



شورای راهبری و سیاستگذاری هوش مصنوعی در نظام سلامت

مجموعه استانداردها و الزامات اعتباربخشی محصولات مبتنی بر هوش مصنوعی در حوزه سلامت

مقدمه: در این سند، مجموعه استانداردها و الزامات مورد استفاده برای ارزیابی و اعتباربخشی محصولات مبتنی بر هوش مصنوعی در حوزه سلامت ارائه می‌شود. هدف از تدوین این مجموعه آن است که ارزیابی این محصولات صرفاً به قضاوت‌های موردی یا الگوهای عمومی تجهیزات پزشکی محدود نماند و با اتکا به معیارهای مشخص، قابل ممیزی و مبتنی بر شواهد انجام شود؛ به گونه‌ای که تصمیم‌گیری درباره صدور، تمدید، تعلیق یا ابطال گواهی، شفاف، قابل پیگیری و قابل دفاع باشد. این مجموعه بر اساس نتایج طرح «طراحی ساختار و شاخص‌های اعتباربخشی محصولات مبتنی بر هوش مصنوعی در حوزه سلامت» و به سفارش سازمان غذا و داروی وزارت بهداشت، درمان و آموزش پزشکی در شورای راهبری و سیاستگذاری هوش مصنوعی دانشگاه علوم پزشکی تهران تدوین شده است و به عنوان مرجع عملیاتی برای ارزیابان فنی، بالینی و امنیت/حریم خصوصی و نیز به عنوان راهنمای آماده‌سازی پرونده فنی برای توسعه‌دهندگان عمل می‌کند و هدف نهایی آن، تسهیل ورود مسئولانه، ایمن و قابل اعتماد محصولات مبتنی بر هوش مصنوعی به نظام سلامت و حفظ کیفیت و ایمنی آن‌ها در طول زمان است.

تعاریف و اختصارات

تعاریف:

۱. اعتبارسنجی تحلیلی/فنی (Analytical/Technical Validation): اطمینان از درست و پایدار بودن عملکرد فنی مدل، پردازش ورودی‌ها با منطق تعریف‌شده و تولید خروجی‌های دقیق، تکرارپذیر و بازتولیدپذیر.
۲. اعتبارسنجی بالینی-کارکردی (Clinical/Functional Validation): نشان دهد مدل برای هدف بالینی مشخص واقعاً کار می‌کند؛ با «معیار طلایی/Outcome» ارتباط دارد و در جمعیت/بافت هدف عملکرد قابل قبول ارائه می‌دهد.
۳. ارزیابی مستقل و پایداری عملکرد (Independent Evaluation & Performance Stability): اثبات تعمیم‌پذیری و پایداری قبل از استقرار (مثلاً با آزمون خارجی/مستقل و تست‌های پایداری).

۴. تحلیل زیرگروه‌ها و عدالت عملکرد (Subgroup Analysis & Fairness): اطمینان از منصفانه و پایدار بودن عملکرد مدل در زیرگروه‌های مرتبط (جمعیتی/بالینی/تجهیزاتی/مرکزی) و شناسایی و مدیریت اختلافات معنادار.
۵. پایش پس‌بازار و عملکرد دنیای واقعی (PMS/PMCF & Real-World Performance): وجود یک سیستم نظام‌مند برای پایش عملکرد مدل پس از صدور مجوز که بتواند داده‌های «عملکرد واقعی» را جمع‌آوری، تحلیل و به اقدامات اصلاحی تبدیل کند؛ با تأکید بر پایش فعال، جمع‌آوری بازخورد بالینی و گزارش‌دهی دوره‌ای.
۶. انحراف/رانش مدل (Model Drift: Data/Concept Drift): افت یا تغییر عملکرد مدل در محیط واقعی که معمولاً به تغییر توزیع داده‌های ورودی/شرایط استفاده یا تغییر رابطه ورودی-خروجی رخ می‌دهد و نیازمند کشف و مدیریت در پایش پس‌بازار است.
۷. برنامه کنترل تغییرات از پایش تعیین‌شده (PCCP): رویکرد/برنامه‌ای برای مدیریت تغییرات و به‌روزرسانی‌های مدل در چرخه عمر (برای اینکه تغییرات، کنترل‌شده، مستند و قابل ارزیابی باشند).
۸. اقدامات اصلاحی و پیشگیرانه (CAPA): فرایند رسیدگی و اصلاح/پیشگیری در برابر مشکلات شناسایی‌شده در پایش پس‌بازار (مثل رخدادهای ایمنی، شکایات، یا drift‌های مرتبط با ایمنی).
۹. ارزیابی اثرات حریم خصوصی (DPIA): ارزیابی ساختاریافته اثرات پردازش داده بر حریم خصوصی و ریسک‌های حقوقی/اخلاقی، به‌عنوان جزء الزامات حفاظت از داده.

