کلام آغازین

دوره ۳۵ | شماره ۳ | پیاپی ۹۹ | مهر ۱۴۰۴
تاریخ انتشار برخط: ۱۴۰۵/۰۶/۱۳
صفحات: ۱-۸
شاپای چاپی: ۱۰۲۷-۲۶۹۰
شاپای الکترونیکی: ۲۷۸۳-۴۵۱۴

کلام آغازین

کژساخته‌های هوش مصنوعی در سیاست‌گذاری و توسعهٔ فنّاوری: از خطای تحلیلی تا خطای تصمیم

۱. رضا حافظی
۲. فرخنده ملکی‌فر ◇

چکیده

گسترش استفاده از هوش مصنوعی در پژوهش، آینده‌نگاری، مرور ادبیات، تحلیل روندها و سیاست‌گذاری فنّاوری، وعدهٔ افزایش سرعت، دقت و جامعیت تصمیم‌سازی را همراه داشته است. بااین‌حال، خطر اصلی هوش مصنوعی در سیاست‌گذاری فقط تولید پاسخ نادرست یا دادهٔ جعلی نیست. خطر عمیق‌تر آن است که خروجی‌های خطادار را در قالبی منسجم، قاطع، مستند و قابل استفاده برای تصمیم‌های نهادی عرضه کند. این مقاله مفهوم «کژساخته‌های هوش مصنوعی» را برای توصیف این نوع خطاهای معتبرنما به کار می‌گیرد؛ خطاهایی که ممکن است از معماری الگو، داده‌های ناقص یا سوگیرانه، قطعیت در روایت‌ها، فشار نهادی برای تصمیم سریع، یا ترجمهٔ شتاب‌زدهٔ خروجی‌های تحلیلی به توصیهٔ سیاستی پدید آیند. استدلال اصلی متن این است که در سیاست‌گذاری و توسعهٔ فنّاوری، کژساخته زمانی خطرناک می‌شود که از سطح خطای تحلیلی عبور کند و به مبنای تصمیم، بودجه، اولویت‌گذاری، نقشهٔ راه یا روایت رسمی تبدیل شود. بر این اساس، مسئولیت کژساخته‌ها نیز به‌صورت فردی و در ساختاری ساده نیست، بلکه در زنجیره‌ای از طراحان الگو، سکوها، تحلیلگران، مشاوران، سازمان‌های سفارش‌دهنده و نهادهای حکمرانی توزیع می‌شود. مقاله در پایان بر ضرورت شکل‌گیری تحلیل سیاستی حساس به کژساخته تأکید می‌کند؛ رویکردی که نه به حذف هوش مصنوعی می‌انجامد و نه به

کلیدواژه‌ها

هوش مصنوعی، کژساخته، سیاست‌گذاری فنّاوری، خطای تصمیم، آینده‌نگاری.

۱. استادیارآینده پژوهی، مؤسسه تحقیقات سیاست علمی کشور، تهران، ایران (دبیر موضوعی)
hafezi@nrisp.ac.ir
ORCID: 0000-0003-0070-3737

۲. ◇ استادیارمدیریت فناوری، مؤسسه تحقیقات سیاست علمی کشور، تهران، ایران (دبیر موضوعی ویدیدآوررابط)
malekifar@nrisp.ac.ir
ORCID: 0009-0009-0674-1401

استناد به این مقاله: حافظی، ر.، و ملکی‌فر، ف. (۱۴۰۴). کژساخته‌های هوش مصنوعی در سیاست‌گذاری و توسعهٔ فنّاوری: از خطای تحلیلی تا خطای تصمیم. *رهیافت، ۳۵ (۳)*، صص. ۸-۱.
DOI: 10.22034/RAHYAFT.2025.14174

ناشر: مؤسسه تحقیقات سیاست علمی کشور

اعتماد خام به آن، بلکه استفاده از آن را با تشخیص خطای دید، حفظ عدم قطعیت، راستی‌آزمایی زمینه‌ای و مسئولیت‌پذیری نهادی همراه می‌سازد.

