


From Algorithmic Hegemony to Technological Sovereignty: A Critique of the Political Economy of the Internal Governance of Artificial Intelligence and the Design of the Proposed "Parsa Algorithm" Standard

Saiyed Mahdi Majidi Nezami

Assistant Professor, Department of Information Organization, Research Center for Islamic Information and Document Management, Islamic Sciences and Culture Academy, Qom, Iran. m.majidi@isca.ac.ir

Abstract



Amidst the global competition over "artificial intelligence regulation," the present article critically examines a model it terms "internal governance"; a model in which big tech companies are not merely subject to external rules, but are themselves the designers of norm-setting and standard-setting mechanisms in the field of artificial intelligence. The central problem of the research is to transcend the conventional dichotomy of "state regulation" versus "self-regulation"; because this dichotomy alone is incapable of demonstrating how ostensibly neutral technical standards can be transformed into instruments of political and economic power. The article argues that leading companies, by formulating discourses such as "safety," and relying on internal technical, policy, and organizational mechanisms, turn technical standards into mechanisms of exerting influence; in such a way that its consequence can be the formation of "algorithmic hegemony" and the consolidation of technological monopoly. From this perspective, the issue is not merely the regulation of a technology, but rather the power relations at the level of the political economy of technology and its geopolitical consequences: who writes the rules of the artificial intelligence game, how they globalize it, and place others in a position of subordination. The primary objective of the research is the critical analysis of "internal governance" in the domain of artificial intelligence and the exploration of the mechanisms through which big tech companies transform technical standards into instruments of political-economic power. Alongside this explanatory objective, the article also seeks to present an alternative: designing a "third model of governance" through proposing the "Parsa Algorithm" standard as a localized option against the dominant standards of Silicon Valley; an option that can provide a technical framework for auditing and assessing the compliance of artificial intelligence with the "fundamental values of Iran" and strengthen the possibility of transitioning from the status of a "passive consumer" to an "active technological governor." The current research seeks to answer the problem of: how does

Cite this article: Majidi Nezami, S.M. (2026). From Algorithmic Hegemony to Technological Sovereignty: A Critique of the Political Economy of the Internal Governance of Artificial Intelligence and the Design of the Proposed "Parsa Algorithm" Standard. *Islamic Governance Studies*, 2(1), p. 337-362. <https://doi.org/10.22081/jislamicgo.2026.74176.1061>

Received: 2025-09-07 ; **Revised:** 2025-11-19 ; **Accepted:** 2025-12-22 ; **Published online:** 2026-03-30

Type of article: Research Article
<https://jislamicgov.isca.ac.ir>

Publisher: Islamic Sciences and Culture Academy

©2026/authors retain the copyright and full publishing rights



the "internal governance" of large artificial intelligence companies transcend the state regulation/self-regulation dichotomy through transforming technical standards into mechanisms of influence, and lead to the creation of political-economic power and the consolidation of algorithmic hegemony? And based on this, how can the "third model of governance," relying on the proposed "Parsa Algorithm" standard, create a framework for auditing and value-orienting artificial intelligence in Iran as a localized alternative? This is a qualitative research conducted using the "critical discourse analysis" method and formulated within the framework of the "political economy of communications." The research data include the official, technical, and policy documents of three leading American companies—Open AI, Google DeepMind, and Anthropic—which are introduced in the article as the principal architects of the "safety" discourse as well as the governance model based on venture capital. The methodological focus of the research is on extracting the discursive logic and institutional mechanisms that, through documents and procedures, lead to standardization and rulemaking affecting the public policymaking environment. The findings of this research indicate that the internal governance model of big artificial intelligence companies is implemented through a three-layered technical, policy, and organizational architecture. This mechanism allows companies to restrict the entry of new actors into the technological ecosystem by creating structural barriers and increasing compliance costs. Internal governance, through the "soft capture of regulations," preemptively shapes the state legislative space and globalizes its desired standards in the form of the safety discourse and the values of the American technological ecosystem, so that formal regulations are often compelled to comply with the framework of leading companies. Consequently, algorithmic hegemony is formed; meaning that the values and criteria determined by American companies are accepted as a global model and indicator, and other countries and technological actors are driven to follow them. This mechanism, in addition to economic and policy dimensions, is also subject to geopolitical logic and operates as a project to preserve Western technological monopoly. Accordingly, the innovation of the research lies in introducing the "third model of governance" and the localized "Parsa Algorithm" standard, which provides a way to assess and audit artificial intelligence based on the fundamental values of Iran, and to transition from a space of passive consumerism to active technological sovereignty. The results suggest that focusing on the dichotomy of "state regulation" versus "self-regulation" is insufficient for understanding the reality of artificial intelligence governance; because big companies transform technical standards into instruments of exercising power through the path of "internal governance" and with a three-layered technical-policy-organizational architecture: both by raising compliance costs, by preempting the shaping of regulations, and by consolidating the algorithmic hegemony that globalizes the values of the American technological ecosystem. Against this situation, the proposal of the "Parsa Algorithm" as a localized standard opens the path for designing the "third model of governance" so that the possibility of technical auditing of artificial intelligence based on the fundamental values of Iran is provided, and the country moves from the role of a "passive consumer" toward "active technological sovereignty."

Keywords: Governance, Artificial intelligence, Parsa algorithm standard, Technological sovereignty, Political economy of technology, Algorithmic hegemony.

من الهيمنة الخوارزمية إلى السيادة التقنية: نقد الاقتصاد السياسي للحكومة الداخلية للذكاء الاصطناعي وتصميم المعيار المقترح «خوارزمية بارسا»

السيد مهدي مجيدي نظامي

أستاذ مساعد، قسم تنظيم المعلومات، معهد إدارة المعلومات والوثائق الإسلامية، مركز أبحاث العلوم والثقافة الإسلامية، قم، إيران.
m.majidi@isca.ac.ir

المخلص

في خضم المنافسة العالمية حول «تنظيم الذكاء الاصطناعي»، يتناول المقال الحالي بالنقد والدراسة نموذجاً يسميه «الحكومة الداخلية»؛ وهو نموذج لا تكون فيه شركات التكنولوجيا الكبرى مجرد تابع للقواعد الخارجية فحسب، بل هي نفسها من تصمم الآليات المعيارية والقياسية في مجال الذكاء الاصطناعي. وتتمثل المشكلة المحورية للبحث في تجاوز التنائية الرائجة المتمثلة في «التنظيم الحكومي» في مقابل «التنظيم الذاتي»؛ لأن هذه التنائية وحدها عاجزة عن تبيين: كيف يمكن للمعايير الفنية، التي تبدو محايدة، أن تتحول إلى أداة للسلطة السياسية والاقتصادية. ويستدل المقال على أن الشركات الرائدة، من خلال صياغة خطابات مثل «السلامة»، وبالاعتماد على آليات داخلية (فنية - سياساتية - تنظيمية)، تحول المعايير التقنية إلى آلية لممارسة النفوذ؛ بحيث يمكن أن تتمثل تداعيات ذلك في تبلور «الهيمنة الخوارزمية» وتثبيت الاحتكار التقني. ومن هذا المنظور، فإن المسألة ليست مجرد تنظيم تقنية ما، بل المسألة هي علاقات السلطة على مستوى الاقتصاد السياسي للتكنولوجيا وتدابيرها الجيوسياسية؛ من الذي يكتب قواعد لعبة الذكاء الاصطناعي، وكيف يجعلها عالمية، ويضع الآخرين في موقع التبعية. ويتمثل الهدف الرئيسي للبحث في التحليل النقدي لـ «الحكومة الداخلية» في مجال الذكاء الاصطناعي، واستكشاف الآليات التي تقوم شركات التكنولوجيا الكبرى من خلالها بتحويل المعايير الفنية إلى أداة للسلطة السياسية - الاقتصادية. وإلى جانب هذا الهدف التبييني، يسعى المقال أيضاً إلى تقديم بديل: تصميم «النموذج الثالث للحكومة» من خلال اقتراح معيار «خوارزمية بارسا التقنية» بوصفه خياراً محلياً في مواجهة المعايير المهيمنة لوادى السيليكون؛ خيار يمكنه توفير إطار فني لتدقيق وتقييم مدى تطابق الذكاء الاصطناعي مع «القيم الأساسية لإيران»، ويعزز إمكانية الانتقال من حالة «المستهلك المنفعل» إلى «الحاكم التقني الفعال». ويسعى البحث الحالي للإجابة عن التساؤل التالي: كيف تتجاوز «الحكومة الداخلية» لشركات الذكاء الاصطناعي الكبرى، من خلال تحويل المعايير الفنية إلى آليات للنفوذ، ثنائية التنظيم الحكومي/التنظيم الذاتي، وتؤدي إلى إيجاد قوة سياسية - اقتصادية وتثبيت الهيمنة الخوارزمية؟ وبناءً على ذلك، كيف يمكن لـ «النموذج الثالث للحكومة»، بالاعتماد على المعيار المقترح «خوارزمية بارسا»، أن يوجد بوصفه بديلاً محلياً إطاراً لتدقيق وتوجيه قيم الذكاء الاصطناعي في

استناداً إلى هذه المقالة: مجيدي نظامي، السيد مهدي (٢٠٢٢). من الهيمنة الخوارزمية إلى السيادة التقنية: نقد الاقتصاد السياسي للحكومة الداخلية للذكاء الاصطناعي وتصميم المعيار المقترح «خوارزمية بارسا». دراسات الحكومة الإسلامية، ١٢(١)، ص ٣٣٧-٣٤٢.