اختصارات:

- AI:** هوش مصنوعی
- ML:** یادگیری ماشین
- SaMD:** نرم‌افزار به‌عنوان ابزار/تجهیز پزشکی (Software as a Medical Device)
- FDA:** سازمان غذا و داروی آمریکا
- IMDRF:** انجمن هماهنگی مقررات تجهیزات پزشکی
- EU:** اتحادیه اروپا
- MDR/IVDR:** مقررات تجهیزات پزشکی/تشخیص آزمایشگاهی اتحادیه اروپا
- AI Act:** قانون هوش مصنوعی اتحادیه اروپا
- GMLP:** اصول «عملکرد خوب یادگیری ماشین» (Good Machine Learning Practice)
- PMS/PMCF:** پایش پس‌بازار/پیگیری بالینی پس‌بازار
- RWP:** عملکرد دنیای واقعی (Real-World Performance)
- PCCP:** برنامه کنترل تغییرات از پایش تعیین‌شده
- DPIA:** ارزیابی اثرات حریم خصوصی
- CAPA:** اقدامات اصلاحی و پیشگیرانه

استانداردها و الزامات

استانداردها و الزامات در قالب چند حیطه اصلی سازمان‌دهی شده‌اند تا با منطق چک‌لیست‌های طرح هم‌راستا باشند. ارزیابی هر استاندارد از طریق پاسخ‌گویی توسعه‌دهنده به پرسش‌های چک‌لیست‌ها، ارائه شواهد مستند، و قضاوت ساختاریافته ارزیاب انجام می‌شود و نتیجه در مقیاس امتیازدهی طرح ثبت می‌گردد.

حیطه ۱) حاکمیت داده (کیفیت و منشأ داده)

هدف این حیطه اطمینان از کیفیت، تمامیت، منشأ، امنیت و انطباق داده‌های مورد استفاده در توسعه، اعتبارسنجی و استقرار محصولات هوش مصنوعی پزشکی است. تحقق الزامات این حیطه باید از طریق مستندات نسخه‌دار، شواهد قابل ممیزی، و قابلیت بازتولید نشان داده شود.

استاندارد ۱-۱: سیاست کیفیت داده، پاک‌سازی، تمامیت و امنیت

متقاضی باید یک سند رسمی نسخه‌دار برای کیفیت داده ارائه کند که حداقل سه محور را پوشش دهد: ۱) قواعد پاک‌سازی/اعتبارسنجی داده، ۲) تمامیت و نسخه‌بندی/ردیابی تغییرات داده، ۳) امنیت و کنترل دسترسی. همراه سند، باید شواهد اجرایی (گزارش‌های QC قبل/بعد، نسخه‌بندی مجموعه‌ها، IAM/Role-Based Access، لاگ ورود/دسترسی، رمزنگاری و پشتیبان‌گیری) ارائه شود.

استاندارد ۱-۲: مستندسازی پیش‌پردازش و بازتولیدپذیری

متقاضی باید نشان دهد پیش‌پردازش داده‌ها قابل بازتولید است؛ یعنی مراحل پیش‌پردازش، کد/اسکرپت‌ها، تنظیمات، seed و نسخه کتابخانه‌ها/محیط اجرا مستند باشد و امکان اجرای نمونه ورودی/خروجی وجود داشته باشد. همچنین KPI‌های مرتبط با عملکرد/کیفیت (مطابق هدف محصول و نوع داده) باید با فرمول، آستانه قابل قبول و دوره اندازه‌گیری تعریف شوند تا در پایش و کنترل کیفیت قابل استفاده باشند.

