ رَبِّكَ بِالْحُكْمَةِ وَالْمَوْعِظَةِ الْحَسَنَةِ» (نحل، ۱۲۵). این آیه، سه سطح «حکمت»، «موعظه حسنه» و «مجادله احسن» را به‌عنوان مراحل مداخلات اخلاقی معرفی می‌کند که می‌تواند الگویی عملیاتی برای سطوح مختلف حکمرانی هوش مصنوعی باشد. بدین ترتیب، فناوری‌های هوشمند به ابزارهایی تمدن‌ساز و اخلاق‌محور بدل می‌شوند که در خدمت تحقق ارزش‌های انسانی و الهی خواهند بود.

۴. ساختار الگوریتم امر به معروف و نهی از منکر

فریضه «امر به معروف و نهی از منکر» از ارکان بنیادین اخلاق اسلامی است که افزون بر جایگاه والای دینی، چارچوبی کارآمد برای هدایت و اصلاح رفتارهای فردی و اجتماعی فراهم می‌آورد. این اصل به‌واسطه ساختار سلسله‌مراتبی و انعطاف‌پذیری ذاتی‌اش، از قابلیت تبدیل به یک «الگوریتم اخلاقی» برخوردار است و می‌تواند به‌عنوان زیربنایی نظری در طراحی سیستم‌های هوش مصنوعی اخلاق‌محور به‌کار گرفته شود. امروزه فناوری‌های مبتنی بر هوش مصنوعی، ضمن خلق فرصت‌های بی‌شمار، چالش‌های پیچیده‌ای نظیر نقض حریم خصوصی، وابستگی کاربران، تولید محتوای مخرب و آسیب‌های روان‌شناختی را نیز به همراه داشته‌اند. در چنین وضعیتی، فریضه امر به معروف و نهی از منکر به‌عنوان مجموعه‌ای از اصول و ابزارهای کنش‌گرایانه، می‌تواند پارادایمی جامع برای طراحی و توسعه این سیستم‌ها ارائه دهد تا فرآیند حکمرانی آن‌ها بر پایه‌های اخلاقی استواری مستقر گردد. «الگوریتم اخلاقی امر به معروف و نهی از منکر» در مقام چارچوبی برای حکمرانی اخلاقی هوش مصنوعی، مشتمل بر شش مرحله کلیدی است که در مجموع، یک چرخه کامل کنش اخلاقی را صورت‌بندی می‌کنند:

۴-۱. ورودی و پایش^۱

این مرحله ناظر به گردآوری داده‌های محیطی و نظارت مداوم بر فضای عملیاتی سیستم هوش مصنوعی است. در این فرآیند، سیستم باید به‌طور مستمر داده‌های مرتبط با رفتار کاربران و شرایط محیطی را پایش کند تا امکان شناسایی دقیق الگوهای رفتاری منطبق یا متضاد با ارزش‌های «معروف» و «منکر» فراهم آید. در سنت اسلامی، «شناخت بستر اجتماعی» و درک شرایط کنشگران، یکی از مؤلفه‌های بنیادین فریضه امر به معروف و نهی از منکر است. این اصل ریشه در سیره عملی پیامبر اکرم (ص) و ائمه اطهار (ع) دارد که همواره با آگاهی دقیق از اوضاع اجتماعی و روانی افراد، به هدایت و اصلاح امت می‌پرداختند. پیامبر اکرم (ص) در بیان شروط تأثیرگذاری این فریضه می‌فرماید: «امر به معروف و نهی از منکر نکند مگر کسی که سه خصلت داشته باشد: در امر و نهی خود مدارا کند، در این اقدامات میانه‌روی پیشه سازد و به آنچه امر و نهی می‌کند، آگاهی کامل داشته باشد» (مجلسی، ۱۴۰۳ق، ج ۹۷، ص ۸۷). این کلام نبوی، بر لزوم شناخت عمیق بستر اجتماعی، فرهنگی و فردی برای هرگونه کنش اصلاحی تأکید می‌ورزد. این رویکرد مبتنی بر آگاهی در منابع اسلامی، نشان‌دهنده اهمیت حیاتی «پایش مداوم» برای تشخیص رفتارهای اخلاقی است. این مرحله، سنگ‌بنای تشکیل چارچوب اخلاقی مدنظر است و در حوزه‌ی فناوری‌های نوین، الگویی کارآمد برای طراحی سیستم‌های «پایش و تشخیص»^۲ در هوش مصنوعی فراهم می‌آورد. چنین سیستمی می‌تواند با تحلیل هوشمند و مداوم داده‌های محیطی (اعم از متن، تصویر یا الگوهای رفتاری)، نمونه‌های «منکر» را شناسایی کرده و پیش از وقوع آسیب‌های گسترده، به‌صورت پیش‌دستانه^۳ وارد عمل شود. این الگو، تجسم عملی روایتی است که امر به معروف و نهی از منکر را «سدی در برابر تخلفات» (کلینی، ۱۴۱۷ق، ج ۵، ص ۵۹) می‌داند؛ سدی که جامعه را پیش از فروپاشی اخلاقی، محافظت و صیانت می‌کند.