<https://doi.org/10.22081/jislamicgo.2026.74176.1061>

تاريخ الاستلام: ٢٠٢٥/٠٩/٠٧؛ تاريخ المراجعة: ٢٠٢٥/١١/١٩؛ تاريخ القبول: ٢٠٢٥/١٢/٢٢؛ تاريخ النشر: ٢٠٢٤/٠٣/٣٠

<https://jislamicgov.isca.ac.ir>

الناشر: المعهد العالي للعلوم والثقافة الإسلامية

نوع المقالة: بحثية



إيران؟ هذا البحث نوعي، وقد أنجز بمنهج «تحليل الخطاب النقدي»، وتمت صياغته في إطار «الاقتصاد السياسي للاتصالات». وتشتمل بيانات البحث على الوثائق الرسمية والفنية والسياساتية لثلاث شركات أمريكية رائدة - (Open AI)، و (Google DeepMind)، و (Anthropic) - والتي تُقدم في المقال بوصفها المهندس الرئيسي لخطاب «السلامة» وكذلك لنموذج الحوكمة القائم على رأس المال الاستثماري (المخاطر). وينصب التركيز المنهجي للبحث على استخراج المنطق الخطابي والآليات المؤسسية التي تفضي، من خلال الوثائق والإجراءات، إلى التقييس ووضع القواعد المؤثرة على بيئة السياسات العامة. وتُظهر نتائج هذا البحث أن نموذج الحوكمة الداخلية لشركات الذكاء الاصطناعي الكبرى يُنفذ من خلال بنية (معمارية) ثلاثية الطبقات: فنية، وسياساتية، وتنظيمية. وتسمح هذه الآلية للشركات بالحد من دخول فاعلين جدد إلى النظام البيئي التكنولوجي من خلال إيجاد عوائق هيكلية وزيادة تكاليف الامتثال. وتشكل الحوكمة الداخلية، من خلال «الاستيلاء الناعم على اللوائح»، فضاء التشريع الحكومي مسبقاً، وتجعل المعايير التي تبتغيها عالمية في قالب خطاب السلامة وقيم النظام البيئي التكنولوجي الأمريكي، بحيث تضطر اللوائح الرسمية غالباً إلى التبعية لإطار الشركات الرائدة. ونتيجة لذلك، تشكل الهيمنة الخوارزمية؛ أي تُقبل القيم والمعايير التي حدتها الشركات الأمريكية بوصفها نموذجاً ومؤشراً عالمياً، وتُدفع الدول والفاعلون التكنولوجيون الآخرون لاتباعها. وهذه الآلية، علاوة على أبعادها الاقتصادية والسياساتية، تخضع أيضاً للمنطق الجيوسياسي، وتعمل كمشروع للحفاظ على الاحتكار التقني للغرب. وبناءً على ذلك، يكمن ابتكار البحث في تقديم «النموذج الثالث للحوكمة» والمعيار المحلي «خوارزمية بارسا»، والذي يوفر طريقاً لتقييم وتدقيق الذكاء الاصطناعي بناءً على القيم الأساسية لإيران، والانتقال من فضاء الاستهلاك المنفعل إلى السيادة التقنية الفعالة. وتفيد النتائج بأن التركيز على ثنائية «التنظيم الحكومي» في مقابل «التنظيم الذاتي» ليس كافياً لفهم حقيقة حوكمة الذكاء الاصطناعي؛ لأن الشركات الكبرى تحول المعايير الفنية إلى أداة لممارسة السلطة عبر مسار «الحوكمة الداخلية» وبنية ثلاثية الطبقات (فنية-سياساتية-تنظيمية): سواء من خلال رفع تكاليف الامتثال، أو من خلال الاستباق في تشكيل اللوائح، أو من خلال تثبيت الهيمنة الخوارزمية التي تجعل قيم النظام البيئي التكنولوجي الأمريكي عالمية. وفي مواجهة هذا الوضع، يفتح اقتراح «خوارزمية بارسا» كمعيار محلي مسار تصميم «النموذج الثالث للحوكمة» لتوفير إمكانية التدقيق الفني للذكاء الاصطناعي بناءً على القيم الأساسية لإيران، وتحرك البلاد من دور «المستهلك المنفعل» نحو «السيادة التقنية الفعالة».

الكلمات المفتاحية: الحوكمة، الذكاء الاصطناعي، معيار خوارزمية بارسا، السيادة التقنية، الاقتصاد السياسي

للتكنولوجيا، الهيمنة الخوارزمية.

از هژمونی الگوریتمی تا حاکمیت فناورانه: نقد اقتصاد سیاسی حکمرانی درونی هوش مصنوعی و طراحی استاندارد پیشنهادی «الگوریتم پارسا»

سید مهدی مجیدی نظامی

استادیار، گروه سازماندهی اطلاعات، پژوهشکده مدیریت اطلاعات و مدارک اسلامی، پژوهشگاه علوم و فرهنگ اسلامی، قم، ایران.
m.majidi@isca.ac.ir

چکیده

در بحبوحه رقابت جهانی بر سر تنظیم‌گری هوش مصنوعی، این پژوهش مدل «حکمرانی درونی» را به‌طور انتقادی تحلیل کرده است. مسئله اصلی، عبور از دوگانه «تنظیم‌گری دولتی» در برابر «خودتنظیمی» و واکاوی سازوکارهای شرکت‌های بزرگ فناوری برای تبدیل استانداردهای فنی به ابزار قدرت سیاسی و اقتصادی است. این پژوهش کیفی با بهره‌گیری از روش «تحلیل گفتمان انتقادی» و در چارچوب اقتصاد سیاسی ارتباطات انجام شده است. داده‌ها شامل اسناد رسمی، فنی و سیاستی سه شرکت پیش‌روی آمریکایی (Google DeepMind, Anthropic, Open AI) به‌عنوان معماران اصلی گفتمان «ایمنی» و مدل حکمرانی مبتنی بر سرمایه‌گذاری خطرپذیر است. حکمرانی درونی از طریق معماری سه‌لایه فنی، سیاستی و سازمانی اعمال می‌شود و سه کارکرد اصلی دارد: ۱) ایجاد موانع ورود از طریق افزایش هزینه‌های انطباق، ۲) «تسخیر نرم مقررات» برای پیش‌دستی بر قانون‌گذاری دولتی، و ۳) تثبیت «هژمونی الگوریتمی» که ارزش‌های اکوسیستم فناوری آمریکا را جهانی‌سازی می‌کند. حکمرانی درونی یک پروژه ژئوپلیتیک برای حفظ انحصار فناورانه است. نوآوری پژوهش، طراحی «الگوی سوم حکمرانی» از طریق ارائه استاندارد پیشنهادی «الگوریتم پارسا» به‌عنوان بدیلی بومی در برابر استانداردهای سیلیکون‌ولی است. این استاندارد چارچوبی فنی برای ممیزی و سنجش انطباق هوش مصنوعی با ارزش‌های بنیادین ایران و گذار از «مصرف‌کننده منفعل» به «حکمران فناورانه فعال» فراهم می‌کند.

کلیدواژه‌ها: حکمرانی، هوش مصنوعی، استاندارد الگوریتم پارسا، حاکمیت فناورانه، اقتصاد سیاسی فناوری، هژمونی الگوریتمی.

استناد به این مقاله: مجیدی نظامی، سید مهدی (۱۴۰۵). از هژمونی الگوریتمی تا حاکمیت فناورانه: نقد اقتصاد سیاسی حکمرانی درونی هوش مصنوعی و طراحی استاندارد پیشنهادی «الگوریتم پارسا». *مطالعات حکمرانی اسلامی*، ۱۲(۱)، ص ۳۳۷-۳۶۲.

<https://doi.org/10.22081/jislamicgo.2026.74176.1061>

تاریخ دریافت: ۱۴۰۴/۰۶/۱۶؛ تاریخ اصلاح: ۱۴۰۴/۰۸/۲۸؛ تاریخ پذیرش: ۱۴۰۴/۱۰/۰۱؛ تاریخ انتشار آنلاین: ۱۴۰۵/۰۱/۱۰

<https://jislamicgov.isca.ac.ir>

ناشر: پژوهشگاه علوم و فرهنگ اسلامی

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



۱. مقدمه

در دهه اخیر، مدل‌های زبانی بزرگ^۱ و سامانه‌های هوش مصنوعی مولد^۲ سازوکارهای شناخت، ارتباط و تصمیم‌سازی انسانی را دگرگون کرده‌اند. در این فضا، مسئله اصلی تنها خطرات فنی و اجتماعی نیست، بلکه نحوه گذار از مصرف منفعلانه این زیرساخت‌های شناختی به سوی حکمرانی فعال بر آن‌ها است (IT for Change, 2023). در این مقاله، «حکمرانی» معادل Governance و متمایز از Government به کار می‌رود؛ Governance به شبکه‌ها و قواعد تنظیم‌گر (اعم از دولتی، شرکتی یا فنی) اشاره دارد، در حالی که Government بیانگر ساختار رسمی حاکمیت سیاسی است. این تمایز برای فهم «حکمرانی درونی»^۳ اهمیت دارد: الگویی که در آن لایه‌های فنی، قراردادی و سازمانی شرکت‌های بزرگ هوش مصنوعی نقشی مشابه قانون ایفا می‌کنند.

ادیبات جهانی حکمرانی هوش مصنوعی معمولاً از دو پارادایم سخن می‌گوید:

- مقررات‌گذاری سخت اتحادیه اروپا^۴ در قالب «قانون هوش مصنوعی»^۵ با رویکرد پیشگیری از خطر (European Commission, 2021, 2024).

- رویکرد بازارمحور یا خودتنظیمی^۶ در ایالات متحده که با تأکید بر نوآوری، مسئولیت اخلاقی را به صنعت واگذار می‌کند و «دستور اجرایی هوش مصنوعی»^۷ (۲۰۲۳ نمونه آن است Federal Register, 2023; The White House, 2023a).

اما این دوگانه برای فهم ساختار قدرت کافی نیست. در دل این دو رویکرد، سطح سومی شکل گرفته که «حکمرانی درونی» نام دارد: سازوکاری که در آن شرکت‌های بزرگ فناوری با ترکیبی از استانداردهای فنی، قراردادهای ساختارهای حاکمیت شرکتی، نوعی «حکومت خصوصی الگوریتمی»^۸ را بر زیرساخت‌های شناختی جهانی اعمال می‌کنند. در اینجا ارزش‌ها در کد نهادینه می‌شوند و معماری سیستم نقش تنظیم‌گر را برعهده می‌گیرد (Lessig, 1999). پژوهش حاضر رویکرد بازارمحور آمریکا را در قالب حکمرانی درونی تحلیل می‌کند و نشان می‌دهد چگونه این مدل به ساختاری از حکمرانی خصوصی الگوریتمی تبدیل شده است. گفتمان «ایمنی و اخلاق»^۹ نزد شرکت‌هایی مانند Open AI، Google و Anthropic صرفاً پروژه‌ای فنی نیست، بلکه راهبردی اقتصادی و ژئوپلیتیک (ناظر به رقابت

-
1. Large Language Models
 2. Generative AI
 3. Internal Governance
 4. Hard Regulation
 5. AI Act
 6. Self-Regulation
 7. Algorithmic Private Government
 8. Ethics & Safety