استاندارد ۱-۳: مدیریت داده‌های گم‌شده و تحلیل سوگیری ناشی از آن

متقاضی باید روش برخورد با داده‌های گم‌شده را شفاف بیان کند (حذف، جانشینی، imputation، استفاده از مقادیر جایگزین و...)، و اثر این تصمیم بر توزیع داده، سوگیری و نتایج احتمالی را تحلیل و مستند نماید.

استاندارد ۱-۴: حداقل چک‌فهرست حاکمیت داده و امنیت

(SOP/Data Flow/Owner/Controls/Logging/Monitoring/DPIA/IR)

متقاضی موظف است حداقل عناصر پایه زیر را به‌صورت «وجود دارد/وجود ندارد» همراه با ارجاع به شواهد تکمیل و ارائه کند:

- SOP‌های مرتبط (مسیر/نسخه)

- دیاگرام جریان داده (Data Flow)
- مالک داده و مسئولیت‌ها (Data Owner)
- کنترل‌های فنی امنیت (IAM/رمزنگاری)
- لاگینگ و نسخه‌بندی
- مانیتورینگ کیفیت/رانش
- DPIA/انطباق قانونی (در صورت اقتضا)
- برنامه پاسخ به حادثه (Incident Response)

در صورت نبود هر مورد، باید دلیل فنی و ریسک ناشی از آن مستند شود.

استاندارد ۱-۵: چارچوب حاکمیت داده

متقاضی باید چارچوب حاکمیت داده را به صورت عملیاتی و نسخه دار ارائه کند؛ شامل نقش‌ها (مالک/سرپرست داده/امنیت)، فرایندها، SLAها، و «manifest» یا فهرست/شناسنامه مجموعه داده‌ها که ردیابی و مسئولیت‌پذیری را ممکن کند. در صورت استفاده از داده با منشأ خارج از کشور، محدودیت‌های دسترسی/کنترل (از جمله محدودیت‌های ارتباطی/تحریم) باید شفاف تشریح شود.

استاندارد ۱-۶: تشخیص و پاسخ به رانش داده

متقاضی باید ساز و کار تشخیص رانش داده/مفهوم را مشخص کند (مثلاً PSI، آزمون KS، KL، یا پایش عملکرد مدل در زمان)، آستانه‌های هشدار را مستدل تعریف نماید و سیاست واکنش را توضیح دهد (مانند alert، محدودسازی، بازآموزی، یا بازنگری).

استاندارد ۱-۷: مشارکت خبرگان بالینی در آماده‌سازی داده

متقاضی باید نشان دهد خبرگان بالینی در مراحل حساس چرخه داده (پاک‌سازی، انتخاب ویژگی‌ها، آنوتیشن/داوری) مشارکت داشته‌اند و این مشارکت قابل ردیابی است؛ همچنین باید مستند کند که ویژگی‌ها/پاک‌سازی‌ها با Use-case بالینی محصول معنادار و همسو هستند.

استاندارد ۱-۸: منشأ داده و تطابق با جمعیت هدف (و ادعای چندمرکزی/چندسخت‌افزاری)

منشأ داده‌ها باید مشخص و تطابق آن‌ها با جمعیت هدف تحلیل شود. اگر ادعای انطباق بین مراکز/دستگاه‌ها (Cross-center/Cross-device) وجود دارد، باید مشخص شود داده‌های آموزشی از کدام مراکز و با چه برند یا مدل تجهیزات جمع‌آوری شده و شواهد مربوط به این ادعا ارائه گردد.

استاندارد ۱-۹: فرآیند برچسب‌گذاری مجموعه آزمون و بازتولیدپذیری آن

متقاضی باید فرایند برچسب‌گذاری مجموعه آزمون را کامل مستند کند: ابزار و نسخه، مراحل دستی/نیمه‌خودکار، استخراج برچسب از گزارش‌ها (در صورت استفاده از NLP)، پروتکل QC برچسب‌ها، و معیار توافق بین ارزیابان (inter-rater). اگر ابزارها/سیستم‌عامل‌های لایسنس‌دار استفاده شده‌اند، وضعیت دسترسی/مالکیت و اثر احتمالی آن بر بازتولیدپذیری باید شفاف شود.