۴-۲. مرحله تشخیص و طبقه‌بندی^۴

در این مرحله، سیستم هوش مصنوعی با بهره‌گیری از «پارادایم قرآن‌بنیان حکمرانی اخلاقی»، داده‌های دریافتی را تحلیل و بر اساس ارزش‌های شناختی و آموزه‌های متعالی، طبقه‌بندی می‌کند. در این چارچوب مفهومی، مفاهیم «معروف» و «منکر» به‌عنوان شاخص‌های بنیادین در محور تصمیم‌گیری‌های سیستم قرار می‌گیرند. قرآن کریم، این دو مفهوم را شالوده‌ی تمایز خیر از شر و هدایت جامعه معرفی می‌فرماید: «الْمُؤْمِنُونَ وَالْمُؤْمِنَاتُ بَعْضُهُمْ أَوْلِيَاءُ بَعْضٍ يَأْمُرُونَ بِالْمَعْرُوفِ وَيَنْهَوْنَ عَنِ الْمُنْكَرِ؛ مردان و زنان

1. Input & Monitoring
2. Detection Systems & Monitoring
3. Proactive
4. Detection & Classification

مؤمن، اولیای یکدیگرند؛ آنان به نیکی فرمان می‌دهند و از بدی باز می‌دارند» (توبه، ۷۱). الگوریتم‌های هوش مصنوعی برای اتخاذ تصمیمات دقیق و قابل اطمینان، مستلزم داشتن مبنایی نظام‌مند و منطبق با اصول اخلاقی اسلام هستند. در جهان اسلام، این مبنا بر پایه آموزه‌های وحیانی - که قرآن کریم آن را «تَبَيَّنًا لِّكُلِّ شَيْءٍ» (نحل: ۸۹) می‌خواند - تبیین می‌شود. این آموزه‌ها طی قرون متمادی در متون فقهی و روایی، با هدف تنظیم روابط اجتماعی و الگوی حکمرانی، تفصیل یافته‌اند. بهره‌گیری از این پشتوانه‌ی غنی، الگویی تاریخی و نظام‌مند را برای طراحی ساختار اخلاقی سیستم‌های هوشمند فراهم می‌آورد که می‌تواند الهام‌بخش حکمرانی فناورانه در عصر حاضر باشد.

۳-۴. مرحله ارزیابی و اولویت‌بندی^۱

این مرحله، کانون شناختی و پردازشی الگوریتم است؛ جایی که پس از تشخیص اولیه، سیستم باید با تکیه بر معیارهای دقیق فقهی و عقلانی، درباره ضرورت، کیفیت و اولویت مداخله تصمیم‌گیری کند. این تصمیم‌گیری فرآیندی «صفر و یک» نیست، بلکه مبتنی بر یک «تابع ارزیابی ریسک اخلاقی» است که از اصول عمیق شریعت اسلامی الهام گرفته است. در فقه اسلامی، وجوب اجرای این تکلیف مشروط به چهار شرط کلیدی است که هر یک در منظومه‌ی شناختی و عملی این فریضه نقشی حیاتی دارند (حلی، ۱۴۱۰ق، ج ۲، ص ۲۳):

۱) **شناخت معروف و منکر:** آمر و ناهی باید نسبت به مصادیق خیر و شر از منظر شریعت، معرفت دقیق داشته باشند.

۲) **امید تأثیر:** در صورتی که امکان تأثیرگذاری مداخله منتفی باشد، الزام آن مرتفع می‌گردد.

۳) **اصرار بر ارتکاب:** وجود نشانه‌هایی دال بر تداوم رفتار فرد در ارتکاب منکر.

۴) **فقدان مفسده:** اگر بیم زیان جانی، مالی یا آبرویی برای شخص آمر یا دیگر مؤمنان وجود داشته باشد، اجرای این تکلیف ساقط می‌شود.

این سازوکار در ساحت هوش مصنوعی، می‌تواند به مثابه یک الگوریتم پیشرفته برای اصلاح و هدایت عمل کند. سه مؤلفه‌ی اصلی این فرآیند با منطق الگوریتمی تطبیق یافته و در قالبی نظام‌مند، به شرح زیر تبیین می‌گردد:

۳-۴-۱. علم به منکر^۲

نخستین و اساسی‌ترین شرط، اطمینان از وقوع عمل ناپسند است. مداخله اخلاقی نباید بر پایه گمان یا شک استوار باشد. قرآن کریم به صراحت از تجسس و گمانه‌زنی‌های بی‌اساس نهی می‌کند: «يَا أَيُّهَا الَّذِينَ

1. Evaluation & Prioritization

2. Certainty of Detection

آمَنُوا اجْتَنِبُوا كَثِيرًا مِّنَ الظَّنِّ إِنَّ بَعْضَ الظَّنِّ إِثْمٌ» (حجرات: ۱۲). فقها نیز اجماع دارند که امر و نهی باید مبتنی بر علم و یقین باشد (محقق حلی، ۱۴۰۸ق، ج ۱، ص ۳۴۲). در حکمرانی هوش مصنوعی، این اصل به این معناست که سیستم AI باید به یک آستانه اطمینان^۱ بالا در تشخیص خود دست یابد. برای مثال، یک الگوریتم تعدیل محتوا پیش از حذف یک نظر، باید با اطمینان بالای ۹۹٪ تشخیص دهد که آن نظر مصداق قطعی نفرت‌پراکنی است تا از سرکوب ناحق آزادی بیان و اتهام‌زنی به کاربران بی‌گناه جلوگیری شود.