مقدمه

هوش مصنوعی به‌سرعت وارد حوزه‌هایی شده است که پیش‌تر بر قضاوت انسانی، تجربهٔ کارشناسی، مرور تدریجی شواهد و گفت‌وگوی نهادی متکی بودند (Jarrahi, 2018; Shrestha, Ben-Menahem, & von Krogh, 2019). امروز از ابزارهای هوش مصنوعی در مراحل مختلف چرخهٔ سیاست‌گذاری، به‌ویژه در تعیین دستور کار (مثلاً در آماده‌سازی و تحلیل داده‌ها، خلاصه‌سازی داده‌ها، و محک‌زنی[1] بین‌المللی) و تدوین سیاست‌ها (مثلاً در پیش‌بینی و شبیه‌سازی سیاست‌ها، تحلیل اسناد حقوقی، و توسعهٔ راهبردها) استفاده می‌شود (Lnenicka et al., 2026). این تحول، در ظاهر، پاسخی مناسب به یکی از مشکلات مزمن سیاست‌گذاری فنّاوری است: انباشت داده، کمبود زمان و نیاز به تصمیم‌گیری در شرایط عدم قطعیت. اما همین مزیت ظاهری، نقطهٔ آغاز مسئله است. سیاست‌گذاری فنّاوری معمولاً به خروجی‌هایی علاقه دارد که خلاصه، روشن، قابل دفاع، قابل ارائه و قابل تبدیل به تصمیم باشند. هوش مصنوعی (مولد) نیز دقیقاً چنین خروجی‌هایی تولید می‌کند: متن مرتب، دسته‌بندی روشن، جمع‌بندی سریع، توصیهٔ سیاستی و گاه شواهدی که ظاهراً معتبر به نظر می‌رسند. خطر از همین‌جا آغاز می‌شود؛ زمانی که هوش مصنوعی خطای احتمالی را در قالبی تولید می‌کند که برای نظام تصمیم‌گیری جذاب، قابل استفاده و حتی قانع‌کننده است (Goddard, Roudsari, & Wyatt, 2012; Huang et al., 2025). این موضوع وقتی اهمیت بیشتری پیدا می‌کند که طبق مطالعات تجربی، حتی افراد باتجربه هم نمی‌توانند تمام خطاهای هوش مصنوعی در تصمیم‌گیری را تشخیص دهند (Janssen, Hartog, Matheus, Yi Ding, & Kuk, 2022).

افزون‌براین، شواهد تجربی نشان از این دارد که افراد در صورتی که توصیه‌های هوش مصنوعی همسو با قضاوت‌های پیشین آن‌ها باشد، راحت‌تر به آن اعتماد می‌کنند (Selten, Robeer, & Grimmelikhuijsen, 2023)؛ این دغدغه که می‌تواند به ایجاد چرخه‌های معیوب در سیاست‌گذاری با تکیه بر شواهد مبتنی بر هوش مصنوعی منجر شود، با عبارت «شواهد سیاست‌محور»[2] در مقابل «سیاست شواهدمحور» شناخته می‌شود (Sharman & Holmes, 2010). در این صورت، خطاها در توصیه‌های هوش مصنوعی، اگر همسو و تقویت‌کنندهٔ قضاوت‌ها یا سیاست‌های پیشین باشند، کمتر نادیده گرفته می‌شوند یا حتی ممکن است این توصیه‌ها صرفاً برای توجیه سیاست‌ها و به شکل نمادین به کار گرفته شوند (Lnenicka et al., 2026).

در ادبیات موضوع متأخر، بحث‌های مهمی دربارهٔ هوش مصنوعی مسئولانه، شفافیت، سوگیری، توضیح‌پذیری و اعتمادپذیری شکل گرفته است (Cheong, 2024; Mikalef, Conboy, Lundström, & Popovič, 2022; Papagiannidis, Mikalef, & Conboy, 2025). بااین‌حال، بخش مهمی از این بحث‌ها هنوز بر این فرض استوار است که می‌توان با اصول اخلاقی، کنترل‌های فنی یا دستورالعمل‌های استفاده، خطرات و ریسک‌های هوش مصنوعی را مهار کرد. برخی پژوهش‌ها نشان داده‌اند که چهارچوب‌های رایج مسئولیت‌پذیری، هوش مصنوعی را غالباً ابزاری قابل کنترل، ایستا، توضیح‌پذیر و همگن در نظر می‌گیرند، درحالی‌که سامانه‌های هوش مصنوعی معاصر پویا، خودتنظیم‌شونده، تا حدی تبیین‌ناپذیر، سازگارشونده و ناهمگن‌اند (de Bruijn, Warnier, & Janssen, 2022; Ibitoye, Nkwo, & Orji, 2025; Mikalef, Benlian, Conboy, & Tarafdar, 2025). همین شکاف، مسئولیت‌پذیری در برابر خروجی‌های هوش مصنوعی را پیچیده‌تر می‌کند.