استاندارد ۱-۱۰: دسترسی و شرایط انتشار/اشتراک داده

اگر داده‌ها برای عموم یا افراد ثالث قابل دسترسی‌اند (یا قرار است باشند)، شرایط دسترسی، مجوزها، محل میزبانی (repository)، و محدودیت‌های استفاده (تحقیقاتی/تجاری) باید شفاف باشد و با الزامات محرمانگی/قراردادی سازگار ارائه شود.

استاندارد ۱-۱۱: تفکیک داده‌ها و کنترل نشتی (Train/Val/Test Split Integrity)

روش تفکیک داده‌ها به آموزش/اعتبارسنجی/آزمون باید شفاف، قابل بازتولید و ضدنشتی باشد (مثلاً patient-level یا center-level). اسکرپت split, seed و فهرست شناسه‌ها/ایندکس‌ها باید ارائه شود تا امکان راستی‌آزمایی وجود داشته باشد.

استاندارد ۱-۱۲: تعیین اندازه نمونه آزمون و تحلیل توان (Sample Size/Power)

متقاضی باید روش تعیین اندازه نمونه آزمون را ارائه کند (تحلیل توان یا برآورد مبتنی بر متریک هدف مانند AUC/حساسیت/ویژگی). تفسیر کفایت متریک‌ها باید متناسب با کاربرد بالینی ادعا شده باشد (مثلاً تصمیم‌یار در برابر تشخیص قطعی).

استاندارد ۱-۱۳: مستندسازی پروتکل جمع‌آوری داده و مشخصات تجهیزات

پروتکل جمع‌آوری داده باید کامل و قابل بازتولید باشد و مشخصات تجهیزات/تنظیمات کلیدی (مثلاً CT/MRI sequence یا kVp، نرخ نمونه‌برداری صوت، نوع حسگر و دوره نمونه‌برداری IoT و...) را پوشش دهد. اگر تجهیزات In-house توسط توسعه‌دهنده ساخته شده‌اند، شواهد فنی و کالیبراسیون/اعتبارسنجی عملکرد تجهیزات باید ضمیمه شود.

حیطه ۲) اعتبارسنجی تحلیلی/فنی (Analytical/Technical Validation)

هدف این حیطه اطمینان از صحت، پایداری، تکرارپذیری و بازتولیدپذیری عملکرد مدل در سطح فنی پیش از ورود به ارزیابی بالینی است.

استاندارد ۲-۱: تفکیک داده‌های آموزش و آزمون

ساختار تفکیک داده‌ها (train/validation/test) باید شفاف، قابل بازتولید و ضدنشتی باشد و تدابیر کنترل data leakage و کنترل سوگیری/توازن جمعیتی در تفکیک مستند شود.

استاندارد ۲-۲: تعریف و کنترل ورودی/خروجی

ورودی‌ها و خروجی‌های مدل باید کامل، بدون ابهام و قابل آزمون تعریف شوند؛ حداقل شرایط لازم ورودی (فرمت/کیفیت/رزولوشن/متادیتا و...) مشخص باشد و سازوکار تشخیص/رد ورودی نامعتبر پیاده‌سازی و تست شده باشد.

استاندارد ۲-۳: درستی پردازش پایپ‌لاین

مقتضای باید شواهدی ارائه کند که نشان دهد پایپ‌لاین پردازش داده (صرف‌نظر از یادگیری مدل) بدون باگ فنی کار می‌کند و داده ورودی بدون تحریف/خطای پنهان به مدل می‌رسد (sanity check). همچنین باید متریک‌های ارزیابی مناسب و نتایج عملکرد (و در صورت ارائه، شواهدی از تعمیم‌پذیری روی داده‌های دیده‌نشده) گزارش شود.

استاندارد ۲-۴: تکرارپذیری و بازتولیدپذیری

اگر مؤلفه‌های تصادفی وجود دارد، پایداری نتایج در تکرارهای مختلف با seed ثابت باید مستند شود؛ اگر مدل/استنتاج قطعی است، باید نشان داده شود در محیط‌ها/سخت‌افزارهای متفاوت (CPU/GPU، کتابخانه‌ها) خروجی‌های سازگار تولید می‌شود و اختلاف‌ها در محدوده قابل قبول است.

استاندارد ۲-۵: مدیریت نسخه‌ها و ردیابی تغییرات

نسخه‌های مدل، داده و کد باید به‌صورت شفاف و قابل پیگیری مدیریت شوند؛ هر نسخه باید شماره/تاریخ/تغییرات کلیدی داشته باشد و اثر تغییرات بر متریک‌های عملکرد قابل ردیابی باشد (Changelog).