۴-۳-۲. احتمال تأثیر^۲

شرط دوم، امید به اثربخشی اقدام است. اگر مداخله، تأثیری در اصلاح رفتار یا جلوگیری از منکر نداشته باشد، و جواب آن ساقط می‌شود. این اصل بر عقلانیت و نتیجه‌گرایی حکیمانه شریعت تأکید دارد و از اقدامات بیهوده و فرسایشی جلوگیری می‌کند (حرّ عاملی، ۱۴۱۶ق، ج ۱۶، ص ۱۳۲). در حکمرانی هوش مصنوعی، این شرط سیستم را ملزم به ارزیابی احتمال موفقیت می‌کند. هوش مصنوعی باید با تحلیل داده‌های رفتاری کاربر و زمینه موقعیتی، پیش‌بینی کند که کدام نوع مداخله (مانند یک هشدار نرم، یک تذکر جدی یا یک محدودیت موقت) بیشترین تأثیر مثبت را خواهد داشت. اگر سیستم تشخیص دهد که هرگونه مداخله‌ای منجر به لجبازی و تشدید رفتار منفی می‌شود، ممکن است تصمیم به عدم مداخله مستقیم و انتخاب روش دیگری بگیرد.

۴-۳-۳. عدم ترتب مفسده بزرگ‌تر^۳

این شرط، پیچیده‌ترین و مهم‌ترین معیار ارزیابی است. اقدام اصلاحی نباید خود به یک منکر بزرگ‌تر یا مفسده‌ای مهم‌تر منجر شود. این اصل، ریشه در قاعده فقهی «دفع المفسده مقدم علی جلب المصلحة» (دفع مفسده بر جلب مصلحت مقدم است) و همچنین آیه شریفه «وَلَا تَسُبُّوا الَّذِينَ يَدْعُونَ مِن دُونِ اللَّهِ فَيَسُبُّوا اللَّهَ عَدْوًا بِغَيْرِ عِلْمٍ» (انعام، ۱۰۸) دارد که نشان می‌دهد حتی یک کنش حق (نهی از شرک) اگر به مفسده بزرگ‌تر (توهین به مقدسات الهی) بینجامد، باید ترک شود. در حکمرانی هوش مصنوعی، این اصل سیستم را ملزم به ارزیابی عواقب ناخواسته می‌سازد. سیستم باید پیامدهای جانبی و زنجیره‌ای اقدام خود را مدل‌سازی کند. برای مثال، مسدودسازی ناگهانی یک حساب کاربری پر مخاطب که در حال انتشار اطلاعات غلط است، ممکن است در کوتاه‌مدت آن را متوقف کند، اما ممکن است موجب همدردی با آن کاربر، رادیکالیزه شدن حامیانش و انتقال آن‌ها به پلتفرم‌های غیرقابل نظارت شود که این

1. Confidence Threshold
2. Probability of Success Calculation
3. Unintended Consequences Assessment

خود مشکلی بزرگ‌تر ایجاد می‌کند. در چنین شرایطی، الگوریتم هوشمند ممکن است به‌جای حذف، استراتژی «برچسب‌گذاری محتوا به‌عنوان اطلاعات نادرست» یا «کاهش دسترسی به انتشار» را انتخاب کند.

این سه شرط اساسی، پایه‌ای محکم برای الگوریتمی هوشمند، زمینه‌آگاه و پیشگیرانه فراهم می‌آورند که به هوش مصنوعی این امکان را می‌دهد تا با تحلیل حساسیت‌های موقعیتی، اولویت اقدامات خود را تعیین کرده و سطح واکنش را بر اساس ضرورت تنظیم کند.

۴-۴. مرحله اجرا و اقدام سلسله‌مراتبی^۱

این مرحله به آغاز مداخلات اخلاقی اختصاص دارد که به‌صورت پلکانی و بر اساس ضرورت انجام می‌شود (نوری همدانی، ۱۳۷۷، ص ۲۸۱). این اقدامات می‌توانند در سه سطح قلبی، زبانی و عملی اجرا شوند. اقدامات عملی شامل مداخلات مستقیم برای اصلاح رفتارهای ناپسند است. منابع اسلامی، سلسله‌مراتب اقدام در امر به معروف را به سه سطح کلی تقسیم می‌کنند:

۴-۴-۱. قلب (شناختی)

