



OWASP AIBOM

راهنمای مبانی

سپتامبر ۲۰۲۶

فهرست مطالب

خلاصه اجرایی.....	۶
۱. درک صورت حساب مواد هوش مصنوعی	۷
درک صورت حساب مواد هوش مصنوعی به عنوان اسناد حاکمیتی.....	۸
۲. AIBOM چیست؟	۱۰
۳. چرا AIBOM ها اهمیت دارند؟	۱۲
محرك‌های امنیتی	۱۲
محرك‌های مقرراتی	۱۳
محرك‌های حاکمیتی.....	۱۳
تداوم کسب و کار.....	۱۴
تهدیدهای زنجیره تأمین.....	۱۴
۴. مؤلفه‌های اصلی یک AIBOM.....	۱۶
سرآیند.....	۱۶
فهرست مؤلفه‌ها	۱۷
جریان‌های داده	۱۸
مناطق اعتماد و مرزها	۱۹
رفتارها.....	۱۹
آنچه یک AIBOM امکان پذیر می‌کند.....	۲۰
۵. ساختار AIBOM.....	۲۱
آنچه هر AIBOM اعلام می‌کند.....	۲۱
مثال کاربردی: یک سامانه استنتاج میزبانی شده.....	۲۲
مثال کاربردی: یک خط لوله تنظیم دقیق	۲۵
آنچه AIBOM بیان نمی‌کند	۲۶
یک AIBOM چگونه نوشته می‌شود؛ برای خوانندگانی که عمیق تر می‌روند	۲۷
۶. AIBOM برای سامانه‌های هوش مصنوعی عامل محور	۲۸
وابستگی‌های چندعاملی	۲۸
ابزارهای مرتبط	۲۹
اتصال دهنده‌های شخص ثالث	۳۰
ردیابی اکوسیستم MCP/A2A.....	۳۱
۷. نقشه راه پذیرش سازمانی.....	۳۲
مرحله ۱ – آگاهی و خط مبنا	۳۳
مرحله ۲ – آزمایش آزمایشی (پایلوت)	۳۴
مرحله ۳ – یکپارچه سازی با حاکمیت	۳۵
مرحله ۴ – عملیات مستمر.....	۳۶
قدردانی	۳۸

از طرف رهبر پروژه AIBOM

مدتی ریاست مشترک ابتکار SBOM را در CISA وابسته به وزارت امنیت داخلی آمریکا (DHS) بر عهده داشتیم، و درسی که بیش از همه در ذهنم ماند درباره قالب‌های فایل نبود. درباره اتفاقی بود که پس از عرضه قالب می‌افتد. ما صنعت نرم‌افزار را به تولید SBOM در مقیاس وسیع رساندیم و سپس دیدیم که پذیرش درست در همان نقطه متوقف شد، زیرا تولید هرگز بخش دشوار کار نبود. بخش دشوار، ساختن لایه عملیاتی پیرامون آن بود: چه کسی این سند را مصرف می‌کند، به چه تصمیمی خوراک می‌دهد، وقتی کهنه شد چه می‌شود، و وقتی اصلاً وجود ندارد چه کسی پاسخ‌گوست. AIBOM 101 با در نظر داشتن همین درس نوشته شده است. این سند مرجع طرح‌واره (schema) نیست. این همان لایه عملیاتی برای سامانه‌های هوش مصنوعی است که هرگز آن را به‌طور کامل برای نرم‌افزار نساختم.

یک صورت‌حساب مواد هوش مصنوعی تنها زمانی جایگاه خود را به دست می‌آورد که آنچه یک تیم امنیتی، یک حسابرس یا یک بازبین تدارکات واقعاً می‌تواند انجام دهد را تغییر دهد. به همین دلیل این سند به همان اندازه که به مؤلفه‌ها و جریان‌های داده می‌پردازد، به دامنه، ادعاهای کامل‌بودن و شواهد نیز زمان اختصاص می‌دهد. سابقه‌ای که هیچ‌کس نتواند به آن اعتماد کند و هیچ‌کس ملزم به مصرف آن نباشد، زیرساخت نیست؛ یک محصول جانبی مهندسی است که برچسب حاکمیت بر خود زده است. هدف ما از این 101 این است که چیزی به دست رهبران امنیت سازمانی بدهیم که بتوانند فوراً به کار گیرند: راهی برای پرسیدن اینکه چه هوش مصنوعی‌ای دارند، به چه چیزی وابسته است، و آیا کسی می‌تواند برای آن وابستگی‌ها ضمانت کند، با زبانی که هم در اتاق هیئت‌مدیره و هم در خط لوله ساخت معتبر بماند.

ما این سند را برای خوانده‌شدن نوشته‌ایم، نه صرفاً برای ارجاع. مثال‌های کاربردی درون آن، یعنی یک سامانه استنتاج میزبانی‌شده و یک خط لوله تنظیم دقیق، استدلالی را که یک تولیدکننده واقعاً باید انجام دهد گام‌به‌گام نشان می‌دهند: مرز سامانه کجاست، چه چیزی شاهد به شمار می‌آید، و یک ادعای کامل‌بودن اجازه دلالت بر چه چیزی را دارد. نقشه راه پذیرش سازمانی همین کار را برای جنبه سازمانی انجام می‌دهد: از یک فهرست نام‌دار از سامانه‌های هوش مصنوعی، از طریق یک آزمایش پایلوت، به یکپارچه‌سازی با حاکمیت، و سرانجام به عملیات مستمر که در آن AIBOM‌ها بر اساس محرک‌های تعریف‌شده به‌طور خودکار بازتولید می‌شوند. هر مرحله یک معیار خروج دارد که می‌توانید آن را بررسی کنید، و این عمده است، زیرا نقشه راهی که راهی برای دانستن عبور از یک مرحله نداشته باشد، تنها یک آرزوست.

ما این سند را همچنین برای حرکت‌کردن ساخته‌ایم. هوش مصنوعی عامل‌محور یک مورد حاشیه‌ای آینده نیست؛ هم‌اکنون معنای «وابستگی» را دگرگون می‌کند، زیرا فراخوانی ابزارها، واگذاری به زیرعامل‌ها و اتصال‌های MCP یک سامانه اکنون می‌تواند در زمان اجرا تعیین شود و نه در زمان طراحی. به‌جای آنکه این را استثنایی بدانیم که بعداً به آن الحاق شود، AIBOM را از ابتدا به‌صورت یک گراف طراحی کردیم؛ گرافی که به‌طور طبیعی به چیزی گسترش می‌یابد که آن را AgBOM می‌نامیم: هویت عامل‌ها، مجوزهای ابزار و الگوهای تعامل چندعاملی که با همان دقت مدل‌ها و مجموعه داده‌ها ثبت می‌شوند. این معماری به‌گونه‌ای است که با تحول این حوزه انعطاف پیدا کند، بی‌آنکه با ظهور هر الگوی جدید به بازنویسی نیاز باشد.

هیچ‌یک از این‌ها به‌تنهایی اتفاق نمی‌افتد. این سند حاصل جریان‌های کاری پروژه OWASP AIBOM است: الزامات، بهترین روش‌ها، یکپارچگی و کیفیت، هوش تهدید، سیاست‌گذاری، محتوا، اتحادها، مشارکت، حمایت مالی و AgBOM، که به‌صورت یک نظام به‌هم‌پیوسته اداره می‌شوند و نه تلاش‌هایی جداگانه. هر جریان کاری ما را واداشته است تا راهنمایی را درباره آنچه یک AIBOM می‌تواند و نمی‌تواند ادعا کند صادق نگه داریم. ما کوشیده‌ایم در سراسر سند درباره این مرز صریح باشیم: یک AIBOM آنچه را یک سامانه در بر دارد ثبت می‌کند، نه اینکه آن سامانه ایمن یا منطبق با مقررات است یا نه. این داوری با خواننده و با استفاده از چارچوب ریسک خود او باقی می‌ماند، و با بلوغ این پروژه نیز همین‌گونه خواهد ماند. پروژه، کار طرح‌واره و انتشارهای جاری آن در owaspaiбом.org قرار دارند و ما از مشارکت‌کنندگان هر یک از جریان‌های کاری بالا استقبال می‌کنیم.

اگر SBOM چیزی به صنعت نرم‌افزار آموخته باشد، این است که یک قالب خوب بدون انضباط عملیاتی پیرامون آن، سندی است که در قفسه خاک می‌خورد. هدف ما برای AIBOM این است که هرگز به چنین سندی تبدیل نشود. امیدوارم این سند نقطه آغازی به شما بدهد تا این هدف را در سازمان خودتان محقق کنید.

Arunesh Salhotra
رهبر پروژه، OWASP AIBOM
www.owaspai bom.org

دامنه و مخاطبان

این سند مقدمه‌ای بر صورت‌حساب مواد هوش مصنوعی (AIBOM) است: اینکه چیست، چه چیزی باید در بر داشته باشد، چگونه تولید و اعتبارسنجی می‌شود، و یک سازمان چگونه آن را می‌پذیرد. مخاطبان اصلی این سند کسانی هستند که AIBOM ها را تولید و مصرف می‌کنند: مهندسان هوش مصنوعی و یادگیری ماشین، تیم‌های MLOps و پلتفرم، معماران امنیت، و متخصصان امنیت محصول و ریسک زنجیره تأمین. این سند به‌گونه‌ای نوشته شده که کسی با تجربه اندک یا بدون تجربه قبلی در AIBOM بتواند آن را از ابتدا تا انتها بخواند؛ آشنایی با روش‌های SBOM نرم‌افزاری مفید است اما پیش‌فرض گرفته نشده است.

همچنین هدف ما آگاه‌سازی مدیران ارشد امنیت اطلاعات (CISO)، رهبران حاکمیت هوش مصنوعی و مدیران ارشدی است که مسئول ریسک زنجیره تأمین هوش مصنوعی هستند؛ این افراد از چارچوب‌بندی این سند درباره اهمیت AIBOM، محرک‌های امنیتی، مقرراتی و حاکمیتی پشت آن، و نقشه راه پذیرش مرحله‌ای و شاخص‌های بلوغ بهره خواهند برد. افزون بر این، این سند به بستر گسترده‌تر مقرراتی و حاکمیتی پیرامون سامانه‌های هوش مصنوعی نیز می‌پردازد و ممکن است برای تیم‌های انطباق، حقوقی، و تدارکات یا ریسک اشخاص ثالث که الزامات افشای AIBOM را برای تأمین‌کنندگان تعیین می‌کنند مفید باشد.

این سند ساختار AIBOM را به‌صورت مفهومی توصیف می‌کند و نشان می‌دهد این ساختار چگونه بر یک قالب BOM ماشین‌خوان و تثبیت‌شده نگاشت می‌شود. این سند مرجع طرح‌واره در سطح فیلد نیست؛ مشخصات قالب همچنان مرجع معتبر برای جزئیات کدگذاری است و خوانندگان باید به نسخه جاری آن مراجعه کنند و مثال‌های این سند را طرح‌واره تلقی نکنند. این سند طبقه‌بندی تهدید برای سامانه‌های هوش مصنوعی، روش‌شناسی مدیریت ریسک هوش مصنوعی، یا دستورالعمل‌های پیاده‌سازی ویژه فروشندگان و ابزارها را ارائه نمی‌کند؛ این موارد در منابع همراه OWASP و در استانداردهای ارجاع‌شده نگهداری می‌شوند و در سراسر سند به آن‌ها استناد شده است.

حق نشر و مجوز

© بنیاد OWASP. این سند تحت مجوز Creative Commons CC BY-SA 4.0 منتشر شده است. OWASP و نشان OWASP علائم تجاری بنیاد OWASP هستند. مشارکت‌کنندگان در بخش قدردانی معرفی شده‌اند؛ مشارکت به معنای تأیید از سوی کارفرمای هیچ‌یک از مشارکت‌کنندگان نیست. متن کامل مجوز را در نشانی زیر ببینید: <https://creativecommons.org/licenses/by-sa/4.0/legalcode>

یادداشت ترجمه: این متن ترجمه فارسی «OWASP AIBOM Foundations Guide v1.0» است که مطابق مجوز CC BY-SA 4.0 تهیه شده و تحت همان مجوز در دسترس است. در صورت هرگونه اختلاف، نسخه انگلیسی مرجع معتبر است. متن درون تصاویر و نمودارها به زبان اصلی حفظ شده است.

خلاصه اجرایی

بیشتر سازمان‌ها نمی‌توانند با اطمینان بگویند کدام مدل در محیط تولید در حال اجراست، داده‌ها و تأمین‌کنندگان بالادستی پشت آن را ردیابی کنند، یا نشان دهند که آیا یک آسیب‌پذیری تازه افشا شده به سامانه‌ای که به آن وابسته‌اند می‌رسد یا نه. سامانه‌های هوش مصنوعی مدرن از مدل‌های شخص ثالث، داده‌های دارای مجوز و داخلی، چارچوب‌های متن‌باز، خطوط لوله بازیابی، عامل‌ها، ابزارها و سرویس‌های تحت مدیریت ارائه‌دهنده سرهم می‌شوند؛ که بسیاری از آن‌ها بدون هیچ انتشار محلی تغییر می‌کنند. فهرست‌برداری سنتی نرم‌افزار برای ثبت این مصنوعات غیرکدی و سرویس‌های خارجی ساخته نشده است، در حالی که رفتار یک سامانه هوش مصنوعی، و ریسک آن، عمدتاً در همین‌جا تعیین می‌شود.

صورت‌حساب مواد هوش مصنوعی (AIBOM) این شکاف را پر می‌کند: شرحی ساخت‌یافته و ماشین‌خوان از آنچه یک سامانه هوش مصنوعی واقعاً در بر دارد و اینکه اطلاعات چگونه در آن جریان می‌یابد. AIBOM بنا بر طراحی، واقعیت را ثبت می‌کند و نه امتیاز یا حکم انطباق را؛ و همین است که به آن اجازه می‌دهد به‌عنوان بنیانی مشترک و قابل‌راستی‌آزمایی برای تصمیم‌های امنیتی، حاکمیتی، تدارکاتی و حسابرسی به کار رود.

به‌ویژه برای رهبران امنیت، این اهرمی بر تصمیم‌هایی است که هم‌اکنون مسئولیت آن‌ها را بر عهده دارید. یک AIBOM سامانه‌های هوش مصنوعی را از وابستگی‌هایی مبهم به وابستگی‌هایی پاسخ‌گو تبدیل می‌کند؛ زمان واکنش به آسیب‌پذیری را از چند هفته به چند ساعت کاهش می‌دهد، به تدارکات و مدیریت ریسک اشخاص ثالث یک استاندارد افشای مشخص برای الزام تأمین‌کنندگان می‌دهد، و یک پایگاه شواهد واحد و قابل‌دفاع برای بازبینی‌های حاکمیتی و گزارش‌دهی به هیئت‌مدیره فراهم می‌کند. مؤثرترین شیوه پذیرش، راه‌اندازی یک برنامه موازی جدید نیست، بلکه تبدیل AIBOM به ورودی الزامی کنترل‌هایی است که هم‌اکنون اجرا می‌کنید: دروازه‌های انتشار، ارزیابی فروشندگان، بازبینی ریسک مدل و واکنش به حوادث؛ تا یک سابقه بدون آنکه سه بار تهیه شود، هم به امنیت، هم به انطباق و هم به حسابرسی خدمت کند.

این سند نقطه ورودی عملی برای تولید چنین سندی است. توضیح می‌دهد AIBOM چیست و چرا در حال تبدیل‌شدن به یک الزام عملیاتی است و نه یک روش اختیاری؛ با مثال‌های کاربردی نشان می‌دهد یک AIBOM چگونه ساختاربندی و تولید می‌شود؛ مدل را به الزامات افزوده سامانه‌های هوش مصنوعی عامل‌محور گسترش می‌دهد؛ و نقشه راهی مرحله‌ای برای پذیرش AIBOM‌ها در سراسر یک سازمان ارائه می‌کند.

۱. درک صورت حساب مواد هوش مصنوعی

سامانه‌های هوش مصنوعی به چیزی بیش از کد وابسته‌اند. یک سامانه تولیدی از وزن‌های مدل، مجموعه داده‌های آموزش و تنظیم دقیق، آدایتورها و نقاط بازرسی (checkpoint)، پرامپت‌ها و قالب‌ها، پیکره‌های بازیابی، سرویس‌های حفاظ (guardrail) و تعدیل محتوا، چارچوب‌های هماهنگ‌سازی (orchestration) و API‌های تحت مدیریت ارائه‌دهنده بهره می‌گیرد. این‌ها با هم زنجیره تأمین هوش مصنوعی را تشکیل می‌دهند، و این زنجیره از یک جنبه تعیین‌کننده با زنجیره تأمین نرم‌افزار متفاوت است: مهم‌ترین وابستگی‌های آن اغلب اصلاً بسته نرم‌افزاری نیستند. آن‌ها مصنوعات غیرکدی و سرویس‌هایی هستند که از بیرون اداره می‌شوند و رفتارشان می‌تواند بدون هیچ انتشار محلی تغییر کند.