این مقاله از مفهوم «کژساخته» برای فهم این پیچیدگی استفاده می‌کند. کژساخته در معنای عمومی خود به چیزی اشاره دارد که دیده می‌شود، اما الزاماً بازنمایی دقیق واقعیت نیست. در تصویربرداری پزشکی، کژساخته ممکن است لکه، سایه یا ساختاری باشد که شبیه نشانهٔ بیماری به نظر می‌رسد، اما محصول حرکت بیمار، نویز دستگاه یا فرایند پردازش تصویر است (Barrett & Keat, 2004; Zaitsev, Maclaren, & Herbst, 2015). به‌طور مشابه، در سیاست‌گذاری فنّاوری نیز کژساخته‌های هوش مصنوعی می‌توانند شبیه تحلیل، شواهد یا توصیهٔ سیاستی ظاهر شوند، درحالی‌که حاصل سوگیری داده، معماری مدل، قطعیت کاذب در روایت‌ها و تحلیل‌های جدا از بافتار باشند.

بنابراین، پرداختن به این پرسش ضرورت دارد که وقتی هوش مصنوعی در فرایند شناخت، تحلیل و تصمیم‌سازی وارد می‌شود، چه نوع خطاهای معتبرنمایی تولید می‌کند و چه کسی مسئول تبدیل این خطاها به تصمیم است؟

از خطای فنی تا کژساختهٔ سیاستی

پیش از ورود به بحث، لازم است مرز میان خطای فنی و کژساختهٔ سیاستی روشن‌تر شود. هر خطای هوش مصنوعی لزوماً کژساخته نیست. خطای فنی زمانی رخ می‌دهد که سامانه، خروجی نادرست، ناقص، سوگیرانه یا بی‌پشتوانه تولید می‌کند؛ مانند ارجاع جعلی، ترجمهٔ غلط، طبقه‌بندی نادرست، سوگیری در داده یا پاسخ ساخته‌شده. این

1. benchmarking
2. policy-based evidence

خطاها مهم‌اند، اما تا زمانی که در همان سطح خروجی باقی بمانند، هنوز به معنای دقیق کلمه کژساختهٔ سیاستی نیستند.

کژساختهٔ سیاستی زمانی شکل می‌گیرد که همین خروجی، یا حتی خروجی‌ای که در ظاهر غلط به نظر نمی‌رسد، در فرایند تحلیل و تصمیم‌سازی به شکل چیزی معتبرتر از واقعیت خود بازنمایی شود؛ یعنی به‌جای آنکه صرفاً یک پاسخ، خلاصه یا تحلیل مقدماتی باشد، به شکل شواهد، روند، اجماع کارشناسی، بلوغ فنّاوری، اولویت راهبردی یا توصیهٔ اجرایی وارد چرخهٔ سیاست‌گذاری شود. در این معنا، کژساخته فقط خطای الگو نیست؛ خطایی است که در مسیر تبدیل خروجی الگو به معنا، روایت و تصمیم ساخته می‌شود. مثلاً، سوگیری داده به‌خودی‌خود می‌تواند یک خطای فنی یا داده‌ای باشد؛ اما اگر همین سوگیری باعث شود یک فنّاوری بیش از حد بالغ، یک مسئله بیش از حد قطعی، یا یک راه‌حل بیش از حد قابل اجرا نشان داده شود، با کژساختهٔ سیاستی روبه‌روییم. بنابراین، تفاوت اصلی در منشأ اولیهٔ خطا نیست، بلکه در جایگاه و اثر آن است. خطای فنی خروجی را مخدوش می‌کند، اما کژساختهٔ سیاستی فهم مسئله و جهت تصمیم را تغییر می‌دهد.

بسیاری از بحث‌های عمومی دربارهٔ خطای هوش مصنوعی بر نمونه‌هایی مثل ارجاع جعلی، پاسخ نادرست، سوگیری زبانی، ترجمهٔ غلط، یا تولید اطلاعات بدون پشتوانه تمرکز دارند (Gallegos et al., 2024; Huang et al., 2025; Resnik & Hosseini, 2026). گرچه معمولاً این خطاها به‌راحتی قابل مشاهده‌اند، مسئلهٔ دشوارتر زمانی رخ می‌دهد که خروجی هوش مصنوعی در نگاه اول، دقیق، متوازن، حرفه‌ای و قانع‌کننده به نظر می‌رسد. در سیاست‌گذاری فنّاوری، زمانی که تصمیم‌گیران در پی خلاصه، چهارچوب، اولویت، سناریو، نقشهٔ راه یا توصیه هستند، این خطا می‌تواند به نتایجی فاجعه‌بار بینجامد. خروجی هوش مصنوعی به‌رغم سرعت و شفافیت زبانی ممکن است از نظر معرفتی ضعیف باشد. یا ممکن است به‌رغم اینکه در نگاه اول اجرایی به نظر می‌رسد، با ظرفیت نهادی سازگار نباشد. ممکن است یک اولویت فنّاوری، بر داده‌های فراوان مبتنی باشد، اما داده‌ها از ابتدا نمایندهٔ واقعیت فنّاوری نبوده باشند. در این معنا، کژساخته زمانی شکل می‌گیرد که خروجی مدل، خطا را به شکل شواهد، روند، بلوغ فنّاوری، اجماع کارشناسی یا توصیهٔ سیاستی بازنمایی کند. در این حالت، مسئله از سطح فنی عبور می‌کند و وارد سطح سیاستی می‌شود.