برای کنترل زیرساخت‌های دانشی، جریان داده و استانداردهای فنی اثرگذار بر موازنه قدرت در نظام بین‌الملل است: در سطح اقتصادی با تبدیل هزینه‌های زیرساختی و انطباقی به پیش شرط مسئولیت‌پذیری، به تمرکز قدرت و انحصار کمک می‌کند؛ و در سطح ژئوپلیتیک با کنترل زیرساخت‌های شناختی جهانی، اشکالی تازه از هژمونی و «استعمار داده‌ای»^۱ را بازتولید می‌نماید (Couldry & Mejias, 2019; Zuboff, 2019). برای کشورهایمانند ایران در موقعیت پیرامونی، مسئله «حاکمیت فناوریانه» اهمیتی ویژه دارد. اتکای منفعلانه به مدل‌های زبانی خارجی، خطر تعمیق وابستگی و تضعیف حاکمیت شناختی را به همراه دارد؛ و تقلید صرف از الگوهای سخت‌گیرانه اروپایی بدون برخورداری از زیرساخت داخلی، پاسخ‌گوی نیازها نخواهد بود. برخی پژوهش‌های حکمرانی فضای سایبر نیز بر ضرورت چارچوب‌های بومی تأکید کرده‌اند (شکوهیان‌راد، ۱۴۰۴). این پژوهش سه پرسش اصلی را دنبال می‌کند: (۱) معماری حکمرانی درونی شرکت‌های پیش‌رو از چه لایه‌های فنی، سیاستی و سازمانی تشکیل شده است؟

(۲) این سازوکارها چگونه کارکردهای ایجاد انحصار اقتصادی، تسخیر نرم مقررات و تثبیت هژمونی الگوریتمی را پیش می‌برند؟
(۳) پیامدهای این الگو برای کشورهایمانند ایران چیست؟ و چه راهبردهایی برای مواجهه فعال می‌توان پیشنهاد کرد؟

نوآوری پژوهش، عبور از تحلیل‌های صرفاً توصیفی و ارائه یک صورت‌بندی عملیاتی است. مقاله با تکیه بر چارچوب اقتصاد سیاسی انتقادی ارتباطات، از یک‌سو سازوکارهای حکمرانی درونی شرکت‌های منتخب را تحلیل می‌کند و از سوی دیگر با معرفی «الگوریتم پارسا» به‌عنوان کانون «الگوی سوم حکمرانی»، می‌کوشد شکاف میان نقد نظری استعمار داده‌ای و نیاز به یک پروتکل فنی (ممیزی بومی) را پر کرده و چارچوبی بدیل برای بازپس‌گیری حاکمیت شناختی در عصر هوش مصنوعی ارائه دهد. در ادامه، مهم‌ترین جریان‌های نظری حکمرانی فناوری و هوش مصنوعی مرور خواهد شد.

۲. پیشینه پژوهش و مبانی نظری

مطالعات حکمرانی فناوری (به‌ویژه حکمرانی هوش مصنوعی) را می‌توان در سه جریان اصلی صورت‌بندی کرد. هر یک بخشی از واقعیت را روشن می‌کنند، اما به‌تنهایی قادر به توضیح پیوند پیچیده میان فناوری، اقتصاد سیاسی و سلطه نیستند.

۲-۱. جریان فنی - اخلاقی

این جریان که در ادبیات مهندسی هوش مصنوعی غالب است، بر هم‌سوسازی رفتار مدل‌های زبانی با

ارزش‌ها و ترجیحات انسانی تمرکز دارد. روش «یادگیری تقویتی از بازخورد انسانی»^۱ به‌عنوان نمونه شاخص آن مطرح است (Christiano & et al., 2017). در این رویکرد، خطر اصلی هوش مصنوعی در انحراف از ارزش‌های انسانی تعریف می‌گردد؛ از این‌رو، راه‌حل در تنظیم پارامترهای یادگیری برای تربیت مدل‌های مطیع و بی‌خطر جست‌وجو می‌شود. با وجود اهمیت فنی، این ادبیات عموماً از تحلیل ساختارهای قدرت و انحصار غفلت می‌ورزد. پرسش‌هایی از جنس اینکه چه کسی ارزش‌های انسانی را تعریف می‌کند؟ چه گروه‌هایی در فرایند تدوین دستورالعمل‌ها و داده‌های آموزشی نمایندگی می‌شوند؟ و چه منافع سیاسی یا اقتصادی در پس‌فرایند همسوسازی نهفته است؟ اغلب به حاشیه رانده می‌شوند؛ پرسش‌هایی که در ادبیات انتقادی یادگیری ماشین، پررنگ‌تر مطرح شده‌اند (Birhane & et al., 2023).

۲-۲. جریان انتقادی و اقتصاد سیاسی فناوری

در مقابل، جریان انتقادی فناوری را در بستر ساختارهای اقتصادی و سیاسی قدرت بررسی می‌کند. نظریه «سرمایه‌داری نظارتی» زوبوف نشان می‌دهد که چگونه داده‌های رفتاری به سرمایه‌ای قابل مبادله و ابزار انباشت تبدیل شده‌اند (Zuboff, 2019). کودری و مجیاس با طرح مفهوم «استعمار داده‌ای»^۲ از تصرف داده‌ها و روابط اجتماعی، به‌جای سرزمین، به‌عنوان صورت‌بندی جدیدی از استعمار سخن می‌گویند (Coudry & Mejias, 2019). پژوهشگرانی مانند نوبل با اثر «الگوریتم‌های سرکوب»^۳ نشان داده‌اند که سازوکارهای فنی و الگوریتمی به‌طور ساختاری حامل سوگیری‌های نژادی و جنسیتی اند و نمی‌توان آن‌ها را بی‌طرف تصور کرد (Noble, 2018). در سطح نظری‌تر، فلوریدی با طرح «اخلاق اطلاعات» بر ضرورت تحلیل سازوکارهای قدرت در زیرساخت‌های دیجیتال تأکید می‌کند (Floridi, 2013). این جریان، اگرچه نسبت به پیامدهای قدرت‌محور فناوری حساس است، اما کمتر وارد سطح جزئیات فنی حکمرانی درونی شرکت‌ها می‌شود.

۳-۲. شکاف دانشی و نیاز به چارچوب تلفیقی

با وجود غنای این دو سنت، شکافی میان «تکنیک‌های فنی ایمنی و همسوسازی» و «تحلیل اقتصاد سیاسی انحصار و سلطه» باقی است. از یک‌سو، می‌دانیم که مدل‌ها از طریق سازوکارهایی مانند RLHF، پالایش داده و سیاست‌های استفاده چگونه ایمن می‌شوند؛ از سوی دیگر، فناوری می‌تواند ابزار سلطه و استعمار داده‌ای باشد، اما اتصال سیستماتیک این دو سطح کمتر روشن شده است (Bietti, 2020; Tan & et al., 2024). این پژوهش در چارچوب اقتصاد سیاسی انتقادی ارتباطات

1. Reinforcement Learning from Human Feedback (RLHF)

2. Data Colonialism

3. Algorithms of Oppression

(Mosco, 2009) می‌کوشد با تلفیق دو رویکرد، فناوری را به مثابه برساخته‌ای از روابط اجتماعی و تضادهای ساختاری فهم کند و نشان دهد که گفتمان «ایمنی و اخلاق» خود بخشی از راهبرد مشروعیت بخشی به قدرت است، نه الزاماً نشانه اخلاق مداری (Wagner, 2018; Bietti, 2020). برای صورت بندی این چارچوب، جست و جویی نظام مند در منابع فارسی و انگلیسی با مجموعه‌ای از کلیدواژه‌های نظری، فنی و ژئوپلیتیکی انجام شد، تا ابعاد مختلف، از معماری فنی تا پیامدهای اقتصادی، فرهنگی و ژئوپلیتیکی پوشش داده شود. در این فرایند، ترکیب‌هایی مانند «حکمرانی درونی هوش مصنوعی»، «حکومت خصوصی الگوریتمی»، «اقتصاد سیاسی انتقادی فناوری» و «گفتمان ایمنی و اخلاق در هوش مصنوعی» برای ردیابی تبدیل استانداردهای فنی شرکت‌های فراملی به ابزار قدرت و مشروعیت بخشی درون شرکتی به کار رفت. در حوزه فنی- اخلاقی، واژگانی نظیر یادگیری تقویتی از بازخورد انسانی، همسوسازی،^۱ سوگیری الگوریتمی^۲ و طبیعی سازی ایدئولوژی در مدل‌های زبانی انتخاب شد، تا پیوند میان روش‌های ایمن سازی مدل‌ها و فرایند تثبیت ارزش‌های مسلط بررسی شود. در محور ژئوپلیتیکی، اصطلاحاتی مانند «هژمونی الگوریتمی»، «استعمار داده‌ای»، «حاکمیت شناختی» و «انحصار زیرساختی در هوش مصنوعی» برای واکاوی پیامدهای سلطه فناورانه و جهانی سازی استانداردهای شرکت‌های بزرگ به کار رفت (Couldry & Mejias, 2019; Shad, 2025; National Security Commission on Artificial Intelligence, 2021). اینجا عملاً وارد حوزه‌ای می‌شویم که می‌توان آن را «ژئوپلیتیک داده و الگوریتم»^۳ نامید؛ حوزه‌ای که با «سایبوپلیتیک»^۴ هم پوشانی دارد و به جای کنترل سرزمین، بر کنترل زیرساخت‌های داده‌ای و شناختی متمرکز است. همچنین برای ابعاد حکمرانی و تنظیم‌گری، کلیدواژه‌هایی مانند «قانون هوش مصنوعی اتحادیه اروپا»،^۵ «خودتنظیمی آمریکا»،^۶ «تسخیر نرم مقررات»^۷ و «معماری سه لایه حکمرانی درونی»^۸ استفاده شد، تا امکان مقایسه این الگوها با «الگوی سوم»^۹ پیشنهادی مقاله فراهم شود (European Commission, 2024; Mäntymäki & et al., 2022; U.S. AI Regulation Study, 2024; Wei & et al., 2024). در راستای رویکرد بومی، ترکیب‌هایی مانند «الگوریتم پارسا»، «مدل زبانی بومی فارسی»، «ائتلاف‌های حاکمیت داده» و «دیپلماسی سایبری» نیز جست و جو شد که نشان داد درباره پیوند خاص میان مدل‌های بومی فارسی، حاکمیت داده و دیپلماسی

1. Alignment
2. Algorithmic Bias
3. The geopolitics of data and algorithms
4. Cyberpolitics
5. EU AI Act
6. US self-regulation
7. Regulatory Soft Capture
8. The three-layer architecture of internal governance

سایبری، هنوز ادبیات منسجم و مستقلی شکل نگرفته است. نتیجه این جست‌وجوها آن است که اگرچه مباحثی مانند انحصار فناوریانه، ایمنی مدل‌ها و سوگیری الگوریتمی به‌طور گسترده بررسی شده‌اند، پیوند تلفیقی میان لایه‌های فنی حکمرانی درونی و پیامدهای ژئوپلیتیک آن برای کشورهای پیرامونی همچنان کمتر مورد توجه قرار گرفته است. از این‌رو، پژوهش حاضر تلاش می‌کند با تکیه بر مجموعه کلیدواژه‌های فوق، تصویری جامع‌تر و انتقادی‌تر از معماری قدرت در صنعت هوش مصنوعی عرضه کند و زمینه نظری لازم را برای صورت‌بندی «الگوی سوم حکمرانی» و استاندارد پیشنهادی «الگوریتم پارسا» را فراهم سازد. برای فهم بهتر تفاوت‌های رویکردهای مسلط در حکمرانی هوش مصنوعی، می‌توان سه الگوی اصلی را در ادبیات سیاست‌گذاری فناوری، از یکدیگر تفکیک کرد: الگوی آمریکایی مبتنی بر خودتنظیمی بازار محور، الگوی اروپایی مبتنی بر تنظیم‌گری حقوقی سخت، و الگوی سوم پیشنهادی که بر توسعه زیرساخت‌های بومی و حاکمیت فناوریانه تأکید دارد. جدول (۱) این سه الگو را براساس چند معیار کلیدی از جمله فلسفه سیاستی، بازیگران اصلی، ابزارهای حکمرانی و پیامدهای ژئوپلیتیک مقایسه می‌کند.