این وابستگی‌ها در یک مرحله واحد متمرکز نمی‌شوند؛ در طول چرخه عمر انباشته می‌شوند. در مرحله طراحی، تیم‌ها ارائه‌دهندگان مدل و منابع داده را انتخاب می‌کنند. در مرحله توسعه، داده‌های تنظیم دقیق، پرامپت‌ها و دارایی‌های ارزیابی را وارد می‌کنند. در مرحله استقرار، زیرساخت، کنترل‌های ایمنی و یکپارچه‌سازی‌ها را می‌افزایند. در محیط تولید، سامانه‌ها به پیکره‌های بازیابی‌ای وابسته‌اند که پیوسته تغییر می‌کنند، به سرویس‌های تعدیل محتوا که خروجی‌ها را به‌طور اساسی شکل می‌دهند، و به به‌روزرسانی‌های سمت ارائه‌دهنده که طبق برنامه دیگری از راه می‌رسند. بنابراین یک وابستگی حیاتی می‌تواند در هر مرحله‌ای وارد شود، و بخش بزرگی از آن بیرون از دید مستقیم سازمان مستقرکننده قرار دارد.

یک دستیار سازمانی را در نظر بگیرید که بر پایه API یک مدل پایه میزبانی‌شده ساخته شده است. از دید تیم برنامه هیچ چیز تغییر نکرده است: پرامپت‌ها، پیکربندی بازیابی و زیرساخت همه مانند قبل‌اند. با این حال، ارائه‌دهندگان گزارش تغییرات API منتشر می‌کنند و سیاست‌های چرخه عمری دارند که مدل‌های فعال، منسوخ و بازنشسته را از هم متمایز می‌کند، و یک نام مستعار نسخه‌دار می‌تواند به یک نسخه لحظه‌ای (snapshot) جدیدتر اشاره داده شود. مدلی که ارزیابی و تأیید شده بود ممکن است دیگر مدلی نباشد که رفتار محیط تولید را شکل می‌دهد. وقتی کیفیت خروجی تغییر می‌کند یا حادثه‌ای رخ می‌دهد، تیم نمی‌تواند به سه پرسش نخست که از آن پرسیده می‌شود پاسخ دهد؛ یعنی چه چیزی تغییر کرد، چه زمانی، و آیا محرک آن بالادستی بود؛ زیرا اطلاعات لازم برای پاسخ به آن‌ها هرگز در هیچ کجا ثبت نشده بود.

همین شکاف بدون حضور یک ارائه‌دهنده میزبان نیز پدیدار می‌شود. خطوط لوله تنظیم دقیقی که ردیابی تبار ندارند نمی‌توانند نشان دهند کدام مدل پایه و کدام داده یک مصنوع مستقرشده را تولید کرده است. منابع بازیابی بدون نسخه‌گذاری تغییر می‌کنند. لایه‌های ایمنی و تعدیل محتوا با وجود اثری که بر خروجی‌ها دارند، جزئیات پیاده‌سازی تلقی می‌شوند. الگو ثابت است: سازمان‌ها قابلیت را به ارث می‌برند بی‌آنکه درک را به ارث ببرند. آن‌ها می‌دانند کدام برنامه را مستقر کرده‌اند، اما نمی‌توانند نشان دهند کدام مدل بالادستی، مجموعه داده یا لایه کنترل واقعاً رفتار آن را شکل داده است.

این‌جاست که فهرست‌برداری مرسوم نرم‌افزار به حد خود می‌رسد. یک SBOM مؤلفه‌های نرم‌افزاری و نسخه‌های اعلام‌شده آن‌ها را ثبت می‌کند؛ این کار همچنان ضروری است اما کافی نمی‌شود. تعیین اینکه یک سامانه هوش مصنوعی چه خواهد کرد، مستلزم دانستن مدل‌های ازپیش‌آموزش‌دیده،

داده‌هایی که بر آن‌ها آموزش دیده و ارزیابی شده‌اند، نتایج ارزیابی‌ای که استقرار را توجیه کرده‌اند، و سرویس‌های تحت مدیریت ارائه‌دهنده در مسیر درخواست است؛ که هیچ‌یک از آن‌ها به معنای SBOM مؤلفه نرم‌افزاری نیستند. فهرست نرم‌افزار به شما می‌گوید چه کدی حضور دارد. دیدپذیری زنجیره تأمین هوش مصنوعی باید همچنین به شما بگوید کدام مصنوعات غیرکدی و سرویس‌های بالادستی به‌طور اساسی رفتار را شکل داده‌اند، و با منشأیی کافی برای بازسازی سامانه آن‌گونه که در یک مقطع زمانی معین بوده است.

سامانه‌های عامل‌محور این وضعیت را تشدید می‌کنند، نه اینکه آن را تغییر دهند. وقتی یک عامل در زمان اجرا ابزارها را انتخاب می‌کند، سرویس‌های خارجی را فرامی‌خواند یا کار را به عامل‌های دیگر واگذار می‌کند، مجموعه وابستگی‌ها دیگر در زمان طراحی ثابت نیست، و سه دیدگاه نسبت به آن از هم فاصله می‌گیرند: آنچه پیکربندی اعلام می‌کند، آنچه مجوزها دسترس‌پذیر می‌کنند، و آنچه ردهای اجرا نشان می‌دهند که واقعاً استفاده شده است.

درک صورت‌حساب مواد هوش مصنوعی به‌عنوان اسناد حاکمیتی

یک AIBOM در درجه نخست یک ابزار راحتی مهندسی نیست. یک سند حاکمیتی است: سابقه‌ای که توسط یک طرف پاسخ‌گو تألیف شده و به شکلی است که طرف دیگری بتواند بعداً آن را بخواند و بر اساس آن اقدام کند. این تمایز اهمیت دارد زیرا خواننده معمولاً نویسنده نیست، و معمولاً تحت فشار سراغ آن می‌آید: یک مهندس امنیت که ساعت ۲ بامداد دامنه یک حادثه را تعیین می‌کند، یک بازبین تدارکات که افشای یک فروشنده را ارزیابی می‌کند، یا حسابرسی که می‌پرسد کدام مجموعه داده‌ها مدلی را آموزش داده‌اند که هجده ماه پیش تصمیمی مورد مناقشه گرفته است. چهار ویژگی یک سابقه را در چنین لحظاتی قابل‌استفاده می‌کند.

- **تألیف پاسخ‌گو:** هر ادعا در یک AIBOM توسط کسی مطرح شده است. فراداده‌های سند ثبت می‌کنند چه کسی آن را تألیف، بازبینی و تأیید کرده، چه زمانی و با چه روشی؛ و همین فراداده است که به خواننده اجازه می‌دهد سابقه‌ای را که از داده‌های دورسنجی ساخت (build telemetry) تولید شده از سابقه‌ای که در جریان بازبینی تدارکات دستی نوشته شده تشخیص دهد. دو سند می‌توانند ادعاهای یکسانی داشته باشند و سزاوار میزان اطمینان بسیار متفاوتی باشند.

- **دامنه اعلام‌شده و یک ادعای صریح کامل‌بودن:** یک AIBOM بیان می‌کند چه چیزی را پوشش می‌دهد و تولیدکننده آن چقدر مطمئن است که آن را پوشش می‌دهد. همین است که سکوت را تفسیرپذیر می‌کند. بدون ادعای کامل‌بودن، مجموعه داده‌ای که فهرست نشده هیچ چیزی را اثبات نمی‌کند؛ با وجود آن، غیاب آن به ادعایی از سوی تولیدکننده تبدیل می‌شود که می‌توان به آن اتکا کرد و اگر نادرست باشد، آن را به چالش کشید. فصل ۴ این خودتوصیف‌ها را به تفصیل بسط می‌دهد.

- **ادعاهای متصل به شواهد:** هش‌ها، امضاها، کارت‌های مدل، سوابق مجموعه داده و نتایج ارزیابی به ادعاهای منفرد پیوست می‌شوند، و این کار ادعا را به چیزی تبدیل می‌کند که مصرف‌کننده بتواند مستقلاً بررسی کند. یک AIBOM لازم نیست برای هر ادعا شاهد ارائه کند؛ جایی که شاهد وجود ندارد، همین غیاب خود اطلاعاتی است که خواننده برای تصمیم‌گیری درباره اینکه آن ادعا چه وزنی را می‌تواند تحمل کند از آن استفاده می‌کند.

• **چرخه عمر کنترل شده:** یک سند حاکمیتی نسخه‌گذاری، بازبینی و تأیید می‌شود، در صورت تغییر اساسی به‌روزرسانی می‌شود، نگهداری می‌شود، و هنگامی که سامانه‌ای که توصیف می‌کند بازنشسته می‌شود، حفظ می‌گردد. AIBOM ی که یک بار تولید شده و هرگز تازه‌سازی نشده، سامانه‌ای را توصیف می‌کند که دیگر وجود ندارد، و خواندن آن از نداشتن هیچ چیز بدتر است، زیرا اطمینانی کاذب ایجاد می‌کند. فصل ۷ مالکیت و محرک‌های به‌روزرسانی ناشی از این موضوع را تشریح می‌کند.

این ویژگی‌ها هستند که به AIBOM اجازه می‌دهند به‌عنوان شاهد در تصمیم‌هایی که هم‌اکنون وجود دارند مصرف شود: ارزیابی تأمین‌کننده و الزامات افشای قراردادی، دروازه‌گذاری انتشار، تحلیل اثر آسیب‌پذیری، تعیین دامنه حادثه، بازبینی ریسک مدل، و پاسخ به نهادهای ناظر یا حسابرسی.

درک آنچه یک AIBOM نیست: AIBOM امتیاز، رتبه‌بندی یا درجه بلوغ نیست؛ آنچه را یک سامانه در بر دارد ثبت می‌کند، نه اینکه آن سامانه چقدر خوب یا چقدر ایمن است. انطباق را ادعا نمی‌کند، و جایگزین ارزیابی، مدل‌سازی تهدید یا ارزیابی ریسک نیست؛ بلکه ورودی‌ای است که این فعالیت‌ها مصرف می‌کنند، و ارزش آن به انجام‌شدن آن‌ها بستگی دارد. AIBOM به‌خودی‌خود یک کنترل هم نیست: هیچ چیزی را مسدود نمی‌کند تا زمانی که سازمان یک دروازه، یک بازبینی یا یک پرس‌وجو بسازد که آن را بخواند. AIBOM ی که هیچ‌کس آن را مصرف نکند، یک محصول جانبی مهندسی است که برچسب حاکمیت بر خود زده است.

۲. AIBOM چیست؟

صورت حساب مواد هوش مصنوعی، SBOM برای هوش مصنوعی، یا AIBOM، توصیفی ساخت یافته و ماشین خوان از یک سامانه هوش مصنوعی در چهار بخش است: سرآیندی که اعلام می کند سند چه چیزی را پوشش می دهد و ادعا می کند چقدر کامل است؛ فهرستی از مؤلفه ها و سرویس های سامانه مانند مدل ها، مجموعه داده ها، عامل ها و ابزارها، حفاظ ها و عناصر زمان اجرا؛ جریان های داده جهت دار میان آن ها؛ و شواهد منشأ، حقوق، یکپارچگی و ارزیابی که به این ادعاها پیوست شده اند.

هدف یک AIBOM بهبود شفافیت و اعتماد در سراسر زنجیره تأمین هوش مصنوعی است. AIBOM دیدپذیری فراهم می کند و این امکان را به سازمان ها می دهد که هر مؤلفه از اکوسیستم هوش مصنوعی خود را بررسی و تحلیل کنند. AIBOM از روش ها و فرایندهای مدیریت ریسک در حال توسعه و موجود پشتیبانی می کند. به شناسایی آسیب پذیری ها، ارزیابی ریسک های امنیتی مرتبط با هوش مصنوعی و ردیابی منشأ مدل از توسعه تا استقرار کمک می کند. AIBOM از تلاش های سازمانی برای برآورده کردن الزامات مقرراتی و مدیریت ریسک مربوط پشتیبانی می کند.

اهداف اصلی

یک AIBOM سابقه ای جامع از مصنوعات هوش مصنوعی و مؤلفه های چرخه عمر یک سامانه هوش مصنوعی در سراسر چرخه طراحی، توسعه و استقرار آن فراهم می کند که می تواند به عنوان شاهد اصلی در حسابرسی های هوش مصنوعی به کار رود.

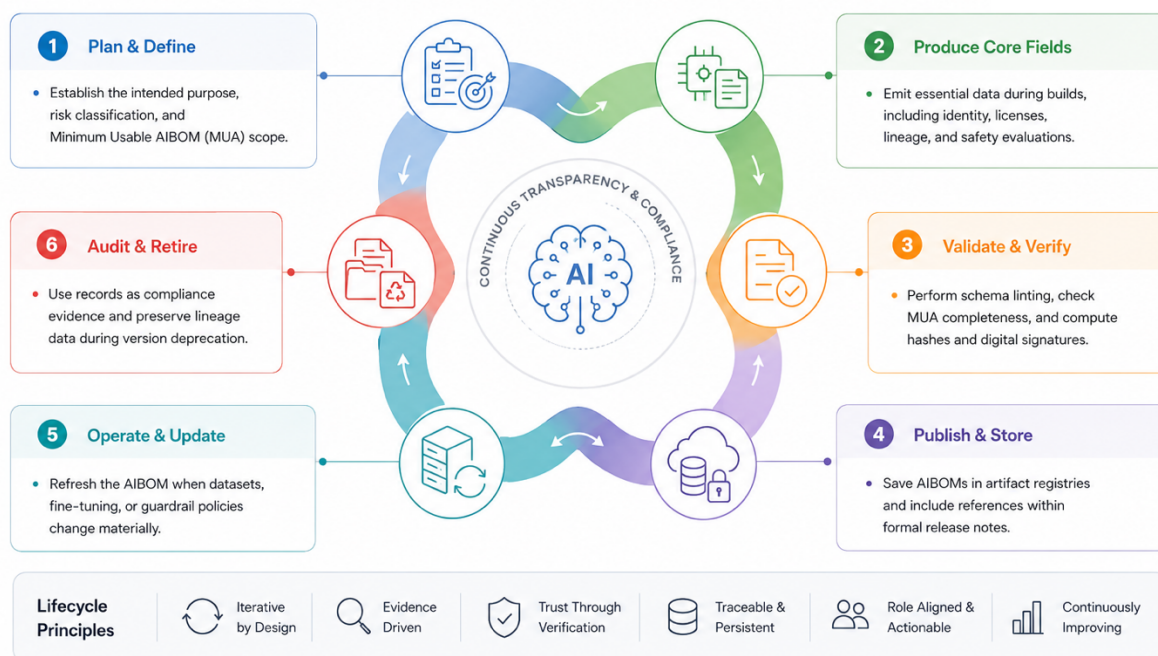
مصنوعات AIBOM به مؤلفه ها و دارایی های سامانه هوش مصنوعی اشاره دارند که شامل مجوزها و تعهدات استفاده برای مدل ها و مجموعه داده ها، منشأ قابل راستی آزمایی و شواهد ارزیابی است. نمونه هایی از مصنوعات عبارت اند از:

- **مدل ها:** چکیده های (digest) مدل پایه، مجوزهای اعلام شده و نتیجه گیری شده، کانال های توزیع، امضاها، مدل های تنظیم دقیق شده، آداپتورهای LoRA/PEFT، نقاط بازرسی
- **مجموعه داده ها:** شناسه/نسخه مجموعه داده، هش، مجوزها، نشانی های منبع، نسخه های لحظه ای، روش های گردآوری، داده های حساس/ناشناس سازی
- **عامل ها و ابزارها:** هماهنگ کننده ها، API ها، افزونه ها، سرورهای MCP
- **پرامپت ها و حفاظ ها:** قالب ها، بسته های سیاست، یادداشت های امنیتی
- **زمان اجرا:** کانتینرها، وابستگی ها، مهرهای زمانی
- **منشأ ساخت:** چکیده های تصاویر ساخته شده، مکان های منبع ورودی، آرگومان های ساخت و مدت زمان ساخت

یکی از اهداف اصلی AIBOM نظام مند کردن یک فرایند تکرار شونده برای مستندسازی مصنوعات مؤلفه های سامانه هوش مصنوعی در سراسر چرخه طراحی، توسعه و استقرار است.

The AIBOM Lifecycle: A Continuous Loop for AI Transparency

A continuous cycle of six iterative stages managing an AI Bill of Materials for compliance, safety, and traceability.