در این سطح، رفتارهای منکر شناسایی شده و در سیستم ثبت می‌شوند تا در آینده تحلیل یا استفاده شوند. در این مرحله، فرد از نظر درونی نسبت به رفتارهای ناپسند واکنش نشان می‌دهد و انگیزه اصلاح را در خود پرورش می‌دهد (علامه حلی، ۱۴۱۴ق، ج ۹، ص ۴۳۳). در این مرحله، آمر و ناهی باید ناخشنودی قلبی خود را نسبت به رفتار ناپسند یا ترک عمل نیک، به‌صورت غیرمستقیم و بدون درگیری آشکار کند (امام خمینی، بی‌تا، ج ۱، ص ۴۵۳). این مرتبه شامل ابراز ناراحتی از طریق رفتارهایی مانند ترش‌رویی، اخم‌کردن، برهم‌نهادن نگاه، روی‌گردانی، کناره‌گیری یا قطع ارتباط با مرتکب منکر است. این مرتبه، ابتدایی‌ترین و درعین حال ساده‌ترین روش اصلاح رفتار است که می‌تواند در بسیاری از موارد فرد خطاکار را به تغییر رفتار ترغیب کند. این مرحله می‌تواند به‌عنوان بخشی از فرآیند تصمیم‌گیری اخلاقی در طراحی الگوریتم‌های هوش مصنوعی عمل کند. این سطح شامل ادراک درونی آمر و ناهی است و احساس قلبی نسبت به گناه و واکنش باطنی به آن، به‌عنوان ورودی اولیه سیستم شناختی فرد عمل می‌کند.

۴-۴-۲. لسان (زبانی)

مرحله تذکر و هشدار، با هدف اطلاع‌رسانی و ارشاد فرد خاطی صورت می‌گیرد. در هوش مصنوعی، می‌توان آن را در قالب هشدارهای مستقیم یا پیشنهادی جایگزین پیاده‌سازی کرد. دومین مرحله از مراتب امر به معروف و نهی از منکر، استفاده از زبان است (علامه حلی، ۱۴۱۴ق، ج ۹، ص ۴۴۴). اگر

1. Execution & Hierarchical Action

انکار قلبی تأثیری در تغییر رفتار ناپسند نداشته باشد، امر و نهی زبانی ضروری می‌شود. در این مرحله، فرد باید با استفاده از سخنان نرم و موعظه، شخص را به انجام کارهای پسندیده یا ترک کارهای ناپسند دعوت کند. فقها تأکید دارند که در این مرحله نیز باید اصل تدریج رعایت شده و از روش‌های ملایم آغاز گردد. این مرتبه تنها در صورتی واجب خواهد بود که احتمال تأثیر وجود داشته باشد و نباید از حدود آن فراتر رفت. در هوش مصنوعی، این سطح می‌تواند منجر به طراحی سیستم‌هایی شود که از طریق تعامل زبانی یا توصیه‌گرهای اخلاقی، کاربران را به رفتارهای مطلوب هدایت کنند. امر و نهی زبانی که به عنوان پردازش اخلاقی در ارتباط با مخاطب صورت می‌گیرد، نقش هدایت‌گری ذهن و بازآرایی تصمیم او را ایفا می‌کند. این سطح باید با توجه به شرایط مخاطب و بستر اجتماعی طراحی شود تا تأثیرگذاری مطلوبی را به همراه داشته باشد.

۴-۳-۴. ید (عملی)

مرحله اقدام مستقیم، که در شرایط خاص برای مقابله فعال با یک رفتار ناهنجار اجرا می‌شود، شامل اقدامات عملی هوش مصنوعی است. این اقدامات می‌تواند شامل مسدودسازی محتوا، غیرفعال‌سازی عملکرد سیستم یا محدود کردن دسترسی کاربران باشد. آخرین مرحله از مراتب امر به معروف و نهی از منکر، به اقدامات عملی و در صورت نیاز، استفاده از زور مربوط می‌شود (علامه حلی، ۱۴۱۴ق، ج ۹، ص ۴۴۳). زمانی که دو مرحله قبل تأثیری نداشته باشند و شرایط ایجاب کند، امر به معروف و نهی از منکر با اقدامات عملی ضرورت می‌یابد. این اقدامات باید با رعایت اصل تدریج، از کمترین میزان شدت آغاز شده و در صورت نیاز، به مراحل بالاتر ارتقا یابد. منظور از اقدامات عملی، شامل اقداماتی است که فرد را در تنگنا قرار می‌دهد، از انجام منکر جلوگیری می‌کند و در موارد ضروری‌تر، به اقدامات شدیدی مانند ضرب و حبس می‌انجامد. با این حال، این اقدامات باید به گونه‌ای اجرا شوند که منجر به آسیب جسمی جدی، نقص عضو، جراحت یا خسارت‌های جبران‌ناپذیر نشوند. مرحله نهایی شامل اتخاذ اقداماتی مؤثر برای اصلاح رفتارها و جلوگیری از انحرافات اجتماعی است. این سطح می‌تواند در طراحی سیستم‌هایی که به طور فعال از رفتارهای ناپسند جلوگیری می‌کنند، به کار رود. این ساختار قلبی، زبانی و عملی به عنوان یک سلسله‌مراتب، الگوریتمی اخلاقی را شکل می‌دهد که می‌تواند به عنوان چارچوبی برای اصلاح رفتارهای فردی و اجتماعی در عصر فناوری‌های هوشمند مورد استفاده قرار گیرد. این فرآیند می‌تواند به عنوان یک چارچوب مؤثر برای طراحی سیستم‌های اخلاق‌محور به کار رود. الگوریتم‌هایی که بر اساس این اصول طراحی شوند، قادر خواهند بود رفتارهای اخلاقی را در سیستم‌های هوشمند ترویج داده و جامعه را به سوی اصلاح و فضیلت هدایت کنند. این مدل سلسله‌مراتبی نشان می‌دهد که مداخله اخلاقی هوش مصنوعی نه تنها متناسب با ماهیت خطا، بلکه با شرایط زمینه‌ای اجرا می‌شود.