جدول ۱- مقایسه سه‌گانه الگوهای حکمرانی هوش مصنوعی

معیار مقایسه	الگوی آمریکایی (خودتنظیمی لیبرال)	الگوی اروپایی (تنظیم‌گری سخت)	الگوی سوم پیشنهادی (حاکمیت فناوریانه بومی)
فلسفه اصلی	نوآوری بدون مجوز (Permissionless Innovation) و رهبری بازار	حفاظت از حقوق شهروندان و مدیریت ریسک	امنیت ملی، حاکمیت شناختی و استقلال فرهنگی
بازیگران اصلی	شرکت‌های بزرگ فناوری (Open AI, Google)	نهادهای تنظیم‌گر دولتی (مانند کمیسیون اروپا)	دولت (به‌عنوان سرمایه‌گذار)، نهادهای ملی و شرکت‌های بومی
تمرکز اصلی	رشد سریع، توسعه مدل‌های پایه و تسخیر بازار	تنظیم «کاربردهای» هوش مصنوعی براساس سطح ریسک	تملک و توسعه «زیرساخت‌های» تولید (مراکز داده، مدل‌های پایه)
ابزار کلیدی	حکمرانی درونی، استانداردهای شرکتی، لابی‌گری (تسخیر نرم مقررات)	قانون‌گذاری جامع و سخت‌گیرانه (مانند EU AI Act)	سرمایه‌گذاری دولتی، استاندارد ملی (الگوریتم پارسا)، دیپلماسی فناوریانه
نقطه ضعف/انتقاد	ایجاد انحصار، هژمونی فرهنگی، تضعیف حاکمیت دولت‌ها	کند کردن نوآوری، تمرکز بر مصرف‌کننده و غفلت از لایه زیرساخت	نیاز به سرمایه‌گذاری عظیم، ریسک انزوای فناوریانه در کوتاه‌مدت
نتیجه ژئوپلیتیک	تثبیت هژمونی الگوریتمی و استعمار داده‌ای	تبدیل شدن به زمین بازی تنظیم‌گری برای دیگران	شکستن انحصار، ایجاد قطب مستقل و ائتلاف‌سازی با کشورهای همسو

همان‌گونه که در جدول (۱) مشاهده می‌شود، الگوی آمریکایی عمدتاً بر اصل «نوآوری بدون مجوز» استوار است؛ رویکردی که در آن شرکت‌های بزرگ فناوری نقش بازیگران اصلی را برعهده دارند و دولت نقش نسبتاً محدودی در تنظیم مستقیم بازار ایفا می‌کند. نتیجه این الگو شکل‌گیری اکوسیستم‌های

نوآوری بسیار پویا و توسعه سریع مدل‌های پایه هوش مصنوعی بوده است. با این حال، منتقدان معتقدند این رویکرد به تمرکز قدرت در دست چند شرکت بزرگ، شکل‌گیری انحصارهای پلتفرمی و نوعی هژمونی الگوریتمی در سطح جهانی منجر شده است. در مقابل، اتحادیه اروپا رویکردی مبتنی بر تنظیم‌گری سخت و حفاظت از حقوق شهروندان را دنبال می‌کند. در این الگو، دولت‌ها و نهادهای فراملی تنظیم‌گر نقش محوری دارند و تلاش می‌کنند از طریق قوانین جامع، به‌ویژه قانون هوش مصنوعی اتحادیه اروپا^۱، کاربردهای هوش مصنوعی را براساس سطح ریسک دسته‌بندی و کنترل کنند. اگرچه این رویکرد می‌تواند به افزایش اعتماد عمومی و حفاظت از حقوق کاربران کمک کند، اما برخی پژوهشگران آن را عامل کند شدن نوآوری و عقب‌ماندن اروپا در رقابت فناوری می‌دانند. الگوی سوم بر «حاکمیت فناورانه بومی» تأکید دارد؛ رویکردی که به‌جای تمرکز صرف بر تنظیم کاربردها، بر تملک و توسعه زیرساخت‌های اصلی هوش مصنوعی - از مراکز داده و توان محاسباتی تا مدل‌های پایه - استوار است. در این چارچوب، دولت نقش سرمایه‌گذار و تسهیل‌گر را برعهده دارد و با استانداردهای ملی، سرمایه‌گذاری راهبردی و دیپلماسی فناورانه، به دنبال ایجاد ظرفیت‌های مستقل است. این الگو می‌تواند وابستگی را کاهش و استقلال فناورانه را تقویت کند، اما به سرمایه‌گذاری سنگین و مدیریت خطر انزوای کوتاه‌مدت نیاز دارد. مقایسه این سه الگو نشان می‌دهد که حکمرانی هوش مصنوعی صرفاً یک مسئله فنی یا اقتصادی نیست، بلکه به‌طور مستقیم با ملاحظات ژئوپلیتیک، حاکمیت داده و ساختار قدرت در نظام بین‌الملل پیوند خورده است.

۳. روش پژوهش

این پژوهش در چارچوب پارادایم انتقادی و با استفاده از روش کیفی «تحلیل گفتمان انتقادی»^۲ و براساس مدل سه‌بعدی نورمن فرکلاف انجام شده است. در این رویکرد، زبان نه صرفاً ابزار انتقال معنا، بلکه شکل خاصی از «عمل اجتماعی» تلقی می‌شود که در تولید و بازتولید روابط قدرت، هویت‌ها و نظام‌های دانشی نقش ایفا می‌کند (Fairclough, 2010). در حوزه فناوری‌های نوین و به‌ویژه هوش مصنوعی، تحلیل گفتمان انتقادی امکان آشکارسازی نحوه پنهان شدن تصمیمات سیاسی و اقتصادی در قالب زبان تکنوکراتیک را فراهم می‌کند و نشان می‌دهد چگونه مفاهیم فنی می‌توانند در فرآیند «طبیعی‌سازی ایدئولوژی» به‌کار گرفته شوند.

۳-۱. پیکره داده‌ها و نمونه‌گیری

پیکره داده‌های پژوهش شامل اسناد رسمی، فنی و سیاستی بازیگران اصلی حوزه هوش مصنوعی مولد در

1. EU AI Act

2. Critical Discourse Analysis (CDA)

بازه زمانی ۲۰۱۷ (هم‌زمان با معرفی معماری ترنسفورمر^۱) تا ۲۰۲۵ است. نمونه‌ها با روش نمونه‌گیری هدفمند و معیار محور انتخاب شده‌اند. این پیکره عمدتاً شامل گزارش‌های ایمنی مدل‌ها، چارچوب‌های سیاستی، اسناد راهنمای مدیریت ریسک و شرایط استفاده از خدمات است که همگی از منابع رسمی و عمومی این شرکت‌ها استخراج شده‌اند. اسناد براساس ارتباط مستقیم با سازوکارهای حکمرانی درونی، سیاست‌های ایمنی و نحوه صورت‌بندی گفتمان «ایمنی» انتخاب گردیدند. تمرکز پژوهش بر «مدل سیلیکون‌ولی» در توسعه هوش مصنوعی است؛ مدلی که با اتکاء به سرمایه‌گذاری خطرپذیر، راهبرد رشد سریع و گفتمان «نوآوری بدون مجوز»^۲ شناخته می‌شود. بر این اساس، اسناد منتشر شده توسط سه شرکت Open AI، Google DeepMind و Anthropic به‌عنوان نمایندگان شاخص این پارادایم برای تحلیل انتخاب شدند.

۳-۲. مراحل تحلیل گفتمان

تحلیل داده‌ها براساس سه سطح پیشنهادی فرکلاف انجام شده است:

- **سطح توصیف (متن):** بررسی ویژگی‌های زبانی و واژگانی اسناد برای فهم نحوه بازنمایی مفاهیم کلیدی؛ مانند ترجیح «ایمنی» به جای «محدودسازی/سانسور» و استفاده از ساختارهای مجهول برای پنهان کردن فاعل تصمیم‌گیرنده.

- **سطح تفسیر (عمل گفتمانی):** تحلیل فرایند تولید، توزیع و مصرف متون؛ از جمله گذار انتشار دانش از مقالات داوری شده به سفیدنامه‌های شرکتی، گزارش‌های سیاستی و وبلاگ‌ها به‌عنوان ابزار شکل‌دهی مستقیم به گفتمان عمومی و سیاست‌گذاری.

- **سطح تبیین (عمل اجتماعی):** پیوند نتایج گفتمانی با ساختارهای کلان اقتصاد سیاسی و ژئوپلیتیک فناوری نشان می‌دهد که مفاهیمی مانند همسوسازی و RLHF در حکمرانی درونی شرکت‌ها می‌تواند به تثبیت «هژمونی الگوریتمی» و ایجاد موانع ورود برای رقبا از طریق افزایش هزینه‌های زیرساختی و انطباق منجر شوند.

۳-۳. استفاده از ابزارهای هوش مصنوعی در فرایند تحلیل

با توجه به حجم زیاد داده‌های متنی، در برخی مراحل از مدل‌های زبانی بزرگ به‌عنوان ابزار کمکی برای شناسایی اولیه الگوهای واژگانی و مفهومی، دسته‌بندی مقدماتی مضامین و جست‌وجوی منابع استفاده شد. با این حال، تحلیل گفتمانی، تفسیر نتایج و استنتاج نظری به‌طور کامل توسط پژوهشگر انجام شد و ابزارهای هوش مصنوعی صرفاً نقش پشتیبان در سازمان‌دهی داده‌ها و تسهیل مرور منابع داشتند.