Source: OWASP AIBOM Foundations Guide v1.0

مراحل توسعه AIBOM:

1. برنامه‌ریزی: تعریف هدف مورد نظر، دامنه، رده ریسک و حداقل AIBOM قابل استفاده (MUA).
2. تولید: صدور فیلدهای اصلی در طول آموزش/ساخت، از جمله (و نه محدود به) هویت، مجوزها، منشأ، تبار، ارزیابی و ایمنی.
3. اعتبارسنجی: بررسی فایل‌ها (طرح‌واره)، تأیید کامل بودن MUA، محاسبه هش‌ها، امضاها و بررسی مجوزها.
4. انتشار: ذخیره در کنار مصنوع (مخزن/انبار مصنوعات)، و نگهداشتن ارجاعی به آن در یادداشتهای انتشار.
5. بهره‌برداری: به‌روزرسانی در صورت تغییرات اساسی مانند مجموعه داده‌ها، تنظیم دقیق، مجوزها، و سیاست‌های حفاظ جدید/به‌روز شده.
6. حسابرسی/بازنشستگی: استفاده از AIBOM به عنوان شاهد اصلی، و حفظ تبار هنگام منسوخ کردن نسخه‌ها.

۳. چرا AIBOM ها اهمیت دارند؟

سامانه‌های هوش مصنوعی مدرن به ندرت به طور کامل توسط یک تیم واحد ساخته می‌شوند. یک پشته تولیدی معمول، یک مدل پایه را از یک ارائه‌دهنده می‌گیرد، آن را بر ترکیبی از داده‌های داخلی و دارای مجوز تنظیم دقیق می‌کند، مؤلفه‌های متن‌باز را یکپارچه می‌کند و از طریق خطوط لوله بازیابی به داده‌های زنده متصل می‌شود. هر یک از این لایه‌ها ریسک‌ها، مجوزها و پرسش‌های منشأ خاص خود را دارد.

بدون یک فهرست رسمی، سازمان‌ها اغلب نمی‌توانند به پرسش‌های بنیادی پاسخ دهند: کدام نسخه مدل در محیط تولید مستقر است؟ بر چه داده‌ای آموزش دیده است؟ آیا هیچ‌یک از مؤلفه‌های آن تحت تأثیر یک آسیب‌پذیری تازه افشا شده قرار دارد؟ آیا مجوز آن با استفاده تجاری سازگار است؟ یک AIBOM با عمل‌کردن به عنوان سابقه مرجع و قابل‌پرس‌وجو از آنچه یک سامانه هوش مصنوعی واقعاً در بر دارد، به این پرسش‌ها پاسخ می‌دهد.

باقی این فصل پنج محرک اصلی را بررسی می‌کند که AIBOM را از یک ایده مفید به یک الزام عملیاتی سوق می‌دهند.

محرک‌های امنیتی

سامانه‌های هوش مصنوعی سطوح حمله جدیدی را معرفی می‌کنند که ابزارهای سنتی امنیت برنامه‌ها برای پوشش آن‌ها طراحی نشده‌اند. AIBOM بنیان هر برنامه معتبر امنیت هوش مصنوعی است، زیرا به مدافعان می‌گوید از چه چیزی محافظت می‌کنند.

• **واکنش به آسیب‌پذیری.** وقتی آسیب‌پذیری‌ای در یک مدل، چارچوب یا قالب سریال‌سازی پرکاربرد افشا می‌شود (مانند سوءاستفاده‌های مکرر از pickle که مخازن مدل میزبانی‌شده عمومی را تحت تأثیر قرار می‌دهند)، یک AIBOM به تیم‌های امنیتی اجازه می‌دهد میزان در معرض بودن را در چند ساعت به جای چند هفته شناسایی کنند.

• **مسموم‌سازی مدل و داده.** عاملان مخرب می‌توانند مدل‌ها یا مجموعه داده‌های آموزشی بالادستی را به خطر بیندازند. یک AIBOM منشأ، هش‌ها و اطلاعات امضا را ثبت می‌کند تا دست‌کاری پیش از استقرار قابل‌تشخیص باشد.

• **تشخیص در پشتی و تروجان.** دانستن تبار هر مصنوع مدل، از جمله اینکه چه کسی و بر چه داده‌ای آن را آموزش داده، برای ارزیابی ریسک رفتارهای پنهانی که با ورودی‌های خاص فعال می‌شوند ضروری است.

• **نشت اسرار و پرامپت.** فهرست‌برداری از پرامپت‌ها، دستورالعمل‌های سامانه و اعتبارنامه‌های تعبیه‌شده به جلوگیری از افشای ناخواسته هنگام به اشتراک‌گذاری یا تنظیم دقیق مدل‌ها در پایین‌دست کمک می‌کند.

• **واکنش به حادثه.** وقتی یک سامانه هوش مصنوعی خروجی زیان‌بار یا ناهنجار تولید می‌کند، یک AIBOM به واکنش‌دهندگان اجازه می‌دهد دقیقاً بازسازی کنند که در زمان حادثه چه چیزی، بر چه داده‌ای و با چه پیکربندی‌ای در حال اجرا بوده است.

محرك‌های مقرراتی

چشم‌انداز مقرراتی هوش مصنوعی در زمانی به‌طرز چشمگیری کوتاه از اصول داوطلبانه به تعهدات لازم‌الاجرا تغییر کرده است. AIBOM ها در حال ظهور به‌عنوان سازوکار عملی برای نشان‌دادن انطباق با چندین چارچوب به‌طور هم‌زمان هستند.

- **مقررات ویژه هوش مصنوعی.** مقررات نوظهور هوش مصنوعی^۱ به‌طور فزاینده‌ای سامانه‌های هوش مصنوعی پریسک را ملزم می‌کنند مستندات فنی شامل داده‌های آموزشی، معماری مدل، معیارهای عملکرد و محدودیت‌های شناخته‌شده را نگهداری کنند. یک AIBOM این الزامات را در قالبی استاندارد عملیاتی می‌کند.
- **چارچوب‌های مدیریت ریسک هوش مصنوعی^۲.** کارکردهای اصلی آن‌ها صراحتاً فهرست‌های مستندی از مؤلفه‌ها، داده‌ها و وابستگی‌های هوش مصنوعی را به‌عنوان پیش‌شرط ارزیابی ریسک طلب می‌کنند.
- **تدارکات دولتی و راهنمایی‌های نهادها^۳.** نهادهای بخش عمومی به‌طور فزاینده‌ای هنگام تدارک یا استقرار هوش مصنوعی، مستندات سامانه هوش مصنوعی را الزامی می‌کنند که بازتاب الزامات SBOM معرفی‌شده برای نرم‌افزارهای سنتی است.
- **قواعد ویژه بخش‌ها.** ناظران خدمات مالی، مراجع سلامت و ناظران حفاظت از داده در حال همگرایی به سمت انتظارات افشای مشابهی برای تصمیم‌های مبتنی بر هوش مصنوعی هستند.
- **انتقال‌های فرامرزی.** جایی که قواعد اقامت داده یا حاکمیت داده اعمال می‌شود، یک AIBOM رد حسابرسی لازم برای اثبات اینکه داده‌های آموزشی و مؤلفه‌های مدل الزامات حوزه قضایی را برآورده می‌کنند فراهم می‌کند.

محرك‌های حاکمیتی

- حاکمیت داخلی در طول تاریخ از پذیرش هوش مصنوعی عقب مانده است. بسیاری از سازمان‌ها، اغلب در جریان یک حسابرسی یا حادثه، درمی‌یابند که هیچ‌کس نمی‌تواند به‌طور قطعی بگوید کدام مدل‌ها در محیط تولیدند، مالک آن‌ها کیست، یا چگونه اعتبارسنجی شده‌اند. AIBOM ها به کارکردهای حاکمیتی سندی مشخص می‌دهند تا سیاست‌ها، بازبینی‌ها و تأییدها را بر آن استوار کنند.
- **مدیریت چرخه عمر مدل.** یک AIBOM هر نسخه، تنظیم دقیق و استقرار یک مدل را ردیابی می‌کند و سوابق قابل‌حسابرسی را جایگزین دانش شفاهی و غیرمکتوب می‌کند.
 - **بازبینی سوگیری، انصاف و اخلاق.** بازبینان بدون دانستن اینکه چه داده‌ای یک مدل را آموزش داده، نمی‌توانند ارزیابی کنند که آیا آن مدل برای استخدام، وام‌دهی یا سلامت مناسب است. AIBOM ورودی این ارزیابی‌هاست.

۱ قانون هوش مصنوعی اتحادیه اروپا، آیین‌نامه (EU) 2024/1689 (ر.ک. ماده ۱۱ و پیوست IV): eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

۲ NIST AI RMF 1.0 (NIST AI 100-1): nist.gov/publications/artificial-intelligence-risk-management-framework-ai-100-1

۳ NTIA، حداقل عناصر برای صورت‌حساب مواد نرم‌افزار (۲۰۲۱): ntia.gov/report/2021/minimum-elements-software-bill-2021؛ و به‌روزرسانی CISA در ۲۰۲۵: cisa.gov/.../2025_CISA_SBOM_Minimum_Elements.pdf

- **انطباق با مجوزها و مالکیت فکری.** مدل‌ها و مجموعه داده‌ها مجوزهای متنوعی دارند؛ برخی آزادانه، برخی محدودکننده، و برخی با پرسش‌های حل‌نشده درباره حق نشر داده‌های آموزشی. آشکارکردن زود هنگام این تعهدات از اصلاحات پرهزینه بعدی جلوگیری می‌کند.
- **ریسک اشخاص ثالث و فروشندگان.** وقتی قابلیت‌های هوش مصنوعی از فروشندگان تأمین می‌شوند، AIBOM به‌طور فزاینده‌ای در کنار گزارش‌های SOC 2 و پرسش‌نامه‌های امنیتی بخشی از بررسی دقیق (due diligence) است.
- **دیدپذیری در سطح هیئت‌مدیره.** داده‌های تجمیع‌شده AIBOM به مدیران اجرایی و هیئت‌مدیره اجازه می‌دهد ردپای هوش مصنوعی سازمان را در یک نمای واحد ببینند و نه به‌صورت مصنوعات پراکنده پروژه‌ها.

تداوم کسب‌وکار

- سامانه‌های هوش مصنوعی اغلب بی‌سروصدا حیاتی می‌شوند. یک موتور توصیه‌گر، یک مدل کشف تقلب یا یک پردازشگر اسناد می‌تواند در عرض چند ماه از مرحله آزمایشی به زیرساختی باربر تبدیل شود. وقتی چیزی خراب می‌شود، هزینه نداشتن AIBOM با زمان ازکارافتادگی، از دست رفتن درآمد و آسیب به اعتبار سنجیده می‌شود.
- **تکرارپذیری.** اگر یک مدل تولیدی از دست برود، خراب شود یا نیاز به آموزش مجدد داشته باشد، یک AIBOM ورودی‌ها، پارامترها و وابستگی‌های لازم برای بازسازی پیش‌بینی‌پذیر آن را ثبت کرده است.
- **شکست یا منسوخ شدن فروشنده.** وقتی یک ارائه‌دهنده مدل سرویسی را تعطیل می‌کند، قیمت‌گذاری را تغییر می‌دهد یا مدلی را منسوخ می‌کند، یک AIBOM به شما می‌گوید کدام سامانه‌ها تحت تأثیرند و چه جایگزین‌هایی سازگارند.
- **رانش و افت کیفیت.** ردیابی نسخه‌های مدل و داده در طول زمان امکان می‌دهد پسرقت‌های کیفی را به تغییرات مشخص نسبت دهیم، به‌جای آنکه حدس بزنیم.
- **توقیف‌های مقرراتی و حسابرسی‌ها.** وقتی ناظران یا دادگاه‌ها شواهدی از آنچه یک سامانه در تاریخی مشخص انجام می‌داده طلب می‌کنند، یک AIBOM سابقه تاریخی را فراهم می‌کند.
- **برنامه‌ریزی بازیابی از فاجعه.** اهداف زمان بازیابی و نقطه بازیابی برای سامانه‌های هوش مصنوعی تنها زمانی معنا دارند که مؤلفه‌های لازم برای بازگرداندن آن‌ها فهرست‌برداری شده و در دسترس باشند.

تهدیدهای زنجیره تأمین

- زنجیره تأمین هوش مصنوعی گسترده‌تر، عمیق‌تر و نابالغ‌تر از زنجیره تأمین سنتی نرم‌افزار است. مؤلفه‌ها از هاب‌های عمومی مدل، مخازن مجموعه داده و API‌های استنتاج دریافت می‌شوند که با

سطوح متفاوتی از بهداشت امنیتی اداره می‌شوند. AIBOM ها زنجیره تأمین را به اندازه کافی دیدپذیر می‌کنند تا بتوان از آن دفاع کرد.

- **مخازن عمومی مدل به‌خطرافتاده.** هاب‌های عمومی بارها میزبان مدل‌هایی بوده‌اند که در قالب‌های سریال‌سازی خود کد مخرب داشته‌اند. یک AIBOM با هش‌های راستی‌آزمایی‌شده این رده حمله را قابل تشخیص می‌کند.

- **تایپواسکوئیتینگ و جعل هویت.** مهاجمان مدل‌ها یا بسته‌هایی با نام مشابه منتشر می‌کنند که طوری طراحی شده‌اند که به‌طور تصادفی دریافت شوند. ردیابی منشأ در یک AIBOM دفاع اصلی است.

- **آلودگی مجموعه داده.** داده‌های آموزش یا تنظیم دقیق که از وب عمومی تأمین می‌شوند ممکن است عمداً با محتوای مسموم آلوده شوند. فهرست کردن منابع مجموعه داده و مراحل پردازش نقطه آغاز تشخیص است.

- **وابستگی‌های گذرا.** هر مدل وابستگی‌های بالادستی دارد: به توکن‌سازها، وزن‌های پایه و کتابخانه‌های پشتیبان. AIBOM ها این روابط گذرا را صریح می‌کنند.

- **ریسک خودی و توسعه دهنده.** خطوط لوله داخلی تنظیم دقیق و برچسب‌گذاری داده می‌توانند از طریق مشارکت‌های بررسی‌نشده ریسک وارد کنند. کنترل‌های مبتنی بر AIBOM با مصنوعات داخلی با همان دقتی رفتار می‌کنند که با مصنوعات خارجی.

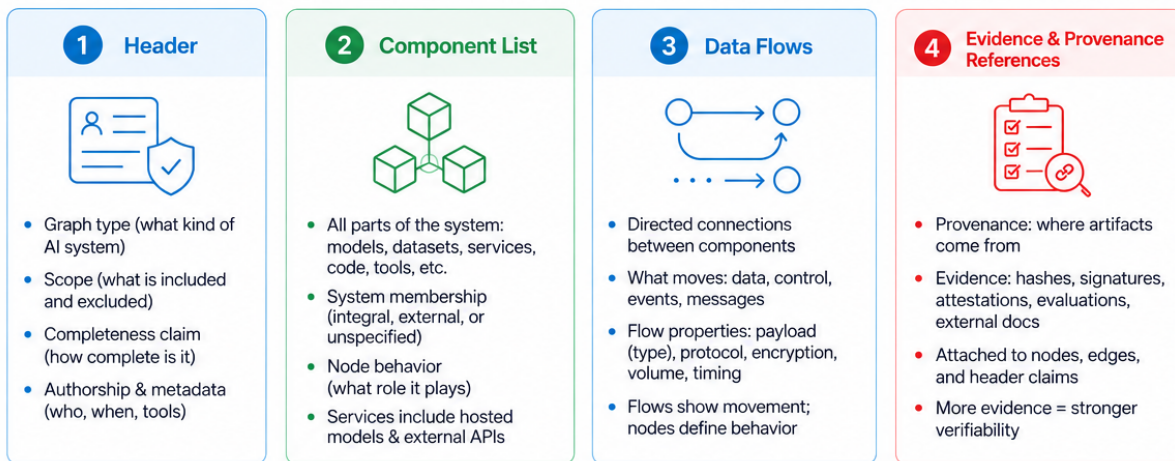
۴. مؤلفه‌های اصلی یک AIBOM

فصل‌های پیشین توضیح دادند AIBOM چیست و چرا اهمیت دارد. این فصل بخش‌های یک AIBOM و کارکرد هر بخش را معرفی می‌کند تا خواننده بداند چه چیزی در آن جای می‌گیرد و چرا.

از نظر ساختاری، یک AIBOM چهار بخش دارد: **یک سرآیند، یک فهرست مؤلفه‌ها، جریان‌های داده جهت‌دار، و ارجاعات شواهد و منشأ.** هر AIBOM هر چهار بخش را دارد. دو نوع محتوای اختیاری، هر جا که سامانه اقتضا کند، تصویر را کامل می‌کنند: بستر اعتماد، که مؤلفه‌ها را در مناطقی گروه‌بندی کرده و مرزهای میان آن‌ها را مشخص می‌کند، و رفتارها، که کنش شرطی را برای سامانه‌هایی توصیف می‌کنند که جابه‌جایی داده در آن‌ها به تصمیم‌های زمان اجرا بستگی دارد. این فصل هر یک را به نوبت بررسی می‌کند، هر کدام را به هدفی که به آن خدمت می‌کند پیوند می‌دهد، و با آنچه یک AIBOM کامل امکان‌پذیر می‌کند به پایان می‌رسد.