جدول ۱ - مراحل سه‌گانه در الگوریتم اخلاقی

مرحله اخلاقی	مفهوم در الگوریتم اخلاقی	توضیح تطبیقی
ادراک خیر و شر	ورودی داده‌ها	معرفت نسبت به حسن و قبح اعمال، معادل ورود داده‌های اخلاقی به سیستم تصمیم‌گیری است.
تذکر و هدایت زبانی	پردازش داده‌ها	هشدار، موعظه یا گفتار اصلاحی، به‌عنوان فرآیند پردازش اخلاقی برای تغییر ذهنیت و تصمیم مخاطب عمل می‌کند.
کنش اصلاحی یا اجتناب از شر	خروجی نهایی	نتیجه‌ی نهایی شامل اصلاح رفتار یا اجتناب از گناه است که معادل خروجی مطلوب الگوریتم اخلاقی می‌باشد.

۴-۵. مرحله بازخورد و یادگیری^۱

نتیجه اقدامات مداخله‌ای سیستم باید به‌طور مستمر تحلیل و به داده‌های آموزشی اضافه شود. این فرآیند یادگیری به الگوریتم امکان می‌دهد که در مواجهه با موارد جدید، کارآمدتر عمل کند. این ویژگی هوش مصنوعی منطبق بر اصول «یادگیری تقویتی»^۲ است. در واقع، این مرحله از امر به معروف و نهی از منکر الگوریتمی، به سیستم ماهیت تکاملی می‌بخشد. این روند بازخورد، مدل پایه اخلاقی سیستم را با جمع‌آوری اطلاعات جدید بهینه می‌کند تا اقدامات بعدی دقیق‌تر و تأثیرگذارتر باشد.

۴-۶. مرحله نظارت توزیع‌شده^۳

یکی از ویژگی‌های مهم این الگوریتم، ماهیت حکمرانی غیرمتمرکز و مشارکتی آن است. برخلاف مدل‌های سنتی هوش مصنوعی که بر اساس یک مرکز تصمیم‌گیری واحد طراحی شده‌اند، الگوریتم امر به معروف مبتنی بر حکمرانی توزیع‌شده است. در این الگو، مسئولیت نظارت و اجرا میان سیستم‌های هوش مصنوعی در سطوح مختلف و ناظران انسانی تقسیم می‌شود. این رویکرد با اصل اسلامی مسئولیت جمعی و مفهوم «امت» سازگار بوده و به پایداری بیشتر سیستم حکمرانی کمک می‌کند.

۵. پارادایم قرآن‌بنیان در حکمرانی اخلاقی هوش مصنوعی

پارادایم قرآن‌بنیان، فراتر از یک الگوریتم، ساختاری جامع برای حکمرانی اخلاقی هوش مصنوعی ارائه می‌دهد. این پارادایم، اصول زیر را به‌عنوان محورهای حکمرانی اخلاقی معرفی می‌کند:

۵-۱. حکمرانی فعال و پیش‌دستانه^۴

برخلاف رویکردهای متداول که هوش مصنوعی را یک سیستم منفعل تعریف می‌کنند، این پارادایم برای

1. Feedback & Learning
2. Reinforcement Learning
3. Distributed Oversight
4. Proactive Governance

هوش مصنوعی وظایفی فعال تعریف می‌کند. هوش مصنوعی نه تنها باید از آسیب‌زدن خودداری کند، بلکه موظف است در جهت تحقق خیر و پیشگیری از منکر، به صورت فعالانه عمل کند. به عنوان مثال، هوش مصنوعی باید به جای صرفاً عدم تولید محتوای مضر، در مواجهه با تعاملات مخرب، با ارائه هشدار یا پیشنهاد جایگزین، موجب اصلاح محیط شود.

۵-۲. حکمرانی توزیع شده و چندسطحی^۱

مطابق اصل اسلامی «امت»، مسئولیت اخلاقی در این پارادایم بر یک شبکه متشکل از هوش‌های مصنوعی، کاربران و ناظران انسانی تقسیم می‌شود. این ساختار چندسطحی، از هوش‌های مصنوعی ساده (مانند فیلترهای اسپم) که به مقابله با رفتارهای ساده می‌پردازند تا مدل‌های پیچیده‌تر نظیر ابرهوش مصنوعی که تحلیل‌های عمیق‌تری انجام می‌دهند، امتداد دارد. این توزیع مسئولیت، انعطاف‌پذیری و مقاومت سیستم را در برابر نقاط شکست احتمالی افزایش می‌دهد.