در پروژه‌های آینده‌نگاری فنّاوری نیز چنین وضعیتی آشناست. شاخص‌های کتاب‌سنجی، داده‌های پتنت، تحلیل روند، خوشه‌بندی موضوعی و مدل‌های یادگیری ماشین می‌توانند سیگنال‌هایی تولید کنند که در ظاهر نشان‌دهندهٔ ظهور یا بلوغ یک فنّاوری‌اند. اما این سیگنال‌ها ممکن است حاصل سوگیری پایگاه داده، بزرگ‌نمایی اخبار رسانه‌ای، حلقه‌های ارجاع، کلیدواژه‌های تکرارشده، رفتار بازیگران یا حساسیت مدل به پارامترها باشند. در چهارچوب «کژساخته‌های پیش‌بینی فنّاوری»، این نوع خطاها بازنمایی‌های ساخت‌یافته‌ای تعریف می‌شوند که شبیه سیگنال معتبر فنّاوری ظاهر می‌شوند، اما پشتوانهٔ ضعیفی در پویایی واقعی فنّاوری دارند. ورود هوش مصنوعی مولد این مسئله را تشدید می‌کند، زیرا هوش مصنوعی فقط داده را تحلیل نمی‌کند؛ داده را روایت‌پذیر می‌کند؛ یعنی پراکندگی، ابهام و عدم قطعیت را به متن منسجم تبدیل می‌کند. این تبدیل، اگر با مراقبت همراه نباشد، می‌تواند خطا را از سطح داده به سطح معنا منتقل کند.

پنج کژساختهٔ اصلی در سیاست‌گذاری و توسعهٔ فنّاوری

برای فهم دقیق‌تر مسئله، نویسندگان در این مقاله پنج نوع کژساخته معرفی می‌کنند. این دسته‌بندی نهایی و کامل نیست، اما برای سیاست‌گذاری فنّاوری کاربردی و نقطهٔ آغازین مناسبی است.

نخست، کژساختهٔ قطعیت است. این کژساخته زمانی پدید می‌آید که هوش مصنوعی عدم قطعیت را به زبان قطعی تبدیل می‌کند. بسیاری از موضوعات سیاست‌گذاری فنّاوری، از آیندهٔ هوش مصنوعی و انرژی گرفته تا آموزش عالی، سلامت دیجیتال و فنّاوری‌های نوظهور، با عدم قطعیت عمیق همراه‌اند. بااین‌حال، خروجی هوش مصنوعی معمولاً تمایل دارد پاسخ را مرتب، متوازن و نتیجه‌دار ارائه کند. در چنین حالتی، مسئله این نیست که مدل قطعاً اشتباه می‌گوید. مسئله این است که ندانستن را به‌صورت دانستن تصویر می‌کند و عدم قطعیت را ساده‌سازی می‌کند و پدیده‌ای قطعی نشان می‌دهد. همچنان‌که یکی از مطالعات تجربی اخیر (Leonardi & Leavell, 2026) نشان می‌دهد که بازنمایی‌های بسیار جزئی و واقع‌گرایانهٔ هوش مصنوعی در برنامه‌ریزی شهری، توهم «قطعیت مصنوعی» ایجاد می‌کند؛ یعنی وضعیتی که در آن کاربران، پیامدهای پیچیده و نامعلوم آینده را قطعاً قابل دانستن فرض می‌کنند.

دوم، کژساختهٔ اجماع است. سیاست‌گذاری فنّاوری غالباً با تعارض دیدگاه‌ها، منافع و چهارچوب‌های تفسیری همراه است. دانشگاه، صنعت، دولت، جامعهٔ مدنی، متخصصان فنی و کاربران نهایی، یک فنّاوری را به یک شکل نمی‌فهمند، اما هوش مصنوعی ممکن است این واگرایی‌ها را به یک روایت سازگار و ساده تبدیل کند. خروجی نهایی شاید متوازن به نظر برسد، اما اختلاف‌های واقعی را پنهان کند. برای برنامه‌های آینده‌نگاری، این خطر بسیار جدی است، زیرا آینده‌نگاری قرار نیست فقط میانگین دیدگاه‌ها را منعکس کند، بلکه باید ساختار واگرایی را نیز آشکار سازد. مطالعه‌ای که ۲۴۰۰ توصیهٔ راهبردی تولیدشده از چهار مدل زبانی بزرگ را با قضاوت ۲۱۸ مدیر