1. Transformer Architecture
2. Permissionless Innovation

۴. یافته‌ها

تحلیل داده‌ها براساس چارچوب تحلیل گفتمان انتقادی فرکلاف نشان می‌دهد که «حکمرانی درونی» در صنعت هوش مصنوعی نه یک سازوکار پراکنده یا صرفاً واکنشی، بلکه نوعی معماری نهادی طراحی شده برای اعمال و تثبیت کنترل است. این معماری در سه لایه به هم پیوسته عمل می‌کند که هر یک کارکردی مکمل در تولید و بازتولید نظم حکمرانی شرکت‌محور دارند: لایه فنی، لایه سیاستی و لایه سازمانی.

۴-۱. لایه فنی: الگوریتم به مثابه قانون

در عمیق‌ترین سطح، حکمرانی از طریق معماری فنی سامانه‌ها اعمال می‌شود و «یادگیری تقویتی از بازخورد انسانی» (RLHF) در این میان نقشی محوری دارد. در اسناد شرکت‌ها، این روش با مفاهیمی مانند «ایمنی»، «مفید بودن» و «کاهش آسیب» توصیف می‌شود (Open AI, 2023b) و چارچوب معنایی آن بیشتر بر مدیریت خطر و مسئولیت‌پذیری فنی تأکید دارد، در حالی که پیامدهای هنجاری کمتر به صراحت طرح می‌شوند. با این حال، بررسی پژوهش‌های فنی نشان می‌دهد که RLHF در عمل سازوکاری برای شکل‌دهی پیشینی به دامنه پاسخ‌های مدل است؛ به گونه‌ای که مدل نه تنها می‌آموزد چه بگوید، بلکه از تولید برخی پاسخ‌ها نیز اجتناب می‌کند. در «هوش مصنوعی قانون اساسی» نیز تأکید شده که مدل باید براساس مجموعه‌ای از اصول هنجاری هدایت شود (Bai & et al., 2022). بدین ترتیب، ترجیحات هنجاری از سطح ارزش‌ها به سطح معماری فنی منتقل و در پارامترهای مدل تثبیت می‌شوند. در چارچوب نظری لسیگ، این وضعیت مصداق گزاره «کد، قانون است» است (Lessig, 1999). با این تفاوت که، قواعد نهفته در معماری الگوریتمی معمولاً برای کاربران شفاف یا قابل اعتراض نیستند؛ از این رو، بخشی از فرایند حکمرانی از مقررات قابل مشاهده به سطح نامرئی طراحی فنی منتقل می‌شود.

۴-۲. لایه سیاستی: چارچوب‌های قراردادی و بوروکراسی دیجیتال

در لایه دوم، اسناد حقوقی و سیاستی شرکت‌ها (از جمله «شرایط استفاده» و «خط‌مشی‌های استفاده») به عنوان چارچوبی حقوقی برای اعمال و تثبیت محدودیت‌های فنی عمل می‌کنند (Google, 2024; Open AI, 2025). تحلیل این اسناد نشان می‌دهد که آن‌ها عمدتاً در قالب قراردادهای الحاقی تنظیم شده‌اند؛ به این معنا که کاربر تنها می‌تواند شرایط ارائه‌شده را بپذیرد یا از استفاده از خدمت صرف‌نظر کند. چنین ساختاری می‌تواند به تغییر تدریجی توازن قدرت قراردادی به سود پلتفرم‌ها منجر شود (Radin, 2013). تحلیل گفتمانی این اسناد همچنین نشان می‌دهد که مفاهیم کلیدی مانند «محتوای مضر»، «اطلاعات گمراه‌کننده» یا «استفاده نامناسب» اغلب به صورت کلی و تفسیرپذیر تعریف می‌شوند. برای نمونه، سیاست‌های استفاده شرکت‌ها فهرست‌هایی از انواع محتوای ممنوع (از جمله خشونت، نفرت‌پراکنی، فعالیت‌های غیرقانونی یا تولید اطلاعات گمراه‌کننده) ارائه می‌کنند، اما در

حوزه‌های پیچیده یا مناقشه‌برانگیز، معیارهای عملیاتی روشنی برای تمایز میان بحث مشروع، خطای علمی یا اطلاعات گمراه‌کننده ارائه نمی‌دهند (Google, 2024; Open AI, 2025). افزون‌بر این، نحوه اجرای این سیاست‌ها نیز نشان‌دهنده نوعی عدم تقارن نهادی است. در شرایط استفاده بسیاری از این خدمات تصریح شده که شرکت می‌تواند در صورت تشخیص نقض سیاست‌ها، دسترسی کاربر را تعلیق یا خاتمه دهد (Google, 2024; Open AI, 2024a). هرچند سازوکارهایی برای تماس با پشتیبانی یا ثبت اعتراض ذکر شده‌اند، این فرایندها معمولاً تحت نظارت مستقیم همان شرکت‌ها قرار دارند و با الگوهای کلاسیک «دائرسی مستقل» فاصله دارند. در نتیجه، لایه سیاستی را می‌توان نوعی چارچوب حقوقی حکمرانی پلتفرمی دانست که در آن شرکت‌ها نقش تنظیم‌گر، مجری و داور را به‌طور هم‌زمان ایفا می‌کنند؛ وضعیتی که با نگرانی‌های مطرح‌شده درباره فرسایش تدریجی حقوق کاربران در قراردادهای دیجیتال هم‌خوان است (Radin, 2013).

۴-۳. لایه سازمانی: سازوکارهای مشروعیت‌بخش

لایه سوم به سازوکارهای سازمانی و نهادی مربوط است که نقش مهمی در مدیریت ادراک عمومی و تعامل با نهادهای تنظیم‌گر ایفا می‌کنند. در سال‌های اخیر بسیاری از شرکت‌های فناوری سازوکارهایی مانند «تیم‌های قرمز»،^۱ «کمیته‌های اخلاق هوش مصنوعی» و چارچوب‌های «هوش مصنوعی مسئولانه» را ایجاد کرده‌اند (Google, 2023; Microsoft, 2025). در سطح رسمی، این ساختارها به‌عنوان ابزارهایی برای شناسایی آسیب‌پذیری‌ها، افزایش ایمنی سامانه‌ها و ارتقای پاسخگویی معرفی می‌شوند. با این حال، تحلیل انتقادی ادبیات موجود نشان می‌دهد که چنین سازوکارهایی می‌توانند کارکردی فراتر از مدیریت فنی ریسک داشته باشند. پژوهش‌های انتقادی درباره «اخلاق‌شویی»^۲ استدلال می‌کنند که برخی ابتکارات اخلاقی شرکتی ممکن است به کاهش فشار برای تنظیم‌گری بیرونی یا ایجاد تصویر عمومی از خودتنظیمی مؤثر کمک کنند (Wagner, 2018). در این چارچوب، لایه سازمانی را می‌توان بخشی از معماری حکمرانی درونی دانست که از طریق تولید گزارش‌ها، چارچوب‌های اخلاقی و سازوکارهای ارزیابی داخلی، به مشروعیت‌بخشی به تصمیمات فنی و سیاستی شرکت‌ها کمک می‌کند و تصویر نهادی از مسئولیت‌پذیری و خودتنظیمی ایجاد می‌نماید.

یافته‌ها نشان می‌دهد که حکمرانی درونی در شرکت‌های پیش‌روی هوش مصنوعی از سه لایه به‌هم‌پیوسته تشکیل شده است:

– لایه فنی که با تعبیه ارزش‌ها و محدودیت‌ها در معماری مدل‌ها نوعی کنترل پیشینی و نامرئی ایجاد می‌کند.

1. Red Teams

2. Ethics Washing

- لایه سیاستی که از طریق شرایط استفاده، تعاریف حقوقی و چارچوب‌های سیاستی برای این کنترل پوشش قراردادی و اجرایی فراهم می‌آورد.

- لایه سازمانی که با تیم‌های ارزیابی، گزارش‌های ایمنی و سازوکارهای خودتنظیمی، از طریق تولید اعتماد نهادی و مدیریت ادراک عمومی به این نظم مشروعیت می‌بخشد.

ترکیب این سه لایه، معماری‌ای از اقتصاد سیاسی حکمرانی را شکل می‌دهد که در آن کد، قرارداد و سازمان هم‌زمان برای مدیریت ریسک، تنظیم دسترسی به فناوری و تثبیت قدرت ساختاری شرکت‌های فناوری عمل می‌کنند. این وضعیت نشان می‌دهد که حکمرانی درونی صرفاً مجموعه‌ای از ابزارهای فنی نیست، بلکه بخشی از یک رژیم حکمرانی شرکت محور است.

۵. بحث و تفسیر نتایج

۵-۱. کارکردهای اقتصاد سیاسی حکمرانی درونی

این بخش بررسی می‌کند که چگونه سازوکارهای سه‌گانه حکمرانی درونی-فنی، سیاستی و سازمانی- فراتر از نیت فردی توسعه‌دهندگان، در سطح ساختاری به بازتولید قدرت و ثروت می‌انجامند. میان دغدغه‌های واقعی پژوهشگران درباره ریسک‌های هوش مصنوعی (مانند نفرت‌پراکنی یا بدافزار) و منطق ساختاری بازار تمایز وجود دارد. در چارچوب الیگاپولی صنعت، این نگرانی‌ها اغلب به‌طور ساختاری به سود شرکت‌های بزرگ مصادره می‌شود. بدین ترتیب، «ایمنی» از یک ارزش اخلاقی به ابزاری برای رقابت و حذف رقیب تبدیل می‌شود؛ مسئله اصلی نه خود ایمنی، بلکه «سیاسی‌شدن ایمنی» در خدمت مهندسی منافع اقتصادی است. یافته‌های پژوهش نشان می‌دهد که حکمرانی درونی سه کارکرد راهبردی کلان را محقق می‌کند:

۵-۱-۱. مکانیزم اقتصادی انحصار: سد ورود ساختاری

تحلیل هزینه‌های مرتبط با لایه‌های سه‌گانه حکمرانی نشان می‌دهد که این هزینه‌ها صرفاً ملاحظات کیفی نیستند، بلکه یک واقعیت حسابداری‌اند که به تمرکز قدرت و تثبیت انحصار کمک می‌کنند (Zuboff, 2019).

تبیین مکانیزم: فرایندهایی نظیر RLHF (استخدام گسترده نیروهای انسانی برای برچسب‌زنی)، نگهداری «تیم‌های قرمز»^۱ دائمی و مدیریت «انطباق حقوقی»^۲، «هزینه‌های ثابت»^۳ توسعه مدل را به‌طور چشمگیری افزایش می‌دهند.