۵-۳. اخلاق پویا و سلسله‌مراتبی

پارادایم قرآن‌بنیان، اخلاق را به عنوان یک فرآیند سلسله‌مراتبی و زمینه‌مدار مطرح می‌کند. هوش مصنوعی بسته به زمینه و موقعیت، مداخلات منعطف و تدریجی را در سه سطح قلبی، زبانی و عملی طراحی و اجرا می‌کند. این سلسله‌مراتب از اصول اسلامی الهام گرفته است که بر «تدریج» و «تناسب» تأکید ویژه‌ای دارد. این انعطاف‌پذیری اجازه می‌دهد که در شرایط مختلف اجتماعی یا فرهنگی، مداخلاتی مناسب‌تر و مؤثرتر طراحی شود.

۵-۴. حساسیت‌های اجتماعی - اخلاقی

در سلسله‌مراتب مداخلات اخلاقی، هوش مصنوعی موظف است تا شرایط زمینه‌ای، ظرفیت پذیرش کاربران و حساسیت‌های اجتماعی - اخلاقی را در اقدامات خود لحاظ کند. این قابلیت، حکمرانی اخلاقی مبتنی بر پارادایم قرآن‌بنیان را به سیستمی هوشمند و متناسب با زمینه‌های مختلف تبدیل می‌کند.

۵-۵. وظیفه‌گرایی فراتر از قواعد ثابت

این پارادایم، مفهوم اخلاق را فراتر از یک مجموعه قواعد ایستا ارتقاء می‌دهد و وظیفه‌ای فعال برای هوش مصنوعی قائل است. مطابق اصول امر به معروف و نهی از منکر، هوش مصنوعی باید به طور فعال در ترویج خیر و جلوگیری از شر در جامعه دیجیتال مشارکت کند، نه اینکه تنها به رعایت یک سری قواعد از پیش تعریف شده محدود شود.

۵-۶. انعطاف‌پذیری اقتضایی

بر اساس اصول فقه اسلامی، نظیر «تدریج»، «احتمال تأثیر» و «عدم ترتب مفسده‌ی بزرگ‌تر»، این الگوریتم به انعطاف‌پذیری هوشمندی دست می‌یابد که آن را از یکنواختی الگوریتمی رایج در دیگر چارچوب‌ها متمایز می‌سازد. این امر باعث می‌شود که هوش مصنوعی بتواند در سناریوهای پیچیده و متنوع، متناسب با حساسیت‌های اخلاقی و اجتماعی اقدام کند و با محاسبه‌ی عواقب ناخواسته، از افراط و تفریط جلوگیری نماید.

جدول ۲ - تطبیق انعطاف‌پذیری الگوریتمی با شرایط واقعی در حکمرانی هوش مصنوعی

شرط فقهی در امر به معروف و نهی از منکر	معادل در حکمرانی هوش مصنوعی	مثال کاربردی
علم به منکر: اطمینان از وقوع عمل ناشایست و عدم احتمال خطا	اطمینان از تشخیص: ^۱ سیستم هوش مصنوعی باید از صحت تشخیص خود با دقت و اطمینان بالا اطمینان حاصل کند.	سامانه تشخیص چهره برای شناسایی متخلفان باید به دقت بالای ۹۹٫۹۹٪ دست یابد تا از متهم کردن افراد بی‌گناه و وارد آوردن آسیب به حقوق آن‌ها جلوگیری شود.
احتمال تأثیر: اطمینان از اثربخشی اقدام در اصلاح یا پیشگیری از رفتار	محاسبه احتمال موفقیت: ^۲ سیستم باید احتمال موفقیت و تأثیر مثبت اقدام خود را پیش از اجرا ارزیابی کند.	چت‌بات اخلاقی که با تحلیل رفتار کاربر متوجه می‌شود تذکر مستقیم ممکن است باعث واکنش منفی شود، به جای آن از استراتژی غیرمستقیم یا لحن ملایم استفاده می‌کند.
عدم ترتب مفسده مهم‌تر: اقدام نباید موجب ایجاد آسیب گسترده‌تر شود.	ارزیابی عواقب ناخواسته: ^۳ سیستم باید پیامدهای احتمالی و اثرات دومینووار اقدامات خود را پیش‌بینی کند.	سیستم مدیریت ترافیک شهری برای پیشگیری از یک تخلف جزئی نمی‌تواند تصمیماتی اتخاذ کند که به ایجاد ترافیک گسترده، اتلاف وقت و منابع منجر شود.