Core Components of an AIBOM

Every AIBOM has four parts. Two optional kinds of content complete the picture.



سرآیند

سرآیند AIBOM برای توصیف اینکه چه نوع سامانه هوش مصنوعی‌ای مدل‌سازی می‌شود و چه ادعاهایی درباره دامن و کامل‌بودن می‌توان مطرح کرد به کار می‌رود. این اطلاعات بستری برای خواندن بقیه سند فراهم می‌کند. AIBOM یک تولیدکننده ممکن است تصویری کامل از یک سامانه مستقرشده واحد باشد که به‌طور خودکار از داده‌های دورسنجی ساخت تولید شده است. AIBOM دیگری ممکن است طرحی ناقص از یک سامانه آینده باشد که در جریان بازبینی تدارکات دستی نوشته شده است. بدون راهی برای تمایز این دو، خواننده نمی‌تواند از هیچ‌یک استنتاج درستی داشته باشد.

بنابراین هر AIBOM با سه خودتوصیف آغاز می‌شود: (۱) **نوع گراف** می‌گوید AIBOM چه نوع سامانه‌ای را به تصویر می‌کشد؛ مانند یک مدل واحد، یک سامانه زمان اجرا، یک افزونه سامانه مانند یک سرور MCP، یا یک خط لوله آموزش. (۲) **دامنه** می‌گوید AIBOM در تلاش است چه چیزی را پوشش دهد:

کدام سامانه، کدام نسخه، کدام استقرار، کدام رابط‌ها. (۳) ادعای کامل بودن می‌گوید تولیدکننده چقدر مطمئن است که AIBOM آنچه را دامنه‌اش می‌گوید پوشش می‌دهد، که از «کامل برای دامنه اعلام شده» تا پوشش ناقص و تا «نمی‌دانیم چه چیز دیگری در این جا هست» متغیر است. در کنار این خودتوصیف‌ها، فراداده‌های استاندارد سند ثبت می‌کنند چه کسی AIBOM را تألیف، بازبینی و تأیید کرده، چه زمانی و با چه ابزارهایی. همین فراداده است که خواننده به کمک آن سابقه‌ای را که به‌طور خودکار تولید شده از طرحی که دستی نوشته شده تشخیص می‌دهد.

ادعای کامل بودن تعیین می‌کند خواننده از غیاب اطلاعات به‌طور معناداری چه نتیجه‌ای می‌تواند بگیرد. اگر یک AIBOM مجموعه داده خاصی را فهرست نکند و ادعای کامل بودن نداشته باشد، این غیاب هیچ چیزی را اثبات نمی‌کند. ممکن است آن مجموعه داده در حال استفاده باشد و صرفاً فهرست نشده باشد. اما اگر AIBOM ادعا کند برای دامنه اعلام شده کامل است، این غیاب معنادار می‌شود: اکنون ادعایی از سوی تولیدکننده است که هیچ مجموعه داده‌ای از این دست در دامنه مشارکت ندارد.

همین عدم تقارن میان استنتاج مثبت و منفی است که ایجاب می‌کند خودتوصیفی گراف پیش از محتوای گراف بیاید. بدون دامنه و کامل بودن، دو خواننده می‌توانند به یک AIBOM نگاه کنند و به نتیجه‌گیری‌هایی متضاد درباره آنچه می‌گوید برسند. بخش‌های باقی‌مانده این فصل فرض می‌کنند که هر AIBOM هر سه خودتوصیف را دارد، حتی جایی که متن آن‌ها را تکرار نمی‌کند.

فهرست مؤلفه‌ها

یک سامانه هوش مصنوعی از شماری مؤلفه به هم مرتبط تشکیل شده است. این‌ها همان «مواد» در صورت حساب مواد هوش مصنوعی‌اند. در یک AIBOM، همه چیز در سامانه یک مؤلفه است، از جمله سرویس‌ها: یک مازول کد مجزا، یک ریزسرویس، یک مجموعه داده و یک نقطه پایانی API میزبانی شده توسط ارائه‌دهنده همگی مؤلفه‌اند که با نوع مؤلفه از هم متمایز می‌شوند و نه با فهرست‌های جداگانه. فهرست مؤلفه‌ها به نخستین پرسشی که هر بازبین درباره یک سامانه مطرح می‌کند پاسخ می‌دهد: به چه چیزی وابسته‌ام؟

دو چیز یک گره را فراتر از هویتش توصیف می‌کند: عضویت در سامانه و رفتار گره. عضویت در سامانه می‌پرسد آیا یک گره بخش جدایی‌ناپذیر سامانه هوش مصنوعی توصیف شده به شمار می‌آید، بیرون از آن است، یا نامشخص رها شده است. این یک گزاره وابسته به دامنه است و به صورت ساختاری ثبت می‌شود: دامنه اعلام شده AIBOM فهرست می‌کند کدام مؤلفه‌ها در سامانه توصیف شده گنجانده شده‌اند و کدام از آن مستثنا شده‌اند. این با مالکیت یا کنترل یکی نیست. یک مدل میزبانی شده شخص ثالث می‌تواند بخش جدایی‌ناپذیر سامانه هوش مصنوعی توصیف شده باشد، و یک پایگاه داده داخلی همچنان می‌تواند بیرون از مرز سامانه هوش مصنوعی باشد، اگر خارج از دامنه اعلام شده قرار گیرد. مؤلفه‌ای که در دامنه اعلام شده گنجانده شده، جزء سازنده سامانه توصیف شده است.

در مقابل، رفتار گره می‌پرسد گره چه نقشی در گراف ایفا می‌کند. برخی گره‌ها کار انجام می‌دهند. یک پردازشگر پرامپت متن را تبدیل می‌کند. یک مدل خروجی تولید می‌کند. یک عامل تصمیم می‌گیرد بعدی چه چیزی را فراخواند. یک حفاظ محتوا را پالایش می‌کند. یک سرویس بازیابی نامزدها را رتبه‌بندی می‌کند. گره‌های دیگر عمدتاً اطلاعات را ذخیره، تأمین یا دریافت می‌کنند؛ مانند مجموعه داده‌ها، نمایه‌ها، لاگ‌ها و انبارهای داده. گره‌ای که مستقیماً اطلاعات را پردازش نمی‌کند ممکن است به‌طور

اطلاع‌رسان به‌عنوان مؤلفه «منفعل» حاشیه‌نویسی شود، یا این وضعیت از الگوی جریان‌های داده‌ای که به آن متصل‌اند استنتاج شود.

مدل‌های میزبانی‌شده برای بسیاری از AIBOM‌ها از مهم‌ترین مؤلفه‌های نوع سرویس‌اند، اما انواع بسیار دیگری نیز وجود دارد. یک مؤلفه سرویس ممکن است به یک مدل یا برنامه هوش مصنوعی اجازه دهد از طریق یک API شبکه به داده‌ها، ابزارها یا سامانه‌ها دسترسی پیدا کند. AIBOM این مؤلفه‌ها را ثبت می‌کند تا خواننده بتواند ببیند کدام بخش‌های سامانه داده مبادله می‌کنند و مرزهای سامانه کجا عبور می‌شوند.

جریان‌های داده

جریان‌های داده در یک AIBOM مؤلفه‌ها را به هم متصل می‌کنند. اگر مؤلفه‌ها «فهرست قطعات» باشند، جریان‌های داده «نقشه ساخت» یا «نقشه سیم‌کشی»‌اند که نشان می‌دهند این قطعات چگونه در کنار هم قرار می‌گیرند. جریان‌های داده همیشه جهت‌دارند: حرکت اطلاعات از یک مؤلفه به مؤلفه دیگر را نشان می‌دهند. وقتی اطلاعات در هر دو جهت میان دو مؤلفه حرکت می‌کند، مانند یک درخواست و پاسخ آن، هر جابه‌جایی اساسی جریانی جداگانه است.

اگر فهرست مؤلفه‌ها به این پرسش پاسخ دهد که به چه چیزی وابسته‌ام، جریان‌های داده به پرسشی پاسخ می‌دهند که سامانه‌های هوش مصنوعی را متفاوت می‌کند: چه چیزی به چه چیزی می‌رسد؟ یک مدل واحد، بسته به اینکه آیا داده‌های مشتری بدون پالایش به آن می‌رسند و خروجی آن به کجا می‌رود، نیمرخ ریسک کاملاً متفاوتی دارد. یک فهرست قطعات نمی‌تواند این تفاوت را بیان کند. در سامانه‌های هوش مصنوعی، ریسک در اتصال‌ها زندگی می‌کند.

هر جریان می‌تواند اطلاعاتی درباره آنچه در امتداد آن حرکت می‌کند داشته باشد. یک جریان نوع خود را اعلام می‌کند (جریان داده، جریان کنترل، رویداد، پیام) و می‌تواند محموله خود را توصیف کند؛ یا با ارجاع به اشیای داده‌ای که در جای دیگری از AIBOM توصیف شده‌اند، یا با یک توصیفگر درون‌خطی، وقتی محموله آن جریان خاص با هر چیزی که در جای دیگر توصیف شده متفاوت است. یک جریان همچنین می‌تواند ثبت کند جابه‌جایی چگونه رخ می‌دهد: آیا رمزگذاری شده است، با چه پروتکلی، با چه حجمی و با چه زمان‌بندی‌ای.

آنچه یک جریان در خود ندارد، توصیف کاری است که مؤلفه دریافت‌کننده با داده انجام می‌دهد. مدلی که خروجی تولید می‌کند، سرویس بازیابی‌ای که نامزدها را رتبه‌بندی می‌کند، حفاظی که محتوا را پالایش می‌کند، یا پایگاه داده‌ای که آنچه را دریافت می‌کند ماندگار می‌سازد، رفتار گره است: این رفتار روی گره، از طریق نوع مؤلفه و فراداده‌های آن توصیف می‌شود، نه به‌صورت حاشیه‌نویسی بر جریان. یک جریان ثبت می‌کند که داده از جایی به جای دیگر حرکت کرده است؛ گره‌ها توضیح می‌دهند که در هر سر چه اتفاقی می‌افتد.

ارجاعات شواهد و منشأ

سرآیند، مؤلفه‌ها و سرویس‌ها، و یال‌های جریان داده می‌گویند سامانه چه چیزی در بر دارد و اطلاعات چگونه در آن حرکت می‌کند. ارجاعات شواهد و منشأ می‌گویند چرا خواننده باید این ادعاها را باور کند و

کجا می‌توان آن‌ها را بررسی کرد. این ارجاعات یک گراف جداگانه نیستند؛ پیوست‌هایی به گره‌ها، یال‌ها و خودتوصیف‌های مشخص‌اند.

دو نوع پیوست بیشترین اهمیت را دارند. منشأ ثبت می‌کند یک مصنوع چگونه به وجود آمده است: یک مدل تنظیم دقیق‌شده از کدام مدل پایه مشتق شده، کدام اجرای آموزش آن را تولید کرده، کدام نسخه لحظه‌ای مجموعه داده یک نمایه بازبایی را تغذیه کرده است. شاهد پشتوانه یک ادعا را ثبت می‌کند: یک هش رمزنگاری که یکپارچگی یک مؤلفه را تثبیت می‌کند، یک امضا یا گواهی (attestation) که یک مصنوع را به ساختی که آن را تولید کرده پیوند می‌دهد، یک نتیجه ارزیابی، یا یک سند خارجی مانند کارت مدل یا فایل مجوز.

از آنجا که این ارجاعات به ادعاهای منفرد محدودند، مستقیماً با ادعای کامل بودن در سرآیند تعامل دارند. مؤلفه‌ای که بدون هیچ ارجاع شاهدهی ادعا شده، ادعایی صرف است؛ همان مؤلفه همراه با یک هش، یک امضا و پیوندی به کارت مدلش، ادعایی است که مصرف‌کننده می‌تواند مستقلاً راستی‌آزمایی کند. یک AIBOM لازم نیست به هر ادعا شاهد پیوست کند، اما حضور یا غیاب شاهد خود اطلاعاتی است که خواننده برای تصمیم‌گیری درباره اینکه یک گره یا یال چه وزنی را می‌تواند تحمل کند از آن استفاده می‌کند.

عناصری که یک تولیدکننده در برابر این چهار بخش ثبت می‌کند (فراداده‌های مدل، منشأ مجموعه داده، محیط آموزش، وابستگی‌ها و کتابخانه‌ها، API‌ها و زنجیره ابزارها، پیکربندی زمان اجرا، و توصیف‌گرهای عامل و زیرعامل برای سامانه‌های عامل محور) فهرست جداگانه‌ای برای نگهداری نیستند. هر یک یا یک گره است، یا ویژگی یک گره، یا یک یال، یا یک ارجاع شاهد که به یکی از آن‌ها پیوست شده است.

مناطق اعتماد و مرزها

همه مؤلفه‌ها در یک مکان یا تحت یک کنترل قرار ندارند. یک AIBOM می‌تواند مؤلفه‌ها را در مناطقی با سطوح اعتماد متفاوت گروه‌بندی کند: زیرساخت خود تیم، شبکه گسترده‌تر سازمان، ابر یک ارائه‌دهنده، اینترنت عمومی. جایی که یک جریان از یک منطقه به منطقه دیگر عبور می‌کند یک مرز است، و مرزها جایی هستند که توجه بازبینی متمرکز می‌شود. یک عبور جایی است که داده دست‌به‌دست می‌شود، جایی که کنترل‌هایی مانند احراز هویت و رمزگذاری اعمال می‌شوند، و جایی که دیدپذیری تولیدکننده پایان می‌یابد.

مناطق برای صریح کردن تفاوتی وجود دارند که یک گراف خالی آن را پنهان می‌کند. جریانی میان دو مؤلفه داخلی و جریانی که همان محموله را به یک API شخص ثالث بیرون می‌برد، یال‌هایی مشابه به نظر می‌رسند، اما واقعیت‌هایی بسیار متفاوت‌اند. بدون مناطق اعلام‌شده، تمایز آن‌ها به دانش پیش‌زمینه خواننده درباره اینکه کدام نام‌ها داخلی‌اند بستگی دارد. با وجود آن‌ها، خود AIBOM می‌گوید کدام جابه‌جایی‌ها در خانه می‌مانند و کدام خارج می‌شوند؛ و هر خروج در مسیر بیرون رفتن ملزم به عبور از چه چیزی است.

رفتارها

جریان‌ها می‌گویند چه چیزی به چه چیزی می‌رسد. رفتارها می‌گویند چه زمانی و تحت چه شرایطی. برای یک خط لوله خطی، جریان‌ها معمولاً کافی‌اند: همان جابه‌جایی‌ها برای هر درخواست رخ می‌دهند.

سامانه‌های عامل‌محور متفاوت‌اند. واقعیت جالب دربارهٔ یک عامل اغلب این است که در زمان اجرا تصمیم می‌گیرد اصلاً اقدام کند یا نه؛ اینکه ابزاری را فراخواند، از پایگاه داده‌ای پرس‌وجو کند، یا کار را به عامل دیگری بسپارد. یک جریان به‌تنهایی یا در آنچه رخ می‌دهد اغراق می‌کند (ابزار هر بار فراخوانده می‌شود) یا آن را کمتر از واقع نشان می‌دهد (اتصال ثبت نمی‌شود). توصیف رفتار بخش شرطی را ثبت می‌کند: چه چیزی یک کنش را برمی‌انگیزد، مؤلفه در پاسخ چه کاری ممکن است انجام دهد، و چه چیزی از آن پیامد می‌شود. بیشتر AIBOM ها به توصیف رفتار نیاز نخواهند داشت؛ سامانه‌هایی که ریسکشان به تصمیم‌های زمان اجرا بستگی دارد نیاز خواهند داشت. فصل سامانه‌های هوش مصنوعی عامل‌محور به این موضوع بازمی‌گردد.

آنچه یک AIBOM امکان‌پذیر می‌کند

همهٔ آنچه این فصل توصیف کرده واقعیت ساخت‌یافته است، و هر بخش از آن (هر مؤلفه، جریان، منطقه و مرز) یک شناسهٔ پایدار دارد. همین است که AIBOM را از یک نمودار معماری جدا می‌کند: مصنوعات دیگر می‌توانند دقیقاً به یک چیز در آن اشاره کنند.