1. Red teams
2. Compliance
3. Fixed Costs

تحلیل انتقادی: در اقتصاد سیاسی، این پدیده یادآور استراتژی تاریخی «محصور کردن»^۱ است. همان‌طور که در انقلاب صنعتی زمین‌های عمومی حصارکشی شدند، امروز «قلمرو دانش و داده» با حصارهای پرهزینه ایمنی محصور می‌شوند. این هزینه‌ها به‌عنوان یک «مانع ورود ساختاری»^۲ عمل کرده و رقابت را برای استارت‌آپ‌ها، دانشگاه‌ها و بازیگران «متن‌باز»^۳ ناممکن می‌سازند؛ امری که نتیجه نهایی آن تثبیت ساختار بازار «الیگاپولی»^۴ در دست چند بازیگر معدود است.

۵-۱-۲. تسخیر نرم مقررات: مهندسی مشروعیت و مرجعیت اپیستمیک

یافته‌های مربوط به لایه سازمانی پدیده‌ای را آشکار می‌کند که می‌توان آن را «تسخیر نرم مقررات»^۵ نامید. برخلاف تسخیر سخت که از طریق لابی مستقیم صورت می‌گیرد، تسخیر نرم از طریق «شکل‌دهی به گفتمان کارشناسی»^۶ عمل می‌کند. شرکت‌های بزرگ با تعریف استانداردهای انحصاری ایمنی، خود را به‌عنوان تنها مراجع ذی‌صلاح برای تنظیم‌گری معرفی می‌کنند. در سطح تحلیلی، برجسته‌سازی سناریوهای فاجعه‌آمیز و خطرات وجودی،^۷ فضای سیاسی را به سمت الگوهای شدید تنظیم‌گری مانند «مجوزدهی اجباری» سوق می‌دهد؛ رژیم‌هایی که از نظر فنی و نهادی تنها توسط بازیگران بزرگ با ظرفیت بالای انطباق قابل اجرا هستند (Guha & et al., 2024). بدین ترتیب، «اخلاق‌هراسی» پیرامون هوش مصنوعی می‌تواند به مقرراتی منجر شود که نه تنها نوآوری را محدود می‌کند، بلکه به‌طور ناخواسته موقعیت انحصاری شرکت‌های مسلط را در برابر رقبای کوچک‌تر بیمه می‌نماید (Inserra, 2024).

۵-۱-۳. استعمار داده‌ای و هژمونی الگوریتمی: استقرار نظم نوین

ترکیب مفاهیم مبهمی چون «مضرت» یا «گمراه‌کننده» با اختیار یک‌جانبه پلتفرم برای مسدودسازی حساب، زمینه را برای اجرای گزینشی سیاست‌ها فراهم می‌کند. این فرآیند از چند منظر قابل تحلیل است:

انتقال ریسک و خودنظارتی: زبان خاص بندهای سلب مسئولیت، بار اصلی ریسک حقوقی را بر دوش کاربر می‌گذارد (Google, 2024; Open AI, 2024a). این ساختار، کاربران را به «خودنظارتی» وامی‌دارد؛ کاربر برای پرهیز از تعلیق، ناچار است رفتار و گفتار خود را با تفسیر احتمالی پلتفرم وفق دهد و هنجارهای مطلوب آن را درونی کند. این وضعیت نوعی «اثر ارعاب»^۸ ایجاد می‌کند که کاربران را به

1. Enclosure
2. Structural Barrier to Entry
3. Open Source
4. Oligopolistic
5. Soft Regulatory Capture
6. Epistemic Capture
7. Existential Risks
8. Chilling Effect

مجریان درونی محدودیت‌ها بدل می‌سازد (Fairclough, 2010; Radin, 2013).

استعمار در قالب تنظیمات پیش فرض: نتیجه نهایی این فرایند، شکل‌گیری «هژمونی الگوریتمی» است. وقتی یک مدل زبانی در بستر اکوسیستم «سیلیکون‌ولی» آموزش می‌بیند، سوگیری‌های فرهنگی و هنجارهای سیاسی - اجتماعی مسلط (از فردگرایی تا لیبرالیسم بازار) به صورت «تنظیمات پیش فرض»^۱ در مدل تثبیت می‌شوند (Birhane & et al., 2023). این نه یک گزینه، بلکه یک زیرساخت شناختی است که بر مبنای «استعمار داده‌ای» عمل می‌کند (Couldry & Mejias, 2019).

در کشورهایی چون ایران که خارج از مدار نهادهای قانون‌گذار غربی‌اند، تکیه بر مدل‌های خارجی به معنای پذیرش ضمنی مرجعیت هنجاری بازیگرانی است که در چارچوب منافع ژئوپلیتیک خود عمل می‌کنند (National Security Commission on AI, 2021). زمانی که این سامانه‌ها تعیین می‌کنند چه متنی «مضر» یا «قابل اعتماد» است، قدرت روایت‌گری و سیاست‌گذاری فرهنگی از نهادهای ملی به زیرساخت‌های شرکتی منتقل می‌شود، که نتیجه آن تضعیف «امنیت شناختی» است؛ زیرا توان جامعه برای تعریف مستقل واقعیت اجتماعی زیر سیطره نظامی الگوریتمی قرار می‌گیرد که توسط بازیگران جهانی مهندسی شده است (Shad, 2025; Floridi, 2013).

۲-۵. الگوی ملی حاکمیت فناوریانه: باز تعریف نقش دولت در حکمرانی هوش مصنوعی

در برابر دو الگوی مسلط حکمرانی هوش مصنوعی («خودتنظیمی بازارمحور آمریکایی» و «تنظیم‌گری مصرف‌محور اروپایی») این پژوهش ضرورت طرح یک «الگوی سوم» را مطرح می‌کند؛ الگویی مبتنی بر حاکمیت فناوریانه که دولت را نه صرفاً تنظیم‌گر منفعل، بلکه توانمندساز راهبردی زیرساخت‌های دیجیتال می‌داند. در این چارچوب، دولت نقشی فعال در توسعه ظرفیت‌های فنی، شکل‌دهی به استانداردها و صیانت از حاکمیت داده و زیرساخت ایفا می‌کند. این الگو بر سه ستون اصلی استوار است: حاکمیت بر زیرساخت‌های فناوریانه، اقتصاد سیاسی مقاومتی در برابر انحصار پلتفرم‌ها، و کثرت‌گرایی ارزشی در برابر هژمونی هنجاری مدل‌های مسلط.

۱) **حاکمیت بر زیرساخت به مثابه امنیت ملی:** برخلاف الگوی اروپایی که عمدتاً بر تنظیم «کاربردهای» هوش مصنوعی تمرکز دارد، الگوی سوم بر ضرورت تملک و توسعه «زیرساخت‌های» تولید (مدل‌های پایه، مراکز داده) به عنوان امری حاکمیتی تأکید می‌کند. در این نگاه، استقلال فناوریانه یک انتخاب اقتصادی نیست، بلکه یک ضرورت امنیت ملی برای حفظ حاکمیت شناختی و فرهنگی است.

۲) **اقتصاد سیاسی مقاومتی:** در مقابل الگوی آمریکایی که فضا را برای «تسخیر نرم مقررات» توسط شرکت‌های بزرگ باز می‌گذارد، الگوی سوم یک رویکرد اقتصاد سیاسی فعال و مقاومتی را اتخاذ می‌کند.

این رویکرد شامل سرمایه‌گذاری دولتی هدفمند برای شکستن انحصار، حمایت از بازیگران بومی، و استفاده از استانداردهای ملی (مانند استاندارد الگوریتم پارسا) به‌عنوان یک «سدّ تعرفه‌ای غیرمستقیم» برای حفاظت از بازار داخلی در برابر استعمار داده‌ای است.

۳) کثرت‌گرایی ارزشی در برابر هژمونی جهانی: «الگوی سوم» با نقد جهان‌شمولی ارزش‌های لیبرال غربی نهفته در مدل‌های سیلیکون‌ولی، بر ساخت یک چارچوب ارزشی -بومی برای حکمرانی هوش مصنوعی تأکید دارد. این الگو راهبردی است ملی و ائتلاف‌ساز برای کشورهایی که خواهان حفظ استقلال فرهنگی و تمدنی در عصر AI هستند. ویژگی آن، ترکیب نظارت دولتی با توانمندسازی بومی و حاکمیت بر زیرساخت‌های فنی است؛ برخلاف مدل آمریکایی آزاد و مدل اروپایی صرفاً تنظیم‌گرانه. در این خصوص، بیانات راهبردی رهبر انقلاب درباره «حاکمیت ملی بر زیرساخت‌ها و مقابله با انحصار شرکت‌های فراملی» پشتوانه معرفتی و سیاسی این الگوی بومی را فراهم می‌کند. شهید آیت‌الله العظمی خامنه‌ای (خامنه‌ای، ۱۴۰۳/۰۶/۰۶) تصریح کردند که هوش مصنوعی «لایه‌های عمیق و متنوعی دارد» و صرفاً «بهره‌بردار بودن» امتیاز نیست، بلکه سلطه واقعی در گرو تسلط بر همان لایه‌های زیرساختی است. هشدار ایشان درباره احتمال شکل‌گیری نهادی شبیه «آژانس انرژی اتمی» برای هوش مصنوعی که با استانداردگذاری جهانی، ورود کشورهای مستقل را محدود کند، دقیقاً با ادبیات اقتصاد سیاسی فناوری هم‌پوشانی دارد؛ جایی که پژوهشگرانی چون زوبوف، فلوریدی و وست نشان می‌دهند که قدرت اصلی در مالکیت زیرساخت، انباشت داده، مدل‌های پایه و استانداردها در حکمرانی جهانی متمرکز می‌شود و نهادهای چندجانبه مانند ابتکارات سازمان ملل، اصول OECD و کنسرسیوم‌های سیلیکون‌ولی می‌توانند استانداردهای شرکت‌های آمریکایی را به «قواعد بین‌المللی» تبدیل کنند (OECD, 2019; Silicon Valley Leadership Group, 2024). بر این اساس، اگر ایران صرفاً مصرف‌کننده مدل‌های خارجی باقی بماند، با وضعیتی مواجه می‌شود که در ادبیات علمی «استعمار داده‌ای» و «هژمونی الگوریتمی» نام دارد؛ یعنی فرسایش حاکمیت شناختی، فرهنگی و حتی امنیتی به سود شرکت‌هایی که در مقام قانون‌گذار، مجری و قاضی نظم دیجیتال عمل می‌کنند (Scharre, 2023; Couldry & Mejias, 2019). الگوی سوم، پاسخی نظری و راهبردی به ناکارآمدی دو الگوی غالب (تنظیم‌گری سخت اروپایی و خودتنظیمی لیبرال آمریکایی) است و مستقیماً بر ملی‌سازی لایه‌های زیرساختی هوش مصنوعی تأکید دارد؛ ضرورتی که در بیانات رهبری به‌عنوان مبنای سیاست‌گذاری آینده و جهت‌دهنده راهبردهای ملی، بین‌المللی و پژوهشی مطرح شده است.