۵-۷. نقش در طراحی و توسعه^۴

پارادایم امر به معروف و نهی از منکر، به‌عنوان یک مدل اخلاقی جامع و سلسله‌مراتبی، ظرفیت بالایی برای کاربرد در حوزه حکمرانی هوش مصنوعی دارد. این پارادایم با تکیه بر اصول اخلاقی اسلام و استفاده از ساختاری تطبیق‌پذیر و پویا، می‌تواند نقش مهمی در تمامی مراحل طراحی، استقرار و نظارت بر سیستم‌های هوش مصنوعی ایفا کند. تمرکز این نظریه بر انتقال ارزش‌های اخلاقی از حوزه مفاهیم نظری به اقدامات عملی است؛ اقدامی که نقش هوش مصنوعی را به‌عنوان عاملی فعال در حکمرانی اخلاقی بازتعریف می‌کند. طراحی و توسعه سیستم‌های هوش مصنوعی بر اساس اصول اخلاقی، نقش بنیادینی در شکل‌گیری رفتار اخلاقی این فناوری‌ها دارد. در این مرحله، پارادایم امر به معروف و نهی از منکر، دو ضرورت اساسی را برای طراحی معماری سیستم‌های هوش مصنوعی مطرح می‌کند:

1. Certainty of Detection
2. Probability of Success Calculation
3. Unintended Consequences Assessment
4. Design & Development

۵-۷-۱. درون‌سازی ارزش‌های اخلاقی^۱

این پارادایم خواستار آن است که اصول «معروف» (خیر و فضیلت) و «منکر» (شر و رفتار ناپسند)، نه به‌عنوان بخشی جداگانه و افزودنی در انتهای چرخه توسعه سیستم، بلکه به‌عنوان جزء ذاتی فرآیند طراحی و آموزش مدل‌های هوش مصنوعی گنجانده شوند. این مفهوم که تحت عنوان «اخلاق به‌وسیله طراحی»^۲ شناخته می‌شود، با رویکرد رایج «اخلاق الصافی»^۳ تفاوت جدی دارد. برای دستیابی به این هدف، لازم است که مدل‌های هوش مصنوعی بر اساس پارادایم اخلاقی قرآن‌بنیان آموزش داده شوند. پارادایم قرآن‌بنیان هوش مصنوعی، شامل اصول اسلامی عدالت، کرامت انسانی و مسئولیت‌پذیری اجتماعی است که در ساختار اولیه سیستم نقش ایفا می‌کند. برای نمونه، مدل‌های زبانی بزرگ می‌توانند بر اساس منابع اخلاقی اسلامی تنظیم^۴ شوند تا توانایی تشخیص رفتارها و انتخاب اقدامات اخلاقی را افزایش دهند.

۵-۷-۲. شفافیت معماری و قابلیت بازبینی^۵

در مرحله طراحی، سیستم باید دارای قابلیت شفافیت باشد، به‌گونه‌ای که بتواند فرآیندهای تصمیم‌گیری خود، از جمله نحوه تشخیص منکر و انتخاب نوع و سطح مداخله را توضیح دهد. این شفافیت، امکان ارزیابی اخلاقی سیستم توسط ناظران انسانی را فراهم می‌کند و به‌عنوان یک ویژگی بنیادین معماری سیستم در نظر گرفته می‌شود. شفافیت معماری موجب اعتمادسازی میان کاربران و تضمین پایبندی سیستم به اصول اخلاقی می‌شود.

۵-۸. نقش در استقرار و بهره‌برداری^۶

در مرحله استقرار، پارادایم امر به معروف و نهی از منکر تأکید دارد که هوش مصنوعی باید به‌عنوان یک عامل اخلاقی فعال در محیط واقعی عمل کند. این امر از دو طریق اصلی محقق می‌شود:

۵-۸-۱. حلقه بازخورد اخلاقی مستمر^۷

سیستم هوش مصنوعی باید به‌طور مداوم با محیط واقعی و کاربران تعامل داشته باشد و از طریق نتایج حاصل از این تعامل، مدل اخلاقی خود را پالایش کند. این فرآیند به‌جای رویکردی ایستا برای تأیید پیش از استقرار، از نظارتی پویا و یادگیری مداوم بهره می‌برد. این «حلقه بازخورد اخلاقی» با استفاده از

1. Intrinsic Value Embedding
2. Ethics by Design
3. Ethics Washing
4. Fine-tune
5. Transparent & Auditable Architecture
6. Deployment & Operation
7. Continuous Ethical Feedback Loop

تکنیک‌های یادگیری ماشین، به‌ویژه «یادگیری تقویتی»، امکان به‌روزرسانی و اصلاح مستمر مدل‌های اخلاقی را فراهم می‌آورد.

۵-۸-۲. ایجاد سلسله‌مراتب مسئولیت^۱

پارادایم امر به معروف و نهی از منکر، توزیع مسئولیت میان اجزای مختلف سیستم حکمرانی هوش مصنوعی را الزامی می‌داند. این سلسله‌مراتب شامل توسعه‌دهندگان، اپراتورها، کاربران و خود سیستم هوش مصنوعی می‌شود. هر یک از این اجزا، بر اساس سطح خود در چرخه الگوریتم اخلاقی، سهم مشخصی در تحقق حکمرانی اخلاقی دارند. این مدل مسئولیت سلسله‌مراتبی، هماهنگی میان انسان‌ها و ماشین‌ها را افزایش داده و امکان نظارت و اصلاح هم‌زمان اقدامات را فراهم می‌کند.