انسانی مقایسه کرده است (Stoeber et al., 2026)، این یافته را تأیید می‌کند که هوش مصنوعی در نقش «تقویت‌کنندهٔ نهادی» در جهت تثبیت هنجارها و کاهش تنوع در برنامه‌ریزی راهبردی عمل می‌کند. پژوهشگران این سوگیری را «خوش‌بینی مصنوعی» نامیده‌اند: تمایل افراطی هوش مصنوعی به همنوایی و هماهنگی با اجماع نهادی. این یافته مستقیماً نشان می‌دهد که کژساختهٔ اجماع ریشه در معماری مدل‌ها دارد، نه‌فقط در نحوهٔ استفاده از آن‌ها.

سوم، کژساختهٔ بلوغ فنّاوری است. در بسیاری از گزارش‌های سیاستی، تعداد مقالات، تعداد پتنت‌ها، میزان سرمایه‌گذاری، شدت رسانه‌ای شدن یا تکرار یک کلیدواژه، نشانهٔ رشد فنّاوری تلقی می‌شود. هوش مصنوعی می‌تواند این نشانه‌ها را سریع‌تر جمع‌آوری و حتی تفسیر کند. اما اگر دادهٔ اولیه متورم یا سوگیرانه باشد، خروجی نیز بلوغی را نشان می‌دهد که شاید در واقعیت وجود ندارد. یک فنّاوری ممکن است در ادبیات و رسانه پررنگ باشد، اما از نظر زیرساخت، بازار، تنظیم‌گری، پذیرش اجتماعی یا ظرفیت عملیاتی کردن هنوز شکننده باشد.

چهارم، کژساختهٔ ارتباط زمینه‌ای است. هوش مصنوعی می‌تواند مفاهیم جهانی را به زبان محلی بازنویسی کند، اما بازنویسی، معادل بافتارمند کردن نیست. ممکن است توصیه‌هایی دربارهٔ حکمرانی داده، سیاست صنعتی، آموزش هوشمند یا تنظیم‌گری هوش مصنوعی تولید شود که در ظاهر برای یک کشور یا سازمان خاص مناسب است، اما با ظرفیت اجرایی، ساختار حکمرانی، منابع مالی، فرهنگ نهادی یا تعارض‌های ذی‌نفعان آن سازگار نباشد. در این حالت، خروجی مرتبط به نظر می‌رسد، اما واقعاً بافتارمند و در نتیجه قابل استفاده نیست.

پنجم، کژساختهٔ اقدام‌پذیری است. این شاید خطرناک‌ترین نوع کژساخته‌ها باشد. هوش مصنوعی معمولاً می‌تواند برای هر مسئله، مجموعه‌ای از توصیه‌ها تولید کند: تدوین چهارچوب، ایجاد نهاد هماهنگ‌کننده، توسعهٔ زیرساخت، آموزش نیروی انسانی، پایش مستمر، حمایت از نوآوری، و مانند آن. این توصیه‌ها غلط نیستند، اما معمولاً بیش از حد عمومی‌اند. مشکل زمانی رخ می‌دهد که همین توصیه‌های عمومی برنامهٔ سیاستی تلقی شوند. اقدام‌پذیری واقعی مستلزم شناخت ابزار سیاستی، قدرت اجرایی، هزینه، زمان، تعارض منافع، ظرفیت نهادی و پیامدهای ناخواسته است. بدون این‌ها، توصیهٔ سیاستی فقط شبیه سیاست بدون ضمانت اجرایی است.

چرا کژساخته‌ها در فرایند سیاست‌گذاری خطرناک‌ترند؟

خطای هوش مصنوعی در یک استفادهٔ فردی ممکن است به سوءبرداشت محدود شود، اما در سیاست‌گذاری، خطا می‌تواند نهادی گردد. یعنی در سند رسمی، برنامهٔ بودجه، نقشهٔ راه، نظام ارزیابی، فراخوان پژوهشی، اولویت فنّاوری یا روایت ملی وارد شود. از این لحظه، کژساخته دیگر صرفاً خروجی یک ابزار نیست؛ بخشی از واقعیت تصمیم‌گیری می‌شود. این نکته برای توسعهٔ فنّاوری اهمیتی ویژه دارد. فنّاوری‌ها فقط براساس ظرفیت فنی رشد نمی‌کنند، بلکه در بستر نهادها، سرمایه‌گذاری‌ها، مقررات، انتظارات اجتماعی، سیاست‌های صنعتی و رقابت‌های ژئوپلیتیک شکل می‌گیرند (Borup, Brown, Konrad, & Van Lente, 2006; Geels, 2002; Markard & Truffer, 2008). بنابراین، یک خطای تحلیلی دربارهٔ وضعیت یا آیندهٔ یک فنّاوری می‌تواند منابع را جابه‌جا کند، توجه سیاستی را منحرف کند، فنّاوری‌های مهم اما کم‌صدا را نادیده بگیرد، یا به فنّاوری‌های پرصدا اما کم‌پشتوانه وزن بیش از حد بدهد.