استاندارد ۲-۶: پایداری عددی و حساسیت به تغییر ورودی

پایداری مدل در برابر تغییرات قابل انتظار ورودی/آرتیفک‌های مرتبط با ماهیت سیستم (نویز دستگاه، افت کیفیت، قطعی شبکه و...) باید ارزیابی شود و متریک‌های قبل و بعد گزارش گردد. اگر ارزیابی حملات خصمانه مطرح است، روش و نتایج باید شفاف ارائه شود.

استاندارد ۲-۷: محیط و قالب رابط کاربری

محیط عملیاتی محصول (وب/دسکتاپ/API و...) باید مشخص باشد؛ حداقل نیازمندی‌های سخت‌افزاری/نرم‌افزاری (System Requirements)، محدودیت‌های دسترسی، و مستندات طراحی UI/UX یا API باید ارائه شود و با نوع کاربرد محصول سازگار باشد.

استاندارد ۲-۸: مدیریت خطا و رفتار ایمن رابط کاربری

در مواجهه با ورودی ناقص/نامعتبر یا خطاهای عملیاتی، سیستم باید رفتار ایمن داشته باشد: پیام خطای قابل فهم، رد ورودی یا تولید خروجی کنترل شده، عدم crash و ثبت رخداد در لاگ. سناریوهای خطا باید آزمون و مستند شوند.

استاندارد ۲-۹: مستندسازی آزمون‌ها و شواهد

نتایج آزمون‌ها باید ساختارمند، نسخه‌دار و قابل ردیابی به اجرای مدل ذخیره شوند (زمان اجرا، نسخه مدل، متریک‌ها، داده آزمون/شناسه‌ها). اگر بازتولید آزمون‌ها به دلیل ماهیت داده‌های واقعی محدود است، حداقل مستندات جایگزین باید به صورت شفاف و قابل ممیزی ارائه شود.

حیطه ۳) اعتبارسنجی بالینی/کارکردی (Clinical/Functional Validation)

هدف این حیطه اطمینان از اینکه خروجی مدل با هدف بالینی ادعا شده ارتباط معنی‌دار دارد و در جمعیت و شرایط واقعی کاربرد، عملکرد قابل قبول و ایمن ارائه می‌دهد. مطالعه پیش از استقرار (گذشته‌نگر یا آینده‌نگر) با مقایسه عملکرد مدل با روش مرجع یا متخصص انسانی و ثبت شواهد قابل ممیزی دامنه این حیطه است.

استاندارد ۳-۱: طراحی مطالعه اعتبارسنجی

متقاضی باید طراحی مطالعه اعتبارسنجی را به صورت کامل و مستند ارائه کند؛ شامل نوع مطالعه (retrospective/prospective)، مراکز و جامعه آماری، حجم نمونه، معیارهای ورود/خروج، و روش‌های کنترل کیفیت داده و کاهش سوگیری (از جمله آنوتیشن، خوانش کور/مستقل، و در صورت لزوم تصادفی‌سازی). طراحی مطالعه باید با سطح ریسک محصول و ادعای کاربرد آن سازگار باشد.

استاندارد ۳-۲: تعریف هدف بالینی، معیار طلایی و مقایسه با استاندارد مراقبت

هدف دقیق بالینی، معیار طلایی و روش مقایسه باید شفاف تعریف شود. باید مشخص باشد عملکرد مدل با چه comparatorی سنجیده شده است (متخصص انسانی/روش مرجع/استاندارد مراقبت)، نوع فرضیه مطالعه (برتری/عدم نقص/non-inferiority/هم‌ارزی) و سازوکار جلوگیری از سوگیری در مقایسه (خوانش مستقل/کور) چگونه بوده است.

استاندارد ۳-۳: جمعیت و شرایط آزمون و نمایندگی محیط واقعی

داده‌های آزمون باید با محیط واقعی استفاده مدل هم‌خوان باشد. متقاضی باید توصیف دقیقی از شرایط آزمون ارائه کند: مراکز، نوع/مدل دستگاه‌ها و نسخه تجهیزات، توزیع دموگرافیک (سن، جنس و...)، ویژگی‌های بالینی

و نرخ شیوع/پیش‌آمد (prevalence) در داده‌ها. نتیجه‌گیری مطالعه باید از نظر تعمیم‌پذیری به محیط هدف «منطقی و مستدل» باشد.