این قابلیت ارجاع‌پذیری همان چیزی است که بقیهٔ پشتهٔ حاکمیت بر آن بنا می‌شود. مدل‌های تهدید، ثبت‌های ریسک، نگاشت‌های کنترل و بیانیه‌های آسیب‌پذیری سامانه را دوباره توصیف نمی‌کنند؛ به گراف AIBOM ارجاع می‌دهند. یک تهدید می‌تواند بگوید که یک مسیر حمله از مرز مشخصی عبور می‌کند. یک کنترل می‌تواند بگوید از کدام جریان محافظت می‌کند. یک تحلیل آسیب‌پذیری می‌تواند بگوید که یک ضعف دسترس‌ناپذیر است زیرا کنترل‌های یک منطقه مسیر رسیدن به آن را مسدود می‌کنند. گراف را یک بار بسازید، و تحلیل‌ها به آن متصل می‌شوند؛ و با تحول سامانه متصل و قابل‌بررسی باقی می‌مانند.

خود AIBOM در سمت خود از این خط باقی می‌ماند. آنچه سامانه هست را ثبت می‌کند، نه اینکه آیا پذیرفتنی است. انطباق، پذیرفتنی‌بودن ریسک و کفایت کنترل داوریهایی هستند که خواننده با استفاده از چارچوب سیاستی خود، بر پایهٔ واقعیت‌هایی که AIBOM فراهم می‌کند، مطرح می‌کند.

۵. ساختار AIBOM

فصل «مؤلفه‌های اصلی یک AIBOM» چهار بخش یک AIBOM (سرآیند، مؤلفه‌ها و سرویس‌ها، یال‌های جریان داده، و ارجاعات شواهد و منشأ) را به صورت یک مدل انتزاعی تشریح کرد. این فصل آن مدل را ملموس می‌کند. مثال‌هایی از یک سامانه استنتاج میزبانی‌شده، یک خط لوله تنظیم دقیق، و یک اجرای سرتاسری که برای یک مدل منتشرشده AIBOM تولید می‌کند را بررسی می‌کند؛ به روشنی بیان می‌کند یک AIBOM چه چیزی را ادعا می‌کند و چه چیزی را ادعا نمی‌کند؛ و ردیف‌ها و لایه‌های کدگذاری‌ای را معرفی می‌کند که تولیدکننده هنگام نشستن برای صدور واقعی یک AIBOM به آن‌ها نیاز دارد.

خودتوصیف‌های سرآیند (نوع گراف، روش تولید، دامنه و کامل بودن) در سراسر این فصل پیش‌فرض گرفته شده‌اند و در این‌جا دوباره استخراج نمی‌شوند؛ ر.ک. «سرآیند» در «مؤلفه‌های اصلی یک AIBOM». بدون آن‌ها هیچ چیز دیگری در AIBOM تفسیرپذیر نیست، بنابراین هر مثال زیر چنان نوشته شده که گویی سرآیندش از پیش در جای خود قرار دارد، حتی جایی که متن آن را تکرار نمی‌کند.

آنچه هر AIBOM اعلام می‌کند

پیش از آنکه کسی بتواند یک AIBOM را بخواند، باید بداند با چه نوع AIBOMی سروکار دارد.

واژه‌های یکسان می‌توانند مصنوعات بسیار متفاوت را توصیف کنند. AIBOM یک تولیدکننده ممکن است تصویری کامل از یک سامانه مستقرشده واحد باشد که به طور خودکار از داده‌های دورسنجی ساخت تولید شده است. AIBOM دیگری ممکن است طرحی ناقص از یک سامانه آینده باشد که در جریان بازبینی تدارکات دستی نوشته شده است. بدون راهی برای تمایز این دو، خواننده نمی‌تواند از هیچ‌یک استنتاج درستی داشته باشد.

بنابراین هر AIBOM با سه خودتوصیف آغاز می‌شود:

نوع گراف می‌گوید AIBOM چه چیزی را به تصویر می‌کشد: یک مدل واحد، یک مجموعه داده واحد، یک سامانه زمان اجرا، یک خط لوله آموزش، یا مصنوع خوش‌تعریف دیگری. به خواننده می‌گوید انتظار چه نوع گره‌ها و جریان‌هایی را داشته باشد.

دامنه می‌گوید AIBOM در تلاش است چه چیزی را پوشش دهد: کدام سامانه، کدام نسخه، کدام استقرار، کدام رابط‌ها. دامنه اهمیت دارد زیرا هر ادعای دیگری در AIBOM نسبت به آن سنجیده می‌شود.

ادعای کامل بودن می‌گوید تولیدکننده چقدر مطمئن است که AIBOM آنچه را دامنه‌اش می‌گوید پوشش می‌دهد. کامل بودن از «کامل برای دامنه اعلام‌شده» تا پوشش ناقص و تا «نمی‌دانیم چه چیز دیگری در این‌جا هست» متغیر است.

در کنار این خودتوصیف‌ها، فراداده‌های سند AIBOM ثبت می‌کنند چه کسی آن را تألیف، بازبینی و تأیید کرده، چه زمانی، با چه ابزارهایی و برای چه دوره اعتباری. همین فراداده است که خواننده به کمک آن AIBOMی را که در جریان بازبینی تدارکات دستی تألیف شده از AIBOMی که از پیکربندی ساخت تولید

شده یا از داده‌های دورسنجی زمان اجرا گرفته شده تشخیص می‌دهد؛ و همین‌جاست که ادعاهای تولیدکننده یک نویسنده پاسخ‌گو پیدا می‌کنند.

ادعای کامل‌بودن کاری بیش از آنچه به نظر می‌رسد انجام می‌دهد. تعیین می‌کند خواننده اجازه دارد از غیاب اطلاعات چه نتیجه‌ای بگیرد. اگر یک AIBOM مجموعه‌داده خاصی را فهرست نکند و ادعای کامل‌بودن نداشته باشد، این غیاب هیچ چیزی را اثبات نمی‌کند؛ ممکن است آن مجموعه‌داده در حال استفاده باشد و صرفاً فهرست نشده باشد. اگر AIBOM ادعا کند برای دامنه اعلام‌شده کامل است، این غیاب معنادار می‌شود؛ اکنون ادعایی از سوی تولیدکننده است که هیچ مجموعه‌داده‌ای از این دست در دامنه مشارکت ندارد.

همین عدم‌تقارن میان استنتاج مثبت و منفی است که ایجاب می‌کند خودتوصیفی گراف پیش از محتوای گراف بیاید. بدون دامنه و کامل‌بودن، دو خواننده می‌توانند به یک AIBOM نگاه کنند و به نتیجه‌گیری‌هایی متضاد درباره آنچه می‌گویند برسند.

بخش‌های باقی‌مانده این فصل فرض می‌کنند که هر AIBOM این سه خودتوصیف را دارد، حتی جایی که متن آن‌ها را تکرار نمی‌کند.

مثال کاربردی: یک سامانه استنتاج میزبانی‌شده

یک دستیار پشتیبانی مشتری سازمانی را در نظر بگیرید. مشتری پرسشی را در یک رابط وب تایپ می‌کند. برنامه یک پرامپت سرهم می‌کند، مستندات داخلی مرتبط را بازیابی می‌کند، ورودی را از نظر محتوای ناایمن بررسی می‌کند، پرامپت را به یک مدل زبانی بزرگ میزبانی‌شده می‌فرستد، پاسخ را پس‌پردازش می‌کند و آن را به مشتری بازمی‌گرداند.

AIBOM این سامانه گره‌های زیر را دارد:

یک **سرویس کاربرمحور** پرسش مشتری را دریافت می‌کند. این یک مؤلفه نوع سرویس است که توسط تیم دستیار اداره می‌شود.

یک **پردازشگر پرامپت** پرامپت نهایی را سرهم می‌کند. قالب‌ها را اعمال می‌کند، اطلاعات شخصی را حذف (redact) می‌کند، بستر نشست را مدیریت می‌کند و تصمیم می‌گیرد چه چیزی بازیابی شود. تیم مالک آن به‌عنوان یک مؤلفه است.

یک **سرویس بازیابی بردارهای تعبیه (embedding)** مستندات پشتیبانی داخلی را نگه می‌دارد و قطعه‌های مرتبط را بازمی‌گرداند. تیم آن را روی زیرساخت داخلی اداره می‌کند. پیکره زیربنایی، یعنی نسخه لحظه‌ای از پایگاه دانش پشتیبانی داخلی در تاریخی مشخص، به‌عنوان یک مؤلفه مجموعه‌داده ارجاع داده می‌شود.

یک **سرویس تعدیل محتوا** ورودی مشتری را از نظر محتوایی که دستیار باید از درگیرشدن با آن امتناع کند بررسی می‌کند. در این مثال، تیم مشترک یک API تعدیل محتوای شخص ثالث است.

یک **مدل زبانی میزبانی‌شده** پاسخ دستیار را تولید می‌کند. تیم از طریق API ارائه‌دهنده به آن دسترسی دارد. این جالب‌ترین گره در گراف است و متن در ادامه به آن بازمی‌گردد.

یک مدیر پاسخ خروجی مدل را پس‌پردازش می‌کند: پاسخ را قالب‌بندی می‌کند، تبادله را برای بازبینی کیفیت ثبت می‌کند و متن نهایی را به سرویس کاربرمحور بازمی‌گرداند.

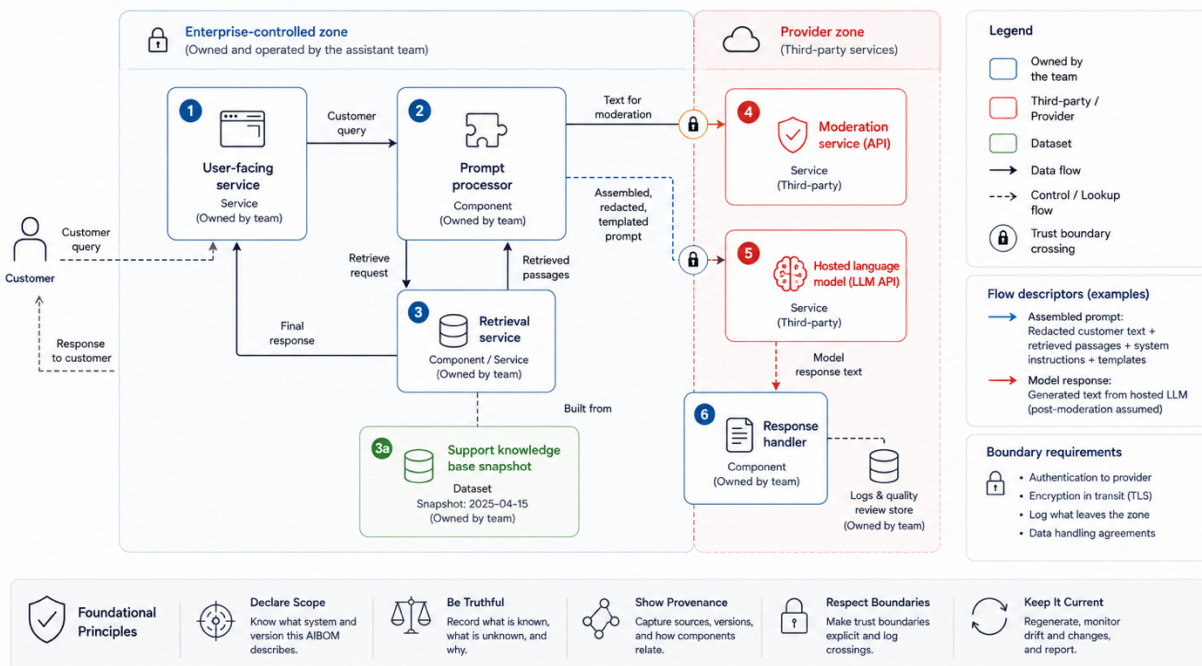
AIBOM این گره‌ها را با جریان‌های داده‌ی جهت‌دار به هم وصل می‌کند. پرسش مشتری از سرویس کاربرمحور به پردازشگر پرامپت می‌رود. پردازشگر متن را به سرویس تعدیل محتوا می‌فرستد، قطعه‌ها را از سرویس بازیابی دریافت می‌کند و پرامپت سرهم‌شده را به مدل زبانی میزبانی‌شده می‌فرستد. مدل متن را به مدیر پاسخ بازمی‌گرداند. مدیر پاسخ متن را به سرویس کاربرمحور بازمی‌گرداند و آن سرویس آن را به مشتری برمی‌گرداند.

هر جریان اطلاعاتی درباره‌ی آنچه در امتداد آن حرکت کرده در خود دارد. پرامپتی که به مدل میزبانی‌شده فرستاده می‌شود پرسش خام مشتری نیست؛ در برابر آن بازیابی انجام شده، حذف اطلاعات صورت گرفته و قالب‌بندی شده است. جریانی که آن را حمل می‌کند می‌تواند توصیفی درون‌خطی از محموله‌ای که در آن جریان خاص جابه‌جا شده داشته باشد، به‌جای آنکه به توصیفی کلی از خروجی پردازشگر پرامپت تکیه کند. توصیفگرهای هر جریان زمانی اهمیت دارند که محموله در یک مرحله از سامانه به‌طور اساسی با محموله در مرحله‌ی دیگر متفاوت باشد.

این مثال بستر اعتماد را نیز در بر دارد. سرویس کاربرمحور، پردازشگر پرامپت، سرویس بازیابی و مدیر پاسخ در منطقه‌ای قرار دارند که تیم کنترل می‌کند. API تعدیل محتوا و مدل زبانی میزبانی‌شده در مناطق ارائه‌دهنده قرار دارند. دو جریانی که منطقه‌ی تیم را ترک می‌کنند (بررسی تعدیل محتوا و پرامپت سرهم‌شده) از مرزهای اعلام‌شده عبور می‌کنند، و هر مرز ثبت می‌کند عبور از آن چه الزاماتی دارد: احراز هویت در برابر ارائه‌دهنده، رمزگذاری در حین انتقال، ثبت آنچه خارج شده است. جریان پرامپت همان جریانی است که بیشترین موشکافی را می‌طلبد، و خود گراف دلیل آن را می‌گوید: متن حذف‌شده مشتری و مستندات داخلی بازیابی‌شده را از یک مرز به زیرساختی می‌برد که تیم نمی‌تواند آن را ببیند.

Worked example: a hosted inference system

An enterprise customer support assistant that retrieves internal knowledge, moderates input, and uses a hosted LLM to generate responses.



Source: OWASP AIBOM Foundations Guide v1.0

دو انتخاب ساختاری در این مثال شایسته بحث است.

مدل زبانی میزبانی شده به صورت یک مؤلفه نوع سرویس نمایش داده شده که مرز آن نقطه پایانی ارائه دهنده است. دلیل آن این است که جابه جایی داده در نقطه پایانی ارائه دهنده خاتمه می یابد، نه درون زیرساخت ارائه دهنده. تیم هیچ دیدی به توپولوژی سرویس دهی ارائه دهنده ندارد، و AIBOM که خلاف این وانمود کند، واقعیت هایی را ادعا می کند که تیم نمی تواند نشان دهد. مؤلفه نقطه پایانی همان چیزی است که تیم واقعاً می بیند: یک نقطه پایانی با یک نشانی، یک روش احراز هویت و یک قرارداد.

با این حال خود مدل از AIBOM غایب نیست. مؤلفه نقطه پایانی ارجاعی به یک مؤلفه مدل یادگیری ماشین دارد، و آن مؤلفه هویت، نسخه، مجوز، کاربرد مورد نظر، ویژگی های آموزش و دیگر فراداده های کارت مدل را در بر دارد. خواننده می تواند با یک گام دریا بد نقطه پایانی کدام مدل را ارائه می کند. اگر ارائه دهنده بعداً نام مستعار را تغییر دهد و آن را به نسخه لحظه ای جدیدتری از مدل اشاره دهد، مؤلفه نقطه پایانی جابه جا نمی شود. ارجاع به مؤلفه مدل تغییر می کند. بازنگری بعدی AIBOM تفاوت نسخه را ثبت می کند.

انتخاب دوم این است که هر گره چگونه به عنوان بخشی از سامانه یا بیرون از آن علامت گذاری شود. سرویس کاربر محور، پردازشگر پرامپت، سرویس بازیابی و مدیر پاسخ بخشی از سامانه ای هستند که AIBOM توصیف می کند: تیم آن ها را اداره می کند و دامنه اعلام شده آن ها را در بر می گیرد. API تعدیل محتوای شخص ثالث محل بحث بیشتری است. اگر تیم قابلیت تعدیل محتوا را بخشی از آنچه دستیار

را برای بهره‌برداری ایمن می‌کند بداند، سرویس تعدیل محتوا جزء سازنده است. اگر تیم آن را وابستگی خارجی‌ای بداند که دستیار صرفاً فرامی‌خواند، بیرونی است. هر دو انتخاب می‌توانند درست باشند؛ AIBOM انتخاب را صریح می‌کند تا بازبین بداند تولیدکننده کدام مدل را در نظر دارد.