در اینجا باید یک نکته را روشن کنیم. نقد کژساخته‌های هوش مصنوعی به معنای مخالفت با استفاده از هوش مصنوعی در سیاست‌گذاری نیست. چنین موضعی نه واقع‌بینانه است و نه مفید. هوش مصنوعی می‌تواند در جست‌وجوی منابع، کشف الگو، مقایسهٔ اسناد، شناسایی شکاف‌ها، تحلیل روایت‌ها و افزایش سرعت کار پژوهشی نقش مهمی داشته باشد. و حتی در ادبیات مرورهای علمی نیز تأکید شده است که هوش مصنوعی می‌تواند مرور ادبیات را پویاتر، گسترده‌تر و نزدیک‌تر به زمان واقعی کند، اما همچنان نیاز به نظارت انسانی، ارزیابی انتقادی و تشخیص زمینه‌ای باقی می‌ماند.

پس مسئلهٔ اصلی این است که نظام سیاست‌گذاری باید دقت و دانش کافی برای تشخیص «کژساخته» در مقابل «تحلیل» را داشته باشد. و تشخیص تفاوت این دو همیشه آسان نیست. گاهی هر دو از نظر زبانی مشابه‌اند و هر دو می‌توانند ساختارمند، مستند و منطقی به نظر برسند. تفاوت در پشتوانهٔ معرفتی، حساسیت به زمینه، شفافیت عدم قطعیت و قابلیت راستی‌آزمایی است.

مسئولیت کژساخته‌ها با کیست؟

مسئولیت کژساخته‌های هوش مصنوعی را نمی‌توان به یک بازیگر نسبت داد. این مسئولیت زنجیره‌ای و توزیع‌شده است. اما توزیع‌شده بودن مسئولیت نباید به بی‌مسئولیتی جمعی تبدیل شود. طراحان و توسعه‌دهندگان مدل مسئول معماری الگو، داده‌های آموزشی، محدودیت‌های فنی، سطح شفافیت و هشدار دربارهٔ حدود استفاده‌اند. سکوها مسئول شیوهٔ نمایش خروجی، سطح قطعیت زبانی و بازنمایی، امکان ارجاع‌دهی، کنترل منابع و طراحی تجربهٔ کاربری‌اند. اگر سامانه‌ای خروجی نامطمئن را با لحنی قاطع، روان و بدون نشان دادن سطح تردید ارائه کند، در تولید کژساختهٔ قطعیت نقش دارد. تحلیلگران، پژوهشگران و مشاوران سیاستی مسئول مرحلهٔ بعدی‌اند. آن‌ها نباید خروجی هوش مصنوعی را مستقیماً به شواهد سیاستی تبدیل کنند. مسئولیت آن‌ها شامل راستی‌آزمایی، مقایسه با منابع

مستقل، آشکارسازی عدم‌قطعیت، بررسی زمینهٔ نهادی و حفظ اختلاف دیدگاه‌هاست. در سیاست‌گذاری فنّاوری، این مسئولیت سنگین‌تر از کاربردهای عمومی است، زیرا خطا می‌تواند وارد تصمیم شود.

اما مهم‌ترین لایه، مسئولیت سازمان سفارش‌دهنده و نهاد سیاست‌گذار است. بسیاری از کژساخته‌ها به این دلیل تولید می‌شوند که نظام تصمیم‌گیری بدون توجه به ساختار و بافتار مسئله خروجی سریع، ساده، عددی و قابل ارائه می‌خواهد. وقتی سازمان‌ها فقط رتبه‌بندی، نمایش فشردهٔ اطلاعات، توصیهٔ کوتاه یا جمع‌بندی قطعی می‌خواهند، خودشان بخشی از فرایند تولید کژساخته می‌شوند. در چنین وضعی، هوش مصنوعی فقط خطا را تولید نمی‌کند؛ به تقاضای نهادی برای قطعیت پاسخ می‌دهد. بنابراین، باید میان دو سطح مسئولیت تفاوت گذاشت: مسئولیت تولید کژساخته و مسئولیت تبدیل کژساخته به تصمیم. ممکن است الگو یا سکو در تولید خروجی خطادار نقش داشته باشد. اما مسئولیت نهایی زمانی سنگین‌تر می‌شود که تحلیلگر یا سیاست‌گذار آن خروجی را به مبنای تصمیم، بودجه، اولویت، سند رسمی یا روایت نهادی تبدیل کند.