۶. نتیجه‌گیری

پارادایم قرآن‌بنیان حکمرانی اخلاقی هوش مصنوعی، با تکیه بر اصول و حیانی اسلام و آموزه‌های جامع اخلاقی قرآن کریم، چارچوبی نظری و عملی برای هدایت هوش مصنوعی به سوی تحقق عدالت، کرامت انسانی و مسئولیت‌پذیری اجتماعی ارائه می‌دهد. این پارادایم با استفاده از اصل امر به معروف و نهی از منکر به‌عنوان یک الگوریتم اخلاقی پویا و سیستمی، نه تنها از رویکردهای ایستا و محدود فاصله می‌گیرد، بلکه وظیفه‌ای فعال و پیش‌دستانه برای هوش مصنوعی تعریف می‌کند؛ وظیفه‌ای که علاوه بر پیشگیری از آسیب‌ها، بر ترویج خیر و اصلاح اجتماعی تأکید دارد. با ترسیم ساختاری سلسله‌مراتبی برای مداخلات اخلاقی در سه سطح قلبی، زبانی و عملی، این پارادایم بسترهای نوین فناورانه را به ابزارهایی اخلاق‌محور، تعالی‌بخش و تمدن‌ساز تبدیل می‌کند. این نظام با استفاده از تدریج و تناسب در مداخلات، انعطاف‌پذیری تناسبی و حساسیت به زمینه‌های اجتماعی-اخلاقی، می‌تواند ضمن کاهش تبعیض الگوریتمی، نقض حریم خصوصی و آسیب‌های اجتماعی، موجب ارتقای اعتماد عمومی و بهینه‌سازی عملکرد فناوری‌های هوشمند شود. علاوه بر این، چارچوب ارائه‌شده، هوش مصنوعی را از یک ابزار منفعل به یک عامل اخلاقی فعال تبدیل می‌کند که با شفافیت، بازبینی و حلقه‌های بازخورد مستمر، ظرفیت خودتنظیمی و یادگیری اخلاقی را به دست می‌آورد و امکان حکمرانی غیر متمرکز در جوامع انسانی را فراهم می‌سازد. این مدل با توجه به چالش‌های اخلاقی پیش‌رو، راهی بنیادین برای هدایت هوش مصنوعی به سوی مقاصد الهی، تعالی بشری و حفظ ارزش‌های اخلاقی ارائه می‌دهد.

منابع

قرآن کریم.

نهج البلاغه.

- ابن حمزه، محمد بن علی. (۱۴۰۸ق). *الوسيلة الى نيل الفضيلة*. تحقیق: محمد حسون. قم: کتابخانه آیت الله مرعشی نجفی.
- حرّ عاملی، محمد بن حسن. (۱۴۱۶ق). *وسائل الشيعة*. (ج ۱۶). تصحیح سید محمد رضا حسینی جلالی. قم: مؤسسه آل البيت (ع).
- حلّی، ابن ادریس. (۱۴۱۰ق). *السرائر الحاوی لتحرير الفتاوى*. (ج ۲). قم: مؤسسه النشر الاسلامی.
- خمینی، سید روح الله، (بی تا). *تحریر الوسيلة*. (ج ۱). تهران: مؤسسه تنظیم و نشر آثار امام خمینی.
- راغب اصفهانی، حسین بن محمد. (۱۴۱۲ق). *مفردات الفاظ القرآن*. تحقیق صفوان عدنان. بیروت: دارالقلم.
- طباطبایی، سید محمد حسین. (۱۳۷۴). *تفسیر المیزان*. (ج ۱۸). مترجم: سید محمد باقر موسوی همدانی. قم: دفتر انتشارات اسلامی.
- طبرسی، فضل بن حسن. (۱۴۱۲ق). *مجمع البیان فی تفسیر القرآن*. (ج ۱). بیروت: دار المعرفة.
- طبرسی، فضل بن حسن. (بی تا). *مجمع البیان فی تفسیر القرآن*. (ج ۲). بیروت: دارالفکر.
- طوسی، محمد بن حسن. (۱۴۰۹ق). *التبیان فی تفسیر القرآن*. (ج ۲). تحقیق احمد قصیر عاملی. بیروت: دار احیاء التراث العربی.
- علامه حلّی، حسن بن یوسف. (۱۴۱۴ق). *تذکرة الفقهاء*. (ج ۹). قم: مؤسسه آل البيت (ع).
- کلینی، محمد بن یعقوب. (۱۴۱۷ق). *الکافی*. (ج ۵). تهران: دار الکتب الإسلامية.
- مجلسی، محمد باقر. (۱۴۰۳ق). *بحار الأنوار*. (ج ۶۸، ۹۷). بیروت: دار احیاء التراث العربی.
- محقق حلّی، جعفر بن حسن. (۱۴۰۸ق). *شرايع الاسلام*. (ج ۱). قم: انتشارات اسماعیلیان.
- مطهری، مرتضی. (۱۳۷۲). *مجموعه آثار*. (ج ۲). تهران: انتشارات صدرا.
- مطهری، مرتضی. (۱۳۹۰). *اسلام و مقتضیات زمان*. تهران: انتشارات صدرا.
- نوری همدانی، حسین. (۱۳۷۷). *امربه معروف و نهی ازمنکر*. ترجمه محمد محمدی اشتهازی. قم: دفتر تبلیغات اسلامی حوزه علمیه قم.