مدل زبانی میزبانی‌شده نیز انتخاب تولیدکننده است. AIBOM ی که دامنه‌اش «دستیار به‌عنوان یک قابلیت مستقرشده» است، نقطه پایانی مدل را جزء سازنده تلقی خواهد کرد؛ مدل بخشی از دستیار است. AIBOM ی که دامنه‌اش «یکپارچه‌سازی‌ای که پیرامون یک مدل خارجی ساختیم» است، آن را بیرونی تلقی خواهد کرد. AIBOM این تمایز را از طریق خود دامنه اعلام‌شده رسمی می‌کند، که مؤلفه‌های گنجانده‌شده در سامانه توصیف‌شده و مستثنائشده از آن را فهرست می‌کند؛ خواننده مقدماتی تنها به ایده زیربنایی نیاز دارد، اینکه عضویت گره یک انتخاب وابسته به دامنه از سوی تولیدکننده است و نه ویژگی خود گره.

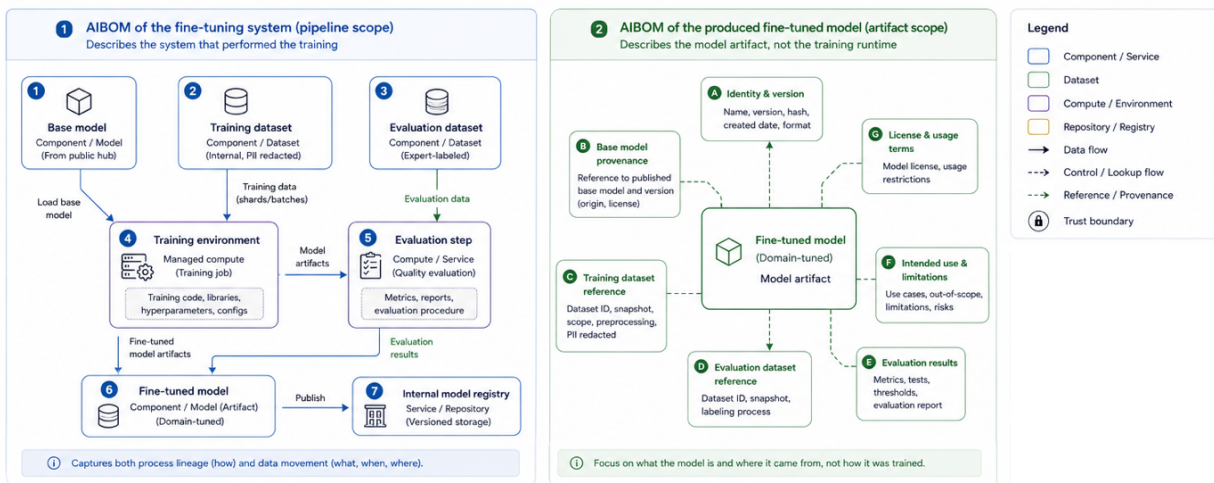
مثال کاربردی: یک خط لوله تنظیم دقیق

اکنون سامانه دیگری را در نظر بگیرید: تیمی که یک مدل پایه با وزن‌های باز را بر داده‌های داخلی تنظیم دقیق می‌کند تا یک مدل تنظیم‌شده برای حوزه خاص جهت استفاده در دستیار بالا تولید کند.

تیم با یک مدل پایه منتشرشده که از یک هاب عمومی مدل دریافت شده آغاز می‌کند. تیم یک مجموعه داده آموزشی دارد که از رونوشت‌های داخلی پشتیبانی مشتری با حذف اطلاعات شخصی گردآوری شده است. تیم یک مجموعه داده ارزیابی دارد که توسط متخصصان حوزه به‌صورت دستی برچسب‌گذاری شده و برای سنجش کیفیت مدل پیش از انتشار مدل تنظیم دقیق‌شده به پایین‌دست به کار می‌رود. تیم کار آموزش را روی یک محیط محاسباتی مدیریت‌شده اجرا می‌کند، یک مدل تنظیم دقیق‌شده تولید می‌کند، آن را ارزیابی می‌کند و در یک مخزن مدل داخلی منتشر می‌کند.

Worked example: a fine-tuning pipeline

Fine-tuning an open-weight base model on internal data to produce a domain-tuned model for use in a customer support assistant.



این سامانه دو AIBOM قابل تمایز تولید می‌کند، و این تمایز اهمیت دارد.

نخستین، AIBOM خود سامانه تنظیم دقیق است: خط لوله‌ای که آموزش را انجام داد. گراف آن شامل مؤلفه مدل پایه، مؤلفه مجموعه داده آموزشی، مؤلفه مجموعه داده ارزیابی، محیط آموزش، گام ارزیابی و مدل تنظیم دقیق شده تولیدی است. جریان‌های داده توصیف می‌کنند داده‌های آموزشی چگونه به گام آموزش رسیدند، مدل پایه چگونه بارگذاری شد، داده‌های ارزیابی چگونه به گام ارزیابی رسیدند، و مدل تولیدی چگونه در مخزن منتشر شد.

دومین، AIBOM خود مدل تنظیم دقیق شده تولیدی است که به عنوان یک مصنوع در نظر گرفته می‌شود. دامنه آن مدل است، نه خط لوله. هویت مدل، منشأ آن تا مدل پایه، ارجاعات به مجموعه داده‌های آموزش و ارزیابی، نتایج ارزیابی، کاربرد مورد نظر و مجوز را در بر دارد. محیط زمان اجرای تنظیم دقیق را توصیف نمی‌کند: تا زمانی که یک سامانه پایین‌دستی این مدل را بارگذاری می‌کند، محیط تنظیم دقیق دیگر حضور ندارد.

هر دو AIBOM به همان مجموعه داده‌ها و همان مدل پایه ارجاع می‌دهند. آن‌ها تکراری نیستند؛ دامنه‌های متفاوتی را توصیف می‌کنند. مصرف‌کننده‌ای که بازبینی ریسک مدل انجام می‌دهد AIBOM مصنوع را می‌خواند. تولیدکننده‌ای که محیط آموزش را حسابرسی می‌کند AIBOM سامانه را می‌خواند. سازمانی که می‌خواهد منشأ سرتاسری را نشان دهد، هر دو را نگهداری کرده و به هم پیوند می‌دهد.

مثال تنظیم دقیق همچنین تمایز میان تبار فرایند و جابه‌جایی داده را آشکار می‌کند. تبار فرایند ثبت می‌کند یک مصنوع چگونه به وجود آمد: از کدام مدل پایه استفاده شد، کدام کد آموزش اجرا شد، کدام ابرپارامترها بر کار حاکم بودند، کدام روال ارزیابی معیارهای گزارش شده را تولید کرد. جابه‌جایی داده ثبت می‌کند کدام داده به کجا رفت، میان کدام گره‌ها و چه زمانی. خطوط لوله آموزش به هر دو نیاز دارند. منشأ مدل تنظیم دقیق شده تا مدل پایه، تبار فرایند است. جابه‌جایی زمان اجرای قطعه‌های (shard) داده‌های آموزشی از یک انبار داده به محاسبات آموزش، جابه‌جایی داده است. AIBOM هر دو را، هر یک در جای خود، در بر دارد. برای خواننده مقدماتی، نکته این است که یک AIBOM مجبور نیست میان این دو دیدگاه یکی را برگزیند، و AIBOM سامانه آموزشی که فقط از یکی استفاده کند، آنچه را رخ داده کمتر از واقع توصیف خواهد کرد.

آنچه AIBOM بیان نمی‌کند

یک AIBOM کامل، واقعیت ساخت‌یافته و منشأ است. حکم نیست.

بیانیه‌ای مبنی بر انطباق یک سامانه با هیچ مقررات خاصی نیست. بیانیه‌ای مبنی بر پذیرفتنی بودن ریسک سامانه نیست. بیانیه‌ای مبنی بر کفایت کنترل‌های موجود در سامانه برای ریسک نیست. بیانیه‌ای درباره اینکه چه کسی از نظر حقوقی مسئول هر جنبه از سامانه است نیست. امتیاز اعتماد، سطح اطمینان یا رتبه‌بندی کیفیت نیست.

این یک انتخاب عامدانه در طراحی AIBOM است، نه یک غفلت. یک ناظر، حسابرس، هیئت بازبینی داخلی یا تیم تدارکات، سیاست‌ها، آستانه‌ها و چارچوب‌های ریسک خود را بر واقعیت‌های درون یک AIBOM اعمال می‌کند. این داورها در پایین‌دست قرار دارند. اگر AIBOM از پیش یک مجموعه از این داورها را کدگذاری کرده بود، این مصنوع فقط در همان نظام مفید بود. با ماندن در سطح واقعیت و منشأ، AIBOM به عنوان بستر مشترکی عمل می‌کند که نظام‌های بسیاری می‌توانند بر آن بنا شوند.

این بستر مشترک اختیاری نیست. AIBOM همان چیزی است که این داوری‌ها را ابطال‌پذیر می‌کند. بدون آن، ارزیابی ریسک ادعایی حسابرسی نشده است. با آن، هر ادعا در ارزیابی می‌تواند تا یک گره، یک جریان، یک ارجاع شاهد یا یک مرز کامل بودن اعلام شده ردیابی شود.

یک AIBOM چگونه نوشته می‌شود؛ برای خوانندگانی که عمیق‌تر می‌روند

مطالب مفهومی این فصل همان چیزی است که یک غریب‌پاده‌ساز نیاز دارد. برای تولیدکنندگانی که واقعاً یک AIBOM صادر خواهند کرد، دانستن یک نکته دیگر ارزشمند است.

جنبه‌هایی که در این فصل بحث شد به‌طور بومی در CycloneDX 2.0^۴ قابل بیان‌اند. گراف در یک نقشه طرح (blueprint) قرار دارد: مؤلفه‌ها گره‌ها هستند، جریان‌ها جابه‌جایی‌های داده جهت‌دارند، و مناطق و مرزها بستر اعتماد را حمل می‌کنند. خودتوصیف‌ها بر انواع مدل اعلام شده و دامنه نقشه طرح نگاشت می‌شوند، و ادعای کامل بودن توسط اعلان‌های ترکیب (composition) سند حمل می‌شود. شواهد از طریق اعلان‌هایی پیوست می‌شوند که با شناسه هر گره یا جریان به آن ارجاع می‌دهند. هیچ قالب فایل جداگانه‌ای برای AIBOM، هیچ طرح‌واره افزونه ویژه هوش مصنوعی و هیچ کدگذاری سازگاری برای انتخاب وجود ندارد: یک AIBOM سندی ساخت‌یافته است که از این ساختارهای استاندارد برای توصیف یک سامانه هوش مصنوعی استفاده می‌کند.

شناسه‌ها کاری بیش از دفترداری داخلی انجام می‌دهند. هر مؤلفه، جریان، منطقه، مرز و ادعا در سند یک ارجاع پایدار دارد، و استاندارد، مدل‌های تهدید، ریسک و کنترل خود را به‌عنوان هم‌سطح‌های نقشه طرح تعریف می‌کند که به همان گراف اشاره می‌کنند: یک تهدید به مرزی ارجاع می‌دهد که مسیر حمله‌اش از آن عبور می‌کند، یک کنترل به جریانی که از آن محافظت می‌کند، و یک تحلیل آسیب‌پذیری به مؤلفه‌هایی که تحت تأثیر قرار می‌دهد. هیچ‌یک سامانه را دوباره توصیف نمی‌کنند. این دلیل ساختاری آن است که AIBOM می‌تواند فارغ از حکم بماند و در عین حال بستری باشد که هر حکمی تا آن قابل ردیابی است؛ AIBOM را حذف کنید و تحلیل‌های پایین‌دستی چیزی برای اشاره کردن ندارند.

تولیدکننده همچنان می‌تواند انتخاب کند سند را با چه غنایی پر کند. یک AIBOM حداقلی (مؤلفه‌ها، جریان‌ها و سه خودتوصیف) کاملاً منطبق است و می‌تواند یک ادعای کامل از سوی تولیدکننده باشد. یک AIBOM با شواهد غنی، اعلان‌ها، ترکیب‌ها و سوابق فرمول‌بندی (formulation) را برای شواهد ماشین‌خوان و تبار فرایند می‌افزاید. غنا سطح اطمینان نیست. یک AIBOM حداقلی کامل و به‌خوبی مستند به شواهد، مفیدتر از یک AIBOM غنی ناقص است. اطمینان به یک AIBOM از کامل بودن، منشأ، کیفیت شواهد و سابقه عملکرد تولیدکننده ناشی می‌شود، نه از اینکه از کدام ساختارها استفاده می‌کند.

^۴ <https://cyclonedx.org/news/cyclonedx-v2.0-coming-soon/>

۶. AIBOM برای سامانه‌های هوش مصنوعی عامل‌محور

سامانه‌های هوش مصنوعی عامل‌محور سطح جدیدی از پیچیدگی را به زنجیره‌های تأمین هوش مصنوعی وارد می‌کنند. برخلاف سامانه‌های هوش مصنوعی سنتی که به‌صورت خطوط لوله نسبتاً ایستا عمل می‌کنند، سامانه‌های عامل‌محور پویا، ترکیب‌پذیر و اغلب خودمختارند. با ابزارهای خارجی تعامل می‌کنند، عامل‌های دیگر را فرامی‌خوانند، رفتار خود را در زمان اجرا تطبیق می‌دهند و به اکوسیستم‌های در حال تحولی از سرویس‌ها و پروتکل‌ها متکی‌اند.

رویکرد سنتی AIBOM بر مستندسازی ایستای مدل، مجموعه‌داده و وابستگی‌ها متمرکز است. در حالی که AIBOM برای سامانه‌های عامل‌محور باید وابستگی‌های پویا، تعاملات زمان اجرا و منطق هماهنگ‌سازی را که با هم تعیین می‌کنند این سامانه‌ها در محیط تولید چگونه رفتار کنند، در نظر بگیرد. این بخش نشان می‌دهد روش‌های AIBOM چگونه به معماری‌های عامل‌محور گسترش می‌یابند، با تأکید بر روابط چندعاملی، اکوسیستم‌های ابزار و اتصال‌پذیری خارجی.

وابستگی‌های چندعاملی

سامانه‌های عامل‌محور اغلب از چندین عامل همکار تشکیل شده‌اند که هر کدام ممکن است نقش‌های متمایزی مانند برنامه‌ریزی، اجرا، بازیابی، اعتبارسنجی یا هماهنگ‌سازی داشته باشند. این عامل‌ها ممکن است:

- بر پایه مدل‌های متفاوتی ساخته شده باشند (برای مثال، برنامه‌ریز مبتنی بر GPT + عامل بازیابی + اجراکننده ابزار).
 - متعلق به تیم‌ها یا فروشندگان مختلف باشند.
 - از طریق پروتکل‌های ساخت‌یافته یا تبادل پیام غیررسمی با هم ارتباط برقرار کنند.
- در این بستر، یک AIBOM باید صراحتاً موارد زیر را ثبت کند:
- هویت‌ها و نقش‌های عامل‌ها (برنامه‌ریز، اجراکننده، منتقد، بازیاب، ناظر، مسیریاب و غیره).
 - وابستگی‌های مدل زیربنایی برای هر عامل (مدل پایه، تنظیم‌های دقیق، آداپتورها).
 - الگوهای تعامل (سلسله‌مراتبی، هم‌تا به هم‌تا، رویدادمحور، هاب و پره).
 - هماهنگ‌سازی جریان کنترل (چه کسی چه کسی را فرامی‌خواند، تحت چه شرایطی، با چه مسیریهای ارجاع به سطح بالاتر).

این موضوع حیاتی است زیرا ریسک دیگر به یک مدل یا سرویس واحد محدود نیست. یک آسیب‌پذیری یا شکست در یک عامل می‌تواند در سراسر سامانه منتشر شود. برای مثال:

- یک عامل بازیابی به‌خطرافتاده می‌تواند استدلال و توصیه‌های پایین‌دستی را مسموم کند.
- یک برنامه‌ریز دچار توهم می‌تواند استفاده نایمن از ابزار یا گردش‌کارهای پرریسک را برانگیزد.
- یک عامل حفاظ نادرست پیکربندی‌شده می‌تواند در مسدودکردن خروجی‌های زیان‌بار یا نامنطبق ناکام بماند.

بنابراین AIBOM باید سامانه‌های عامل‌محور را به‌صورت گراف‌هایی از مؤلفه‌های وابسته به هم نمایش دهد و نه خطوط لوله خطی. BOM به گراف‌هایی از عامل‌ها، مدل‌ها و یال‌های هماهنگ‌سازی تبدیل می‌شود که تحلیل اثر را هنگام تغییر یک گره یا یال امکان‌پذیر می‌کند.

ابزارهای مرتبط

سامانه‌های هوش مصنوعی عامل‌محور بخش بزرگی از توان خود را از استفادهٔ ابزار می‌گیرند. ابزارها ممکن است شامل موارد زیر باشند:

- APIها (جست‌وجو، آب‌وهوا، مالی، منابع انسانی، تیکتینگ و غیره).
- محیط‌های اجرای کد و اجراکننده‌های تابع.
- پایگاه‌های داده و انبارهای برداری. حافظه و وضعیت.
- سرویس‌ها و ریزسرویس‌های داخلی سازمان.
- سامانه‌های فایل یا مخازن اسناد.