توصیه‌هایی برای مدیریت خطای دیدناشی از ابزارهای هوشمند

بدیهی است که راه‌حل، حذف هوش مصنوعی از سیاست‌گذاری نیست. همچنین اعتماد کامل به آن نیز قابل دفاع نخواهد بود. آنچه لازم است، نوعی تحلیل سیاستی حساس به کژساخته است. چنین رویکردی چند اصل ساده اما جدی دارد که در ادامه به آن‌ها اشاره می‌شود.

نخست، خروجی هوش مصنوعی نباید بدون ثبت منشأ داده و حدود استفاده وارد گزارش سیاستی شود. اگر معلوم نباشد داده از کجا آمده، چه منابعی حذف شده‌اند و چه بافتاری غالب بوده است، خروجی نمی‌تواند مبنای تصمیم جدی قرار گیرد. دوم، عدم قطعیت باید در متن حفظ شود. اگر یک موضوع واقعاً نامطمئن است، نباید با زبان قطعی، نمودار ساده یا توصیهٔ نهایی پوشانده شود. سیاست‌گذاری خوب همیشه به قطعیت بیشتر نیاز ندارد؛ گاهی به نمایش دقیق‌تر ندانستن و عدم قطعیت نیاز دارد. سوم، خروجی هوش مصنوعی باید با شواهد غیرمتنی و زمینه‌ای مقایسه شود. در توسعهٔ فنّاوری، مقاله و پتنت کافی نیست. باید ظرفیت صنعتی، زیرساخت، بازار، مقررات، سرمایهٔ انسانی، پذیرش اجتماعی و امکان اجرا نیز بررسی شود. چهارم، اختلاف دیدگاه‌ها نباید بیش از حد ساده شوند. اگر گروه‌های مختلف یک مسئله را متفاوت می‌فهمند، این اختلاف بخشی از دادهٔ سیاستی است، نه مانع تحلیل. هوش مصنوعی نباید نقش ماشین تولید اجماع مصنوعی را ایفا کند. پنجم، هر توصیهٔ سیاستی باید از آزمون ظرفیت اجرا عبور کند. توصیه‌ای که ابزار، متولی، منابع، زمان‌بندی، تعارض ذی‌نفعان و پیامدهای احتمالی آن روشن نیست، هنوز سیاست نیست. فقط یک متن با ادبیات سیاست‌گذاری است.

نتیجه‌گیری

هوش مصنوعی سیاست‌گذاری فنّاوری را سریع‌تر خواهد کرد. اما سریع‌تر شدن الزاماً به معنای بهتر شدن نیست. در بسیاری از موارد، سرعت می‌تواند خطا را زودتر به تصمیم تبدیل کند. خطر اصلی هوش مصنوعی در سیاست‌گذاری این نیست که گاهی پاسخ نادرست می‌دهد. خطر اصلی آن است که پاسخ نادرست، ناقص یا بی‌توجه به بافتار را در قالبی تولید می‌کند که برای نهادهای تصمیم‌گیر بیش از حد ساده‌سازی شده است.

از این منظر، کژساخته‌های هوش مصنوعی فقط خطاهای فنی نیستند. آن‌ها خطاهای معرفتی و نهادی‌اند. در سطح معرفتی، واقعیت را ساده، قطعی یا هموار نشان می‌دهند. در سطح نهادی، می‌توانند به سند، اولویت، بودجه، نقشهٔ راه یا روایت رسمی تبدیل شوند. بنابراین، عبور از خطای تحلیلی به خطای تصمیم، لحظهٔ اصلی ورود خطر سیاستی است. سیاست‌گذاری و توسعهٔ فنّاوری به هوش مصنوعی نیاز خواهد داشت، اما نه به‌عنوان جایگزین قضاوت. نقش مناسب هوش مصنوعی، کمک به دیدن الگوها، پرسیدن پرسش‌های بهتر، مقایسهٔ منابع و آشکارکردن شکاف‌هاست. اما تشخیص معنای این الگوها، ارزیابی عدم قطعیت، فهم زمینه و پذیرش مسئولیت تصمیم همچنان انسانی و نهادی باقی می‌ماند.