استاندارد ۳-۴: متریک‌ها و تحلیل عملکرد

متریک‌های عملکرد باید متناسب با کاربرد بالینی انتخاب و به‌صورت کامل گزارش شوند؛ شامل مقادیر دقیق شاخص‌ها (مثل Sensitivity/Specificity/F1/AUROC و...) همراه با فاصله اطمینان ۹۵٪. در صورت انجام تحلیل زیرگروه (سن، جنس، مرکز، ...)، نتایج باید شفاف ارائه و تفسیر شوند.

استاندارد ۳-۵: تحلیل آماری و کفایت شواهد

کفایت حجم نمونه و توان آماری باید مستند شود. متقاضی باید روش Power analysis و پارامترها (α ، β ، Δ ، prevalence یا اثر مورد انتظار) را ارائه دهد و نشان دهد تعداد واقعی موارد مثبت/منفی با نیاز محاسباتی سازگار است. روش‌های آماری باید معتبر و متناسب با طراحی مطالعه باشند.

استاندارد ۳-۶: گزارش محدودیت‌ها، نقاط ضعف و موارد شکست

محدودیت‌های مدل باید صریح و قابل استفاده برای تصمیم‌گیری بالینی گزارش شود: شرایطی که عملکرد افت می‌کند، سناریوهای شکست و نمونه‌های واقعی failure cases (در حدی که محرمانگی و قوانین اجازه دهد). این محدودیت‌ها باید با دستورالعمل استفاده/IFU سازگار و در آن منعکس شود.

استاندارد ۳-۷: معیارهای عدالت سازگار با کاربرد

در صورتی که معیارهای عدالت/عدم تبعیض گزارش می‌شوند، باید روشن شود «این معیار» چگونه به پیامد بالینی مرتبط می‌شود؛ یعنی اختلاف آماری در عملکرد زیرگروه‌ها چگونه می‌تواند به تصمیم‌گیری/آسیب بالینی ترجمه شود و چه آستانه‌هایی از نظر بالینی قابل قبول/غیرقابل قبول است (متناسب با سناریو؛ مثلاً در تریاژ حساسیت مهم‌تر است).

استاندارد ۳-۸: تحلیل علل و برنامه اقدام اصلاحی

در صورت مشاهده اختلاف عملکرد، خطاهای تکرارشونده یا شکاف زیرگروهی، متقاضی باید تحلیل ریشه‌ای (RCA) ارائه دهد و اقدامات اصلاحی/کاهش‌ی را به‌همراه برنامه اجرا، زمان‌بندی، منابع و معیار سنجش موفقیت مشخص کند (مثلاً جمع‌آوری داده تکمیلی، بازآموزی، وزن‌دهی، تغییر دامنه کاربرد، یا تغییر در IFU).

حیطه ۴) پایش پس‌ازار و عملکرد دنیای واقعی (PMS/PMCF & Real-World Performance)

هدف این حیطه ایجاد یک نظام مستمر برای گردآوری، تحلیل و اقدام اصلاحی بر مبنای داده‌های عملکرد واقعی پس از صدور مجوز، به‌منظور تضمین پایداری، ایمنی و اثربخشی مدل و شناسایی/مدیریت انحرافات عملکرد است.

استاندارد ۴-۱: چارچوب کلی PMS/PMCF پس از صدور مجوز

توسعه‌دهنده باید برنامه‌ای رسمی، مکتوب و قابل اجرا برای پایش پسابازار (PMS) و در صورت نیاز پیگیری بالینی پسابازار (PMCF) ارائه کند که شامل هدف‌ها، مسئولیت‌ها، زمان‌بندی، منابع، روش نگهداری سوابق و مسیر رسیدگی به مشکلات (CAPA) باشد. همچنین باید مدارک آموزشی/راهنمای کاربری ویژه کاربران نهایی (کادر درمان یا سایر کاربران) ارائه شود به‌گونه‌ای که نیاز به آشنایی عمیق با پیچیدگی‌های AI نداشته باشند.

استاندارد ۴-۲: سنجش عملکرد