این ابزارها صرفاً یکپارچه‌سازی نیستند؛ امتدادهای فعال قابلیت‌های عامل‌اند که اغلب به‌طور خودمختار و با بسامد بالا فراخوانده می‌شوند.

یک AIBOM برای سامانه‌های عامل‌محور باید دست‌کم موارد زیر را مستند کند:

- هویت و کارکرد ابزار (ابزار چه می‌کند، با چه حوزه‌ای سروکار دارد).
- سازوکارهای فراخوانی (فراخوانی تابع، APIهای HTTP، افزونه‌ها، ابزارهای بومی چارچوب).
- دامنهٔ دسترسی و مجوزها (فقط‌خواندنی در برابر خواندن/نوشتن در برابر مدیریتی؛ دامنهٔ داده و منابع).
- روش‌های احراز هویت و مدیریت اسرار (کلیدهای OAuth، API، هویت بار کاری، انبارهای اسرار).
- محدودیت‌های نرخ، سهمیه‌ها و محدودیت‌های عملیاتی (به‌ازای هر عامل و سراسری).
- نسخه یا هش تعریف ابزار (طرح‌واره و توصیف دقیقی که عامل برای استفاده از آن تأیید شده است).

از منظر امنیت و حاکمیت، ابزارها گسترش بزرگی در سطح حمله به شمار می‌آیند:

- حملات تزریق پرامپت ممکن است استفاده از ابزار را دست‌کاری کرده و موجب اقدامات زیان‌بار شوند.
- ابزارهای دارای مجوز بیش از حد ممکن است داده‌های حساس را افشا کنند یا حرکت جانبی را ممکن سازند.
- APIهای خارجی ممکن است ریسک‌های خروج غیرمجاز داده، یکپارچگی یا دسترس‌پذیری وارد کنند.
- تعریف و رفتار یک ابزار می‌تواند پس از بازبینی تغییر کند، بنابراین تأیید باید به نسخهٔ مشخصی از تعریف پیوند بخورد و نه به نام ابزار.

ثبت وابستگی‌های ابزار در AIBOM موارد زیر را امکان‌پذیر می‌کند:

- حسابرسی دقیق دسترسی (کدام عامل‌ها می‌توانند کدام ابزارها را با چه مجوزهایی فراخوانند).
- ارزیابی ریسک در سطح ابزار (اهمیت، شعاع انفجار و کنترل‌های جبرانی).
- واکنش سریع‌تر به حادثه هنگام به‌خطرافتادن یک ابزار یا اتصال‌دهنده (چه کسی، کجا و چگونه از آن استفاده می‌کند).

اتصال‌دهنده‌های شخص ثالث

سامانه‌های عامل‌محور مدرن اغلب برای یکپارچه‌شدن با پلتفرم‌های خارجی به اتصال‌دهنده‌های شخص ثالث متکی‌اند، مانند:

- برنامه‌های SaaS (برای مثال، پلتفرم‌های همکاری، CRM، مجموعه‌های بهره‌وری و ابزارهای ITSM مانند ServiceNow و Slack، Salesforce، Google Workspace).
 - ارائه‌دهندگان داده (فیدهای مالی، پایگاه‌های داده حقوقی، API‌های سلامت).
 - سرویس‌های ابری و پلتفرم‌های هوش مصنوعی (ذخیره‌سازی، تحلیل، API‌های مدل).
- این اتصال‌دهنده‌ها وابستگی‌های زنجیره تأمین فراتر از کتابخانه‌ها و مدل‌های نرم‌افزاری سنتی وارد می‌کنند، از جمله:

- API‌ها و توافق‌نامه‌های سطح خدمات (SLA) تحت کنترل فروشنده.
- خطوط لوله و تبدیل‌های مبهم پردازش داده.
- سامانه‌های خارجی مدیریت هویت و دسترسی و موتورهای سیاست.

بنابراین AIBOM باید موارد زیر را ثبت کند:

از منظر امنیت و حاکمیت، اتصال‌دهنده‌ها ریسک‌هایی دارند که عمدتاً بیرون از کنترل مستقیم سازمان‌اند:

- دامنه‌های OAuth که هنگام راه‌اندازی اتصال‌دهنده اعطا می‌شوند ممکن است بسیار گسترده‌تر از نیاز مورد استفاده باشند و به عامل‌ها دسترسی دائمی به کل فضاهای کاری یا مجموعه‌های سوابق بدهند.

- فروشندگان ممکن است رفتار API، شیوه مدیریت داده یا پیمانکاران فرعی پردازش (sub-processors) را بدون اطلاع به دارندگان پایین‌دستی AIBOM تغییر دهند.
- داده‌ها ممکن است از طریق پیمانکاران فرعی‌ای که هرگز افشا نشده‌اند از مرزهای اقامت یا حاکمیت داده عبور کنند.

- جزئیات ارائه‌دهنده و سرویس اتصال‌دهنده (هویت فروشنده، محصول، منطقه، مدل استقرار).
- داده‌های مبادله‌شده (طبقه‌بندی ورودی/خروجی، دسته‌های داده، سطوح حساسیت).
- مدل‌های احراز هویت و مجوزدهی (جریان‌های OAuth، کلیدهای API، حساب‌های سرویس، دسترسی تفویضی).

- ملاحظات اقامت داده، حاکمیت داده و انطباق.

- محدودیت‌های قراردادی، مجوزی و استفاده که ممکن است بر عملیات یا مدیریت داده اثر بگذارند.

یک چالش کلیدی این است که دیدپذیری در سامانه‌های شخص ثالث اغلب محدود است. سازمان‌ها باید:

- در صورت امکان، افشاهای مشابه AIBOM یا گواهی‌های امنیتی را از فروشندگان مطالبه کنند.
- مرزهای اعتماد مفروض و ریسک‌های باقی‌مانده را صراحتاً مستند کنند.
- شکاف‌های شناخته‌شده در شفافیت، از جمله پیمانکاران فرعی ناشناخته یا مراحل پردازش مبهم را ردیابی کنند.

این کار تضمین می‌کند حتی وقتی منشأ کامل یا جزئیات درونی در دسترس نیست، ریسک به‌طور صریح در چارچوب AIBOM پذیرفته، ثبت و مدیریت شود.

ردیابی اکوسیستم MCP/A2A

پروتکل‌های نوظهوری مانند پروتکل بستر مدل (MCP یا Model Context Protocol)^۵ و چارچوب‌های ارتباط عامل به عامل (Agent-to-Agent یا A2A)^۶ به سرعت در حال شکل‌دادن به شیوه همکاری میان سامانه‌های عامل محورند. این اکوسیستم‌ها موارد زیر را ممکن می‌کنند:

- کشف و فراخوانی استاندارد شده ابزار در میان پلتفرم‌ها.
 - ارتباط میان عامل‌ها در سراسر مرزهای سازمانی یا مستأجر (tenant).
 - ترکیب پویای قابلیت‌ها در زمان اجرا بدون استقرار کد.
- این اکوسیستم‌ها با وجود قدرتمند بودن، وابستگی‌هایی بسیار پویا و بالقوه گذرا نیز وارد می‌کنند که ردیابی آن‌ها با روش‌های سنتی BOM دشوار است.
- برای ماندن در عرصه، AIBOM باید تحول یابد تا موارد زیر را در بر گیرد:
- استفاده از پروتکل (برای مثال، MCP، A2A و دیگر استانداردهای همکاری عامل‌ها).
 - نقاط پایانی ثبت شده و ارائه‌دهندگان سرویسی که در این اکوسیستم‌ها مشارکت دارند.
 - سازوکارهای کشف پویای قابلیت (رجیستری‌ها، کاتالوگ‌ها، بازارها).
 - وابستگی‌های سطح نشست یا سطح بستر، و نه فقط پیکربندی ایستا.
 - MCP: هویت سرور، لایه انتقال، تثبیت نسخه، هش‌های تعریف ابزار و دامنه‌های اعطاشده.
 - A2A: کارت‌های عامل (Agent Cards) که هویت، قابلیت‌ها و الزامات احراز هویت یک عامل را اعلام می‌کنند و مستقیماً به عنوان ورودی AIBOM به کار می‌روند.

برای مثال:

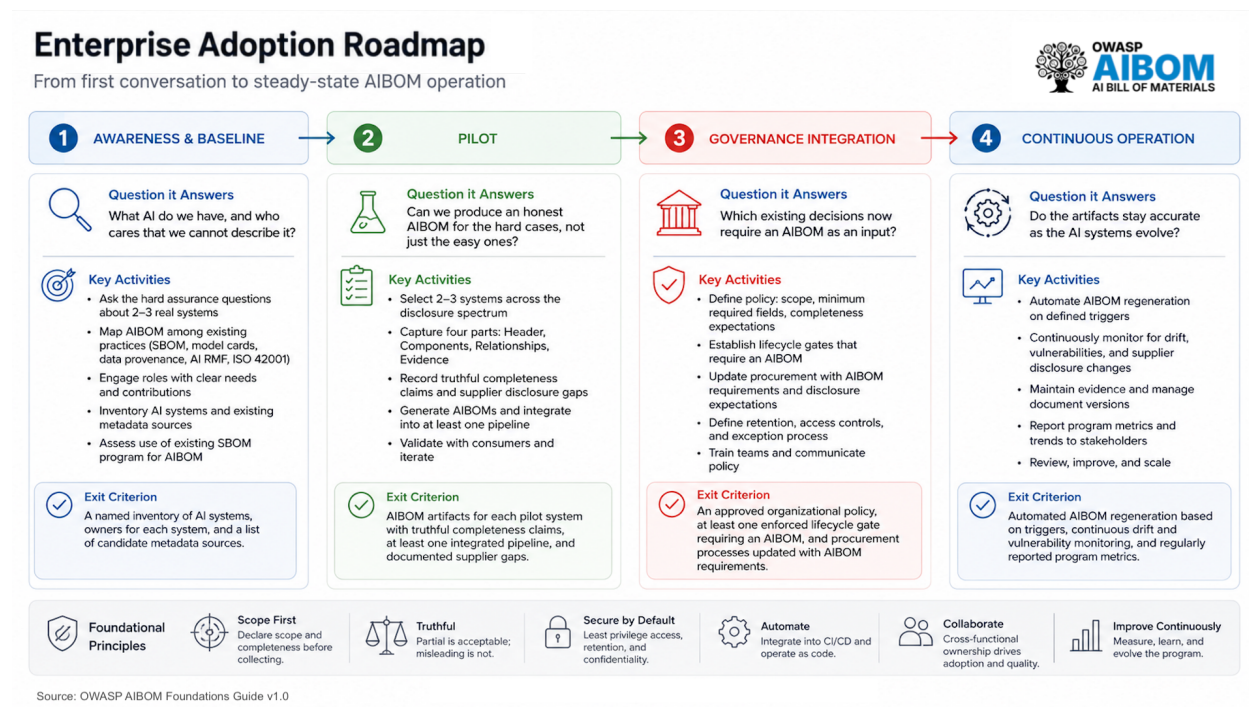
- یک عامل ممکن است یک ابزار را در زمان اجرا از طریق MCP کشف و فراخوانی کند، بدون پیکربندی ایستای قبلی.
 - دو عامل از دو فروشنده مختلف ممکن است از طریق کانال‌های A2A با هم همکاری کنند و وظایف و نتایج را مبادله کنند.
 - قابلیت‌ها و سیاست‌های در دسترس ممکن است بدون هیچ استقرار مجدد برنامه تغییر کنند.
- برای پرداختن به این موضوع، پیاده‌سازی‌های AIBOM باید:

- داده‌های دورسنجی زمان اجرا، لاگ‌ها و ردها را به عنوان منابع حقیقت برای وابستگی‌های واقعی به کار گیرند.
 - رجیستری‌های قابلیت، نقاط پایانی کشف و بازارها را به عنوان مؤلفه‌های درجه یک ردیابی کنند.
 - سوابق زمان‌مند از تعاملات و وابستگی‌ها نگهداری کنند (چه کسی، چه زمانی و از طریق چه پروتکلی با چه کسی ارتباط داشته است).
- این امر AIBOM را از یک مصنوع صرفاً ایستا به بازنمایی زنده‌ای از رفتار سامانه در طول زمان تبدیل می‌کند که پیکربندی زمان طراحی را با مشاهدات زمان اجرا درهم می‌آمیزد.

^۵ MCP: <https://modelcontextprotocol.io/specification/2026-07-28>
^۶ A2A: <https://a2a-protocol.org/latest/>

۷. نقشهٔ راه پذیرش سازمانی

تولید یک AIBOM واحد یک کار مهندسی است، و چرخهٔ عمر مصنوع که پیش‌تر توصیف شد (برنامه‌ریزی، تولید، اعتبارسنجی، انتشار، بهره‌برداری، حساسی و بازنشستگی) شکلی است که این کار به خود می‌گیرد. انجام آن برای هر سامانهٔ هوش مصنوعی در یک سازمان، به‌طور تکرارپذیر و با مصرف واقعی نتایج در تصمیم‌ها، یک برنامهٔ تغییر سازمانی است. این فصل مسیر را از نخستین گفت‌وگو تا عملیات در وضعیت پایدار ترسیم می‌کند.



چهار مرحله زیر بر اساس وابستگی مرتب شده‌اند. هر کدام به یک پرسش پاسخ می‌دهد و با شرطی به پایان می‌رسد که می‌توانید آن را بررسی کنید.

مرحله	پرسشی که پاسخ می‌دهد	معیار خروج
آگاهی و خط مبنا	چه هوش مصنوعی‌ای داریم، و چه کسی اهمیت می‌دهد که نمی‌توانیم آن را توصیف کنیم؟	یک فهرست نام‌دار از سامانه‌های هوش مصنوعی، یک مالک شناسایی‌شده برای هر سامانه، و فهرستی از منابع فراداده نامزد.
آزمایش پایلوت	آیا می‌توانیم برای موارد دشوار، و نه فقط موارد آسان، یک AIBOM صادقانه تولید کنیم؟	مصنوعات AIBOM تولیدشده برای هر سامانه پایلوت با ادعاهای کامل بودن صادقانه، دست‌کم یک خط لوله یکپارچه، و شکاف‌های مستندشده در افشای تأمین‌کنندگان.
یکپارچه‌سازی با حاکمیت	کدام تصمیم‌های موجود اکنون به AIBOM به‌عنوان ورودی نیاز دارند؟	یک سیاست سازمانی مصوب، دست‌کم یک دروازه چرخه عمر اجراشده که AIBOM را الزامی می‌کند، و فرایندهای تدارکاتی به‌روزشده با الزامات AIBOM.
عملیات مستمر	آیا مصنوعات با تحول سامانه‌های هوش مصنوعی دقیق باقی می‌مانند؟	بازتولید خودکار AIBOM بر اساس محرک‌های تعریف‌شده، پایش مستمر رانش و آسیب‌پذیری، و معیارهای برنامه که به‌طور منظم گزارش می‌شوند.

مرحله ۱ – آگاهی و خط مبنا

لازم نیست استدلال شروع کار در این‌جا از نو ساخته شود: محرک‌های امنیتی، مقرراتی و حاکمیتی که پیش‌تر در این سند تشریح شدند همان استدلال‌اند. آنچه این مرحله می‌افزاید، مشخص‌بودن است. پرسش‌های اطمینانی که سازمان در حال حاضر نمی‌تواند به آن‌ها پاسخ دهد (کدام نسخه مدل در محیط تولید است، بر چه چیزی آموزش یا تنظیم دقیق شده، تأمین‌کننده بالادستی کیست، و مجوز آن چه اجازه‌ای می‌دهد) باید درباره دو یا سه سامانه واقعی با صدای بلند پرسیده شوند. ناتوانی در پاسخ به آن‌ها مؤثرترین استدلال برای جلب حمایت است و از هر استناد به چارچوب‌ها متقاعدکننده‌تر است.

آگاهی همچنین به معنای قرارداد دادن درست AIBOM در میان روش‌هایی است که سازمان ممکن است هم‌اکنون اجرا کند. یک SBOM مؤلفه‌های نرم‌افزاری را فهرست می‌کند و همچنان ضروری است؛ یک AIBOM همان انضباط را به مدل‌ها، مجموعه داده‌ها، پرامپت‌ها و حفاظ‌ها، و سرویس‌های مسیر درخواست گسترش می‌دهد. یک کارت مدل کاربرد مورد نظر و عملکرد یک مدل را مستند می‌کند؛ یک AIBOM آن محتوا را به‌عنوان شاهدهی پیوست‌شده به یک مؤلفه، درون سندی نسخه‌دار با دامنه اعلام‌شده، حمل می‌کند. کار منشأ داده، تبار مجموعه داده‌ای را که AIBOM ثبت می‌کند تأمین می‌کند. چارچوب‌های تثبیت‌شده مدیریت ریسک هوش مصنوعی و نظام‌های مدیریتی، مستندسازی ساخت‌یافته ویژگی‌های سامانه را طلب می‌کنند بی‌آنکه شکل آن را مشخص کنند؛ AIBOM شکلی است که این مطالبه را برآورده می‌کند. AIBOM جایگزین هیچ‌یک از این‌ها نمی‌شود و هیچ‌یک را تکرار نمی‌کند.