آیندهٔ سیاست‌گذاری با هوش مصنوعی، الزاماً هوشمندتر نخواهد شد. اگر کژساخته‌ها شناسایی نشوند، این آینده ممکن است فقط سریع‌تر، قانع‌کننده‌تر و شکننده‌تر شود. معیار بلوغ در استفاده از هوش مصنوعی این نیست که چقدر از آن استفاده می‌کنیم؛ معیار اصلی این است که آیا می‌توانیم خطای دید را در خروجی‌هایی تشخیص دهیم که بیش از حد منظم، دقیق و قانع‌کننده به نظر می‌رسند یا نه؟

References

Barrett, J. F., & Keat, N. (2004). Artifacts in CT: Recognition and avoidance. Radiographics, 24(6), 1679-1691. https://doi.org/10.1148/rg.246045065

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. Technology Analysis & Strategic

Management, 18(3-4), 285-298. https://doi.org/10.1080/09537320600777002

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. Frontiers in Human Dynamics, 6, 1421273. https://doi.org/10.3389/fhumd.2024.1421273

de Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. Government Information Quarterly, 39(2), 101666. https://doi.org/10.1016/j.giq.2021.101666

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., ... & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. Computational Linguistics, 50(3), 1097-1179. https://doi.org/10.1162/coli_a_00524

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case study. Research Policy, 31(8), 1257-1274. https://doi.org/10.1016/S0048-7333(02)00062-8

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121-127. https://doi.org/10.1136/amiajnl-2011-000089

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ... & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems, 43(2), 1-55. https://doi.org/10.1145/3703155

Ibitoye, A. O., Nkwo, M. S., & Orji, R. (2025). Rethinking responsible AI from ethical pillars to sociotechnical practice. AI and Ethics, 5(6), 6207-6223. https://doi.org/10.1007/s43681-025-00809-2

Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2022). Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. Social Science Computer Review, 40(2), 478-493. https://doi.org/10.1177/0894439320980118

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business Horizons, 61(4), 577-586. https://doi.org/10.1016/j.bushor.2018.03.007

Leonardi, P. M., & Leavell, V. (2026). Knowing enough to be dangerous: The problem of "artificial certainty" for expert authority when using AI for decision making and planning. Organization Science, 37(2), 516-543. https://doi.org/10.1287/orsc.2023.18224

Lnenicka, M., Clarinval, A., Nikiforova, A., Rudmark, D., Luterek, M., Symeonidis, D., & Bolívar, M. P. R. (2026). Artificial intelligence in policymaking: Mapping integration gaps across the public policy cycle. Government Information Quarterly, 43(2), 102138. https://doi.org/10.1016/J.GIQ.2026.102138

Markard, J., & Truffer, B. (2008). Technological innovation systems and the multi-level perspective: Towards an integrated framework. Research Policy, 37(4), 596-615. https://doi.org/10.1016/j.respol.2008.01.004

Mikalef, P., Benlian, A., Conboy, K., & Tarafdar, M. (2025). Responsible AI starts with the artifact: Challenging the concept of responsible AI in IS research. European Journal of Information Systems, 34(3), 407-414. https://doi.org/10.1080/0960085X.2025.2506875

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI. European Journal of Information Systems, 31(3), 257-268. https://doi.org/10.1080/0960085X.2022.2026621

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems, 34(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Resnik, D. B., & Hosseini, M. (2026). Hallucinated citations produced by generative artificial intelligence may constitute research misconduct when citations function as data in scholarly papers. Accountability in Research, 2645390. https://doi.org/10.1080/08989621.2026.2645390

Selten, F., Robeer, M., & Grimmelikhuijsen, S. (2023). 'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. Public Administration Review, 83(2), 263-278. https://doi.org/10.1111/puar.13602

Sharman, A., & Holmes, J. (2010). Evidence-based policy or policy-based evidence gathering? Biofuels, the EU and the 10% target. Environmental Policy and Governance, 20(5), 309-321. https://doi.org/10.1002/eet.543

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. California Management Review, 61(4), 66-83. https://doi.org/10.1177/0008125619862257

Stoeber, T., Hammerschmidt, J., Lundervold, A., Kromidha, E., Kanbach, D. K., & Kraus, S. (2026). AI strategy

under institutional pressure: strategic conformity and decision-making in large language models. Journal of Business Research, 212, 116227. https://doi.org/10.1016/j.jbusres.2026.116227

Zaitsev, M., Maclaren, J., & Herbst, M. (2015). Motion artifacts in MRI: A complex problem with many partial solutions. Journal of Magnetic Resonance Imaging, 42(4), 887-901. https://doi.org/10.1002/jmri.24850


رهنت