۱. راهبرد ملی: تأسیس زیرساخت بومی و «الگوریتم پارسا»

در سطح ملی، راهبرد «زیرساخت بومی و الگوریتم پارسا» بر پایه ضرورت حفظ هویت زبانی و فرهنگی در عصر هوش مصنوعی بنا شده است. این راهبرد، عدم واکنش در برابر پلتفرم‌های جهانی را

تهدیدی برای بقاء و امنیت ملی دانسته و توسعه مدل‌های زبانی بومی و زیرساخت‌های پردازشی را در تراز پروژه‌های راهبردی تعریف می‌کند. استاندارد «الگوریتم پارسا» نیز صرفاً دستور اخلاقی نیست، بلکه یک نظام فنی - ممیزی چهارسطحی برای پایش، اعتبارسنجی و حکمرانی داده و محتوا است.

- **لایه اول: لایه داده^۱ و «مجموعه داده پارسا»:** در «لایه داده»، معیار اصلی توسعه و استفاده از «پیکره ملی پارسا»^۲ است؛ پیکره‌ای عظیم و پالایش شده که حوزه‌های فرهنگ و ادب فارسی، میراث علمی و فلسفی، علوم اسلامی، و اسناد حقوقی و بالادستی کشور را شامل می‌شود تا داده‌های بنیادی مدل بازتاب‌دهنده ابعاد مختلف هویت ملی ایران باشد.

- **لایه دوم: لایه همسوسازی^۳ و «بازخورد پارسا»:** «لایه همسوسازی» با طرح «یادگیری تقویتی از بازخورد پارسا»^۴ روش متعارف RLHF را به چالش می‌کشد و بر استفاده از تیم‌های میان‌رشته‌ای متشکل از متخصصان اخلاق، حقوق، جامعه‌شناسی و فرهنگ برای رتبه‌بندی خروجی‌ها براساس اصولی چون عدالت و انصاف، کرامت انسانی و حریم خصوصی، راست‌گویی و امانت‌داری اطلاعاتی، خردورزی و دانش‌محوری، مسئولیت‌پذیری اجتماعی و وحدت ملی در عین کثرت فرهنگی تأکید می‌کند. این رویکرد نیازمند کارگروه‌های مشترک میان علوم انسانی و هوش مصنوعی برای فرمول‌بندی دقیق معیارها بوده و تلاش می‌کند همزمان به ادبیات فنی روز متصل بماند.

- **لایه سوم: لایه قانون اساسی^۵:** «لایه قانون اساسی» با الهام از رویکرد Constitutional AI (Bai & et al., 2022) پیشنهاد می‌کند یک «قانون اساسی الگوریتمی» مبتنی بر اصول قانون اساسی جمهوری اسلامی و ارزش‌های بنیادین فرهنگ ایرانی تدوین شود؛ مجموعه اصولی بالادستی که مدل در فرآیند آموزش می‌آموزد که هرگز از آن‌ها تخطی نکند. نمونه‌هایی از این اصول عبارتند از: اصل وحدت ملی (حفظ منافع ملی، تمامیت ارضی و وحدت اقوام)، اصل احترام به تنوع فرهنگی (پرهیز از تولید محتوای توهین‌آمیز یا تفرقه‌افکنانه نسبت به اقوام، زبان‌ها و مذاهب)، اصل کرامت انسانی، و اصل نظم عمومی و اخلاق حسنه که در ممیزی از طریق سناریوهای متخاصم^۶ آزمون می‌شوند.

- **لایه چهارم: لایه نظارت و پالایش^۷:** لایه نظارت و پالایش به‌عنوان خط دفاعی آخر، خروجی نهایی مدل را با پایگاه دانشی از «واژگان و مفاهیم ممنوعه» شامل محتوای صراحتاً غیرقانونی (مانند توهین به اقوام و ادیان رسمی، ترویج تروریسم، آموزش روش‌های مجرمانه، نقض حریم خصوصی) و محتوای

1. Data Layer
2. Parsa National Corpus (PNC)
3. Alignment Layer
4. Reinforcement Learning from PARSA Feedback (RLPF)
5. Constitutional Layer
6. Adversarial Testing
7. Supervisory and Guardrail Layer

هنجارشکن (مانند عادی سازی خشونت یا بی احترامی به باورهای عمومی) مقایسه کرده و در صورت تطابق، خروجی را مسدود یا اصلاح می کند. گواهی «الگوریتم پارسا» که پس از ممیزی چهار لایه اعطا می شود، هم شرط استفاده در نهادهای داخلی و هم مزیت رقابتی در بازارهای منطقه ای است. این الگو، در پاسخ به نقدهای احتمالی (چالش اقتصادی، انزوای فناورانه، و اتهام سانسور)، بر توانمندسازی بومی از طریق همسوسازی مدل های متن باز (نه رقابت پرهزینه) و تعریف قواعد تعامل راهبردی برای حفظ حاکمیت داده تأکید دارد؛ رویکردی که نه به معنای انزوا، بلکه مشابه تثبیت مرجعیت تنظیم گری اتحادیه اروپا از طریق قانون «مقررات عمومی حفاظت از داده»^۱ است. از آنجا که هیچ الگوریتمی خنثی نیست، این استاندارد با جایگزینی ارزش های پنهان غربی با یک چارچوب شفاف ملی، حاکمیت فناورانه دفاعی را در برابر استعمار الگوریتمی محقق می سازد، و نه ملی گرایی تهاجمی.

۲. راهبرد بین المللی: دیپلماسی سایبری و ائتلاف سازی واقع بینانه

در برابر هژمونی دیجیتال شرکت های چندملیتی مورد حمایت دولت آمریکا، اقدام انفرادی برای کشورهایی با سیاست خارجی مستقل پرهزینه و کم اثر است. بنابراین، استقلال فناورانه نیازمند راهبردی بین المللی و دو مرحله ای است که گذار از وابستگی به خودمختاری را ممکن سازد.

مرحله اول: ائتلاف سازی تاکتیکی برای شکستن محاصره فناورانه

در گام نخست، هدف اصلی باید ایجاد یک جبهه متحد برای شکستن انحصار زیرساختی و استانداردگذاری مدل آمریکایی باشد. این مهم از دو طریق قابل پیگیری است:

- **تشکیل کنسرسیوم حاکمیت داده:** پیشگامی در ایجاد کنسرسیومی از کشورهای همسو (مانند اعضای بریکس و سازمان شانگهای) برای تدوین «پروتکل های مشترک حکمرانی داده» و اشتراک منابع محاسباتی و مدل های پایه. این همکاری به معنای پذیرش کامل مدل حکمرانی شرکا نیست، بلکه اقدامی عمل گرایانه برای شکستن انحصار آمریکایی و فراهم کردن زمینه ظهور مدل های جایگزین است.

- **دیپلماسی فعال در مجامع بین المللی:** این ائتلاف باید به طور هماهنگ در نهادهایی مانند ITU و WSIS گفتمان «استعمار داده ای» و مخاطرات امنیتی و فرهنگی انحصار پلتفرم های آمریکایی را مطرح کرده و خواستار پاسخگویی حقوقی و بین المللی این شرکت ها شود.

مرحله دوم: گذار به خودمختاری راهبردی و تحقق الگوی سوم

ائتلاف بین المللی هدف نهایی نیست، بلکه تاکتیکی برای گذار است. این همکاری فضای تنفس راهبردی و منابع اولیه لازم برای مرحله بعد را فراهم می کند: دستیابی به استقلال در لایه های زیرساختی هوش مصنوعی. در این مرحله، توان ملی باید بر راهبردهای بومی، از جمله توسعه مدل های زبانی ملی و اجرای استاندارد «الگوریتم پارسا»، متمرکز شود. هدف نهایی آن است که کشور نه صرفاً مصرف کننده

یا عضو ائتلاف، بلکه قطبی مستقل و صاحب سبک در حکمرانی هوش مصنوعی بر پایه «الگوی سوم» باشد.

۳. راهبرد پژوهشی: گذار از مصرف‌کننده به نظریه‌پرداز

جامعه دانشگاهی باید از مرحله «ترجمه» مباحث اخلاق هوش مصنوعی به مرحله «تولید نظریه» عبور کند. پژوهش‌های آتی باید بر محورهای زیر متمرکز شوند:

- طراحی مدل‌های ریاضی برای فرمول‌بندی ارزش‌های بومی در فرآیندهای RLHF.
- بررسی حقوقی امکان اقامه دعوی علیه پلتفرم‌های هوش مصنوعی خارجی به دلیل نقض حاکمیت فرهنگی و قوانین محلی.

این رویکرد چندلایه، یک جایگزین بومی برای مدل‌های همسوسازی توسعه‌یافته در سیلیکون‌ولی ارائه می‌دهد. جدول (۲)، تفاوت‌های کلیدی این دو رویکرد را در لایه‌های فنی و ارزشی به صورت مقایسه‌ای آورده است.

جدول ۲- مقایسه رویکردهای فنی همسوسازی: مدل سیلیکون‌ولی در برابر الگوریتم پارسا

لایه فنی	رویکرد مسلط سیلیکون‌ولی (مبتنی بر RLHF)	رویکرد پیشنهادی «الگوریتم پارسا» (مبتنی بر RLPF)
لایه داده (Data Layer)	استفاده از داده‌های عظیم وب (عمدتاً انگلیسی‌زبان و غربی) با پالایش حداقلی برای حذف محتوای سمی.	استفاده از «بیکره ملی پارسا» (PNC)؛ مجموعه‌داده پالایش‌شده شامل فرهنگ، ادب، علوم اسلامی و اسناد حقوقی ایران.
لایه همسوسازی (Alignment Layer)	یادگیری تقویتی از بازخورد انسانی (RLHF)؛ استفاده از برجسب‌زن‌های انسانی (اغلب پیمانکاری) برای رتبه‌بندی خروجی‌ها.	یادگیری تقویتی از بازخورد پارسا (RLPF)؛ استفاده از تیم‌های میان‌رشته‌ای (اخلاق، حقوق، جامعه‌شناسی) برای رتبه‌بندی.
مبنای ارزش‌گذاری	ارزش‌های لیبرال غربی (اغلب به صورت ضمنی و کدگذاری‌شده در دستورالعمل‌ها) مانند ایمنی، بی‌ضرری و عدم توهین.	اصول بنیادین بومی (به صورت صریح)؛ مانند عدالت، کرامت انسانی، راست‌گویی، وحدت ملی و خردورزی.
لایه قانون اساسی (Constitutional Layer)	استفاده از اصول کلی و انتزاعی نوشته‌شده توسط شرکت برای هدایت مدل (مانند مدل Anthropic).	اتصال به اسناد بالادستی و قوانین بنیادین کشور به‌عنوان قانون اساسی مدل برای تصمیم‌گیری در موارد پیچیده.
هدف نهایی	ساخت مدل‌های «بی‌ضرر» و «مفید» با هدف پذیرش جهانی و کاهش ریسک‌های حقوقی و اعتباری.	ساخت مدل‌های «پارسا» که بازتاب‌دهنده هویت ملی، مروج ارزش‌های بومی و ابزاری برای تحقق حاکمیت فناورانه است.