مشارکت دادن ذی‌نفعان باید از همان ابتدا متناسب با نقش باشد، زیرا ناهماهنگی در این‌جا به آزمایش‌های پایلوت پراکنده‌ای می‌انجامد که هرگز به حاکمیت نمی‌رسند.

کارکرد	آنچه از AIBOM نیاز دارد	آنچه مشارکت می‌دهد
معماری امنیت و تیم قرمز	اطلاعات سطح حمله و ویژه هوش مصنوعی که بر مؤلفه‌ها نگاشت شده است.	دامنه مدل‌سازی تهدید و تعریف پرسش‌های امنیتی‌ای که AIBOM باید به آن‌ها پاسخ دهد.
مشاور حقوقی و مالکیت فکری	شرایط مجوز مدل و مجموعه داده، حقوق استفاده و محدودیت‌های مربوط.	تفسیر تعهدات حقوقی، الزامات مجوز و انطباق با مالکیت فکری.
انطباق و حسابرسی	شواهد پشتیبان شفافیت، ردیابی‌پذیری و تعهدات مقرراتی هوش مصنوعی.	سیاست‌های نگهداری، کنترل‌های حاکمیتی، الزامات حسابرسی و انتظارات شواهد.
مهندسی پلتفرم و MLOps	الزامات خودکارسازی، سازوکارهای گردآوری فراداده و نقاط یکپارچه‌سازی خط لوله.	پیاده‌سازی تولید خودکار AIBOM، یکپارچه‌سازی CI/CD و نگهداری مستمر.
تیم‌های هوش مصنوعی و یادگیری ماشین	به رسمیت شناختن اینکه ثبت تبار بخش جدایی‌ناپذیر توسعه هوش مصنوعی است.	تبار داده‌های آموزشی، منشأ مدل پایه، تاریخچه تنظیم دقیق، نتایج ارزیابی و فراداده‌های توسعه مدل.

کشف خط مبنا این مرحله را به پایان می‌رساند: فهرست‌برداری از هر سامانه هوش مصنوعی در محیط تولید یا در حال توسعه فعال (مدل‌های آموزش‌دیده داخلی، نسخه‌های تنظیم دقیق‌شده، API‌های مدل تدارک‌شده از بیرون، و قابلیت‌های هوش مصنوعی درون SaaS‌های تدارک‌شده)؛ شناسایی فراداده‌هایی که هم‌اکنون در کارت‌های مدل، لاگ‌های اجرای آموزش، رجیستری‌های مجموعه داده و مانیفست‌های وابستگی وجود دارند؛ و ارزیابی اینکه آیا یک برنامه SBOM موجود می‌تواند کار AIBOM را بر دوش بگیرد، به‌جای آنکه برنامه جدیدی در کنار آن راه‌اندازی شود.

مرحله ۲ – آزمایش آزمایشی (پایلوت)

دو یا سه سامانه را انتخاب کنید که در نقاط مختلفی از طیف افشا قرار دارند، زیرا دشواری تألیف یک AIBOM تقریباً به‌طور کامل با میزان آنچه تولیدکننده مجاز به دانستن آن است تعیین می‌شود. یک مدل آموزش‌دیده و مستقرشده داخلی دسترسی کامل به فراداده می‌دهد و نقطه شروع قابل‌مدیریتی است. یک نسخه تنظیم دقیق‌شده داخلی از یک مدل پایه خارجی، پایلوت را وادار می‌کند دو زنجیره منشأ را ثبت کند که یکی از آن‌ها متعلق به دیگری است، و رایج‌ترین مورد سازمانی است. یک مدل میزبانی‌شده که از طریق API به آن دسترسی می‌شود، می‌آزماید که آیا سازمان می‌تواند به افشای تأمین‌کننده تکیه کند و دقیقاً آشکار می‌کند آن افشا کجا متوقف می‌شود. یک قابلیت هوش مصنوعی

درون یک برنامه SaaS تدارک شده، کف پوشش دست‌یافتنی را تعیین می‌کند و با این کار، مفیدترین ورودی تدارکاتی کل پایلوت را تولید می‌کند.

ثبت اطلاعات همان‌جایی است که پایلوت بیش از همه به‌اشتباه اداره می‌شود، زیرا تیم‌ها فیلدها را گردآوری می‌کنند بی‌آنکه تصمیم بگیرند سند چه چیزی را ادعا می‌کند. کار را حول چهار بخشی که پیش‌تر توصیف شد ساختار بندی کنید:

سرآیند. پیش از گردآوری هر چیز، دامنه را اعلام کنید: کدام سامانه، کدام نسخه، کدام استقرار، کدام رابط‌ها در داخل و بیرون‌اند؛ سپس ادعای کامل‌بودنی را ثبت کنید که صادقانه از آنچه واقعاً در دسترس بوده به دست آمده باشد. ادعای ناقص دربارهٔ یک قابلیت SaaS جعبه‌سیاه یک مصنوع درست است. ادعای کامل دربارهٔ همان سامانه یک ادعای نادرست است.

مؤلفه‌ها. هویت مدل، نسخه، معماری و ارجاع به مدل پایه؛ مجموعه‌داده‌ها با نسخه، منبع، روش گردآوری، مجوز و محدودیت‌های شناخته‌شده؛ ابزارها، هماهنگ‌کننده‌ها و اتصال‌دهنده‌هایی که سامانه فرامی‌خواند؛ پرامپت‌ها و سیاست‌های حفاظ؛ محیط زمان اجرا و وابستگی‌های آن، با ارجاع متقابل به سوابق SBOM موجود به‌جای فهرست‌برداری مجدد.

جریان‌ها. چه داده‌ای میان آن مؤلفه‌ها و از میان کدام مرزها حرکت می‌کند؛ بخشی که هیچ چک‌لیست فیلدی آن را تولید نمی‌کند، و بخشی که واکنش‌دهندگان به حادثه پیش از همه از آن استفاده می‌کنند.

شواهد. هش‌های فایل‌های وزن و نسخه‌های لحظه‌ای مجموعه‌داده، امضاها، کارت‌های مدل، نتایج ارزیابی، و سابقهٔ پاسخ‌گوی تألیف، بازبینی و تأیید که سند را بازبینی‌پذیر می‌کند و نه صرفاً موجود.

در زمینهٔ خودکارسازی، پرسش پایلوت این نیست که از کدام قالب استفاده شود (این موضوع در جای دیگری از این سند حل شده است) بلکه این است که هر فیلد از کجا می‌آید. هر جا اطلاعات هم‌اکنون در خطوط لولهٔ ساخت و آموزش وجود دارد، از آن‌جا استخراج کنید، به هر سامانهٔ ردیابی آزمایش که سازمان اجرا می‌کند متصل شوید، و صریح بگویید کدام فیلدها باید دستی تألیف شوند و مالک آن‌ها کیست. در برابر متعهدکردن برنامه به یک پلتفرم تنها بر پایهٔ شواهد پایلوت مقاومت کنید.

تألیف بر عهدهٔ تیمی است که به هر سامانه نزدیک‌تر است، و بازبینی و تأیید بر عهدهٔ یک کارکرد حاکمیتی یا امنیتی تعیین‌شده. متمرکزکردن تألیف در تیمی بدون دانش مستقیم از سامانه، همان حالت شکستی است که این ترتیب برای جلوگیری از آن وجود دارد.

مرحلهٔ ۳ – یکپارچه‌سازی با حاکمیت

یکپارچه‌سازی به این معناست که AIBOM به ورودی الزامی تصمیم‌هایی تبدیل شود که سازمان هم‌اکنون می‌گیرد. ایجاد یک فرایند موازی AIBOM دلیل آن است که بیشتر برنامه‌ها پس از پایلوت متوقف می‌شوند.

فرایند موجود	آنچه AIBOM تأمین می‌کند	کجا الزام‌آور می‌شود
مدیریت ریسک مدل		ورودی الزامی هر ارزیابی و بازبینی ریسک مدل.

	فهرستی جامع از مؤلفه‌ها، وابستگی‌ها و فراداده‌های پشتیبان هوش مصنوعی که برای ارزیابی ریسک لازم است.	
چرخه عمر توسعه هوش مصنوعی	دامنه اعلام‌شده سامانه، منشأ، تبار و شواهد ارزیابی که در هر مرحله چرخه عمر ثبت می‌شوند.	دروازه الزامی برای تأیید پیش از آموزش، بازبینی پیش از استقرار و ارزیابی پس از حادثه.
تدارکات و ریسک اشخاص ثالث	اطلاعات منشأ اعلام‌شده توسط تأمین‌کننده، وابستگی‌های نرم‌افزاری و مدل، و شناسایی صریح افشاهای ناموجود یا ناقص.	حداقل الزام افشای تأمین‌کننده در ارزیابی فروشندگان و سامانه‌های هوش مصنوعی پریسک.
بازبینی طراحی امنیتی	مؤلفه‌های هوش مصنوعی، جریان‌های داده، وابستگی‌ها و مرزهای اعتماد که برای مدل‌سازی تهدید و بازبینی معماری لازم‌اند.	ورودی استاندارد بازبینی‌های معماری و طراحی امنیت هوش مصنوعی.
آمادگی مقرراتی و حسابرسی	شواهد نسخه‌کنترل‌شده و ردیابی‌پذیر که بر تعهدات مقرراتی، حاکمیتی و انطباقی مربوط نگاشت شده‌اند.	سند حاکمیتی نگهداری‌شده که در بازرسی‌های مقرراتی، حسابرسی‌ها و ارزیابی‌های انطباق ارائه می‌شود.

یک سیاست رسمی این را قابل اجرا می‌کند. این سیاست باید حداقل AIBOM قابل استفاده را برای هر رده ریسک، مسئولیت‌های تألیف و بازبینی، دوره‌های نگهداری، محرک‌های به‌روزرسانی، و مسیر ارجاع به سطح بالاتر در صورت عدم افشای تأمین‌کننده تعریف کند. خود مصنوع باید به‌عنوان یک سند کنترل‌شده در نظر گرفته شود: نسخه‌کنترل‌شده، امضا شده، با ثبت هویت بازبین، تاریخ بازبینی و دامنه بازبینی، و مشمول همان کنترل‌های نگهداری و دسترسی که برای سوابق اعتبارسنجی مدل اعمال می‌شود.

مرحله ۴ – عملیات مستمر

یک AIBOM سامانه را در یک لحظه توصیف می‌کند. سامانه‌ها پیوسته تغییر می‌کنند، و مصنوعی که دیگر با استقرار مطابقت ندارد از نبود آن بدتر است، زیرا اطمینانی را برمی‌انگیزد که نمی‌تواند از آن پشتیبانی کند.

بازتولید باید با این رویدادها برانگیخته شود: آموزش مجدد یا تازه‌سازی مجموعه داده؛ جایگزینی آداپتور یا تغییر پارامترهای تنظیم دقیق؛ تغییر نسخه مجموعه داده، از جمله تغییرات افزایشی؛ تغییرات مدل بالادستی یا نام مستعار که توسط تأمین‌کننده انجام می‌شود؛ ارتقای وابستگی‌ها در محیط زمان اجرای استنتاج یا مسیر پیش‌پردازش؛ تغییرات حفاظ و بسته سیاست؛ آسیب‌پذیری‌های افشاشده‌ای که یک مؤلفه ردیابی‌شده را تحت تأثیر قرار می‌دهند؛ و هر تغییری در کاربرد مورد نظر، دامنه استقرار یا مرزهای دسترسی. تغییرات سمت تأمین‌کننده توجه ویژه‌ای می‌طلبند، زیرا این رده محرک بدون هیچ انتشار محلی فعال می‌شود.

از نظر عملیاتی این به معنای تعبیه تولید AIBOM در خطوط لوله آموزش است تا مصنوعات پس از ساخت موفق تولید شوند؛ بررسی‌های اعتبارسنجی در دروازه‌های استقرار تا سوابق ناقص یا کهنه مانع ارتقا شوند؛ واردکردن سوابق مؤلفه‌ها به همان گردش کارهای تطبیق آسیب‌پذیری که برنامه SBOM هم‌اکنون از آن‌ها استفاده می‌کند، به جای ایجاد یک جزیره جداگانه دوم؛ و راستی‌آزمایی یکپارچگی تا تغییرات پس از تولید قابل تشخیص باشد. مالکیت در امتداد خطوط موجود تقسیم می‌شود: تیم‌های مهندسی یادگیری ماشین و پلتفرم به‌روزرسانی می‌کنند، عملیات امنیت هشدارها را اولویت‌بندی می‌کند، کارکرد حاکمیتی گزارش می‌دهد، و ریسک اشخاص ثالث افشای تأمین‌کنندگان را ردیابی می‌کند.

سلامت برنامه ارزش آن را دارد که با شمار اندکی از شاخص‌ها سنجیده شود: **پوشش**، سهم سامانه‌های هوش مصنوعی تولیدی که مصنوع جاری دارند؛ **تازگی**، فاصله زمانی میان یک رویداد محرک و به‌روزرسانی متناظر؛ **پرش‌دگی فیلدها** در برابر حداقل AIBOM قابل‌استفاده، به تفکیک رده ریسک؛ **واکنش به آسیب‌پذیری**، زمان از افشا تا ارزیابی اثر؛ و **انطباق تأمین‌کنندگان**، سهم تأمین‌کنندگان قراردادی هوش مصنوعی که حداقل استاندارد افشا را برآورده می‌کنند. این شاخص‌ها برنامه را می‌سنجند، نه سامانه‌هایی را که مستند می‌کند. یک سابقه کاملاً پرشده از یک سامانه بدحکمرانی‌شده همچنان تنها یک سابقه کاملاً پرشده است، و هیچ‌یک از این معیارها رتبه‌بندی سامانه هوش مصنوعی یا AIBOM نیست. آن‌ها را همان‌جایی نمایش دهید که ریسک هوش مصنوعی هم‌اکنون گزارش می‌شود، تا همان شواهد بدون آنکه سه بار تهیه شوند، به آمادگی مقرراتی، حسابرسی داخلی و گزارش‌دهی اجرایی خدمت کنند.

قدردانی

مشارکت‌کنندگان

Protectt.ai، هند	Yuvaraj Govindarajulu (رهبر جریان کاری، رهبر مشترک پروژه)
بنیان‌گذار/مدیرعامل، استارت‌آپ در حالت مخفی، آمریکا	Aruneesh Salhotra (رهبر پروژه)
Fabletics	Martin Espinoza
Model Monster، آمریکا	Van Lindberg
Onix، آمریکا	Tanmay Ambegaokar
Steve Madden، آمریکا	Bakul Singhal
	L.Frances. Domingo
استارت‌آپ در حالت مخفی، آمریکا	Anmol Kumar
SISA، هند	Binal Patel
NuGuard AI، آمریکا	Ranjan Goel
Dayforce، آمریکا	Sapna Paul
OWASP، آلمان	Behnaz Karimi
ProCap360، آمریکا	John Cavanaugh
Bajaj Financial Services، هند	Yash Omer
Goldman Sachs، آمریکا	Pratibha Aphale
EY، آمریکا	Deba Mishra
	Karen Bennet
	Chetan Velicheti

حامیان پروژه

از حامیان پروژه که مشارکتهای مالی آنان، افزون بر منابعی که بنیاد OWASP فراهم می‌کند، به پیشبرد اهداف پروژه OWASP AIBOM و پوشش هزینه‌های عملیاتی کمک می‌کند، سپاسگزاریم.

حامیان به پاس مشارکتهایشان در مطالب و وبسایت‌های ما معرفی می‌شوند. همهٔ مطالبی که پروژه تولید می‌کند توسط جامعه توسعه یافته و جامعه‌محورند، و تحت مجوزهای متن‌باز و Creative Commons منتشر می‌شوند. برای آگاهی بیشتر دربارهٔ حامی‌شدن و پشتیبانی از پروژه، لطفاً به بخش حمایت مالی وبسایت ما به نشانی <https://owaspai bom.org/sponsorship/> مراجعه کنید.

حامیان طلایی پروژه



ThreatReaper

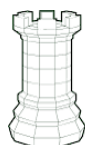
حامیان نقره‌ای پروژه



Codacy



MANIFEST



**Prediction
Guard**



راهنمای مبانی OWASP AIBOM

سپتامبر ۲۰۲۶

نسخه ۱٫۰ (فارسی)