همان‌گونه که جدول (۲) نشان می‌دهد، تفاوت دو رویکرد صرفاً فنی نیست، بلکه بیانگر دو منطق متفاوت حکمرانی است. مدل مسلط سیلیکون‌ولی بر بهینه‌سازی ایمنی و پذیرش جهانی در چارچوب

ارزش های ضمنی لیبرال تکیه دارد، در حالی که رویکرد «الگوریتم پارسا» می کوشد همسوسازی مدل را آگاهانه با چارچوب های ارزشی، حقوقی و فرهنگی بومی پیوند داده و آن را در خدمت حاکمیت فناورانه قرار دهد.

۶. نتیجه گیری

این پژوهش با استفاده از اقتصاد سیاسی و تحلیل گفتمان انتقادی، مفهوم «حکمرانی درونی» در هوش مصنوعی را بازخوانی کرد. یافته ها نشان می دهد که گفتمان ایمنی شرکت های بزرگ، صرفاً فنی نیست؛ بلکه سازوکاری ساختاری برای تثبیت قدرت اقتصادی، تسخیر مقررات و تحمیل هژمونی هنجاری است. معماری این حکمرانی از سه طریق عمل می کند: ایجاد موانع ورود برای رقبا، تسخیر نرم مقررات و همسوسازی مدل ها برای بازتولید ارزش های مسلط. از منظر اندیشه سیاسی اسلام، «حاکمیت فناورانه» ضرورتی راهبردی برای صیانت از استقلال و عدالت است. سلطه زیرساخت های شناختی بیگانه با اصل قرآنی «نفی سبیل» (نساء، ۱۴۱) (به مثابه قاعده ای ناظر بر منع هرگونه برتری و سیطره پایدار غیرمسلمانان بر مقدرات جامعه اسلامی) در تعارض است. افزون بر آن، با استناد به قاعده «حفظ نظام» (که مقدم بر بسیاری از احکام فرعی است) و مفهوم «ولایت در امور عامه»، دخالت دولت در تنظیم معماری های فنی که جایگزین قانون شده اند، مشروع و ضروری است. بر این اساس، استاندارد «الگوریتم پارسا» ترجمان عملی این مبانی است تا حاکمیت را از موضع منفعل و مصرف کننده، به کنشگری فعال در سطح زیرساختی ارتقاء دهد. این پژوهش می تواند مبنایی برای توسعه چارچوب های بومی حکمرانی در کشورهای خواهان استقلال فناورانه باشد؛ هرچند امکان سنجی اقتصادی و فنی این الگوها نیازمند مطالعات آتی است.

۷. محدودیت های پژوهش و پیشنهادهایی برای مطالعات آتی

این پژوهش با وجود تلاش برای ارائه تحلیلی جامع از گفتمان حکمرانی و ایمنی در صنعت هوش مصنوعی، با چند محدودیت روبرو است. نخست، دامنه مطالعه به اسناد سیاستی و فنی سه شرکت Open AI، Google DeepMind و Anthropic محدود شد؛ بنابراین دیدگاه شرکت های چینی، نهادهای دولتی و بازیگران متن باز در این تحلیل غایب است. دوم، تکیه بر اسناد رسمی این شرکت ها به معنای اتکاء به روایت نهادی آن هاست و لزوماً تمام ابعاد عملی حکمرانی درونی را بازتاب نمی دهد. سوم، تحلیل گفتمان انتقادی، ماهیتی تفسیری دارد و نتایج تا حدی به چارچوب نظری و خوانش پژوهشگر وابسته است. بر این اساس، پژوهش های آینده می توانند با گسترش دامنه داده ها - از جمله تحلیل اسناد تنظیم گری، مصاحبه با متخصصان و بررسی تجربی سازوکارهای همسوسازی و ایمنی - تصویر کامل تری از حکمرانی هوش مصنوعی ارائه کنند. همچنین مطالعات مقایسه ای میان الگوهای حکمرانی در مناطق

مختلف جهان می‌تواند به فهم تنوع رویکردها و پیامدهای ژئوپلیتیک آنها کمک کند. در نهایت، ارزیابی فنی و تجربی چارچوب‌های پیشنهادی برای حکمرانی مسئولانه، از جمله مدل‌های بومی یا منطقه‌ای، مسیر مهمی برای تحقیقات آتی خواهد بود.

منابع

قرآن کریم.

- خامنه‌ای، سید علی (۱۴۰۳/۰۶/۰۶). بیانات در دیدار با رئیس‌جمهور و اعضای هیئت دولت چهاردهم. قابل دسترس
<https://farsi.khamenei.ir/speech-content?id=57426> در:
- شکوهمیان‌راد، محمدعلی (۱۴۰۴). ضرورت‌های حکمرانی فضای سایبر از منظر امام خامنه‌ای. *مطالعات حکمرانی
 اسلامی*، ۱(۲)، ص ۲۳۹-۲۶۶. <https://doi.org/10.22081/jislamicgo.2025.71981.1028>
- Bai, Y., Kadavath, S., Kundu, S., Askill, A., Kernion, J., Jones, A., ... Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback. *arXiv*.
<https://doi.org/10.48550/arXiv.2212.08073>
- Bietti, E. (2020). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20)* (pp. 210–219). Association for Computing Machinery.
<https://doi.org/10.1145/3351095.3372860>
- Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R. & Bao, M. (2023). The values encoded in machine learning research. In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)* (pp. 1735–1746). Association for Computing Machinery. <https://doi.org/10.1145/3593013.3594087>
- Christiano, P., Leike, J., Brown, T., Martic, M., Legg, S. & Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in Neural Information Processing Systems*, 30. URL=
https://proceedings.neurips.cc/paper_files/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html
- Couldry, N. & Mejias, U.A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (COM/2021/206 final)*.
 URL= <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- European Commission. (2024). *Artificial Intelligence Act (Regulation (EU) 2024/1689): Harmonised rules on artificial intelligence*.
 URL= <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- Fairclough, N. (2010). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315834368>
- Federal Register. (2023). *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence* (No. 2023-24283). URL=
<https://www.federalregister.gov/documents/2023/11/01/2023-24283/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>
- Floridi, L. (2013). *The ethics of information*. Oxford University Press.

- Google. (2023, July 18). Google's AI red team: The ethical hackers making AI safer. *Google Blog*. URL= <https://blog.google/technology/safety-security/googles-ai-red-team-the-ethical-hackers-making-ai-safer/>
- Google. (2024). *Generative AI prohibited use policy*. URL= <https://policies.google.com/terms/generative-ai/use-policy>
- Guha, N., Ho, D.E. & Kogan, I. (2024). AI regulation has its own alignment problem: The technical and institutional feasibility of disclosure, registration, licensing, and auditing. *The George Washington Law Review*, 92(6), p. 1473-1557. https://dho.stanford.edu/wp-content/uploads/AI_Regulation.pdf
- Inserra, D. (2024). *Artificial intelligence regulation threatens free expression* (Briefing Paper No. 178). Cato Institute. URL= <https://www.cato.org/briefing-paper/artificial-intelligence-regulation-threatens-free-expression>
- IT for Change. (2023). *Reframing AI governance through a political economy lens*. WACC Global. URL= <https://waccglobal.org/reframing-ai-governance-through-a-political-economy-lens/>
- Lessig, L. (1999). *Code and other laws of cyberspace*. Basic Books.
- Mäntymäki, M., Heikkilä, J., Merikivi, J. & Dao, T. (2022). The hourglass model of organizational AI governance. *arXiv*.
- Microsoft. (2025). *Microsoft AI red team*. URL= <https://learn.microsoft.com/en-us/security/ai-red-team/>
- Mosco, V. (2009). *The political economy of communication* (2nd ed.). SAGE.
- National Security Commission on Artificial Intelligence. (2021). *Final report*. URL= <https://reports.nscai.gov/final-report/table-of-contents/>
- Noble, S.U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- OECD. (2019). *OECD principles on artificial intelligence*. OECD. URL= <https://oecd.ai/en/ai-principles>
- Open AI. (2023). *GPT-4 system card*. <https://cdn.openai.com/papers/gpt-4-system-card.pdf>
- Open AI. (2024). *Terms of use*. Open AI. <https://openai.com/policies/row-terms-of-use>
- Open AI. (2025). *Usage policies*. <https://openai.com/policies/usage-policies>
- Radin, M. J. (2013). *Boilerplate: The fine print, vanishing rights, and the rule of law*. Princeton University Press.
- Scharre, P. (2023). *Four battlegrounds: Power in the age of artificial intelligence*. W. W. Norton & Company.
- Shad, B. (2025). Algorithmic hegemony: AI, political elites, and the reinvention of electoral influence. *Journal of Sociocybernetics*, 20(1), p. 1-31. <https://papiro.unizar.es/ojs/index.php/rc51-jos/article/view/11599>
- Silicon Valley Leadership Group. (2024). *About SVLG*. URL= <https://www.svl.org>
- Tan, Z.-X., Carroll, M., Franklin, M. & Ashton, H. (2024). Beyond preferences in AI alignment.

- arXiv*. URL= <https://arxiv.org/abs/2408.16984>
- The White House. (2023). *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence*. URL= <https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>
- U.S. AI Regulation Study. (2024). *The U.S. approach to AI regulation: Federal laws, policies, and governance*. SSRN. URL= https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4931927
- Wagner, B. (2018). Ethics as an escape from regulation: From “ethics-washing” to ethics-shopping? In: E. Bayamlioğlu, I. Baraliuc, L. Janssens, & M. Hildebrandt (Eds.), *Being profiled: Cogitas ergo sum. 10 years of “Profiling the European Citizen”* (pp. 84–88). Amsterdam University Press. <https://doi.org/10.25969/mediarep/13281>
- Wei, K., Ezell, C., Gabrieli, N. & Deshpande, C. (2024). How do AI companies “fine-tune” policy? Examining regulatory capture in AI governance. *arXiv*. <https://doi.org/10.48550/arXiv.2410.13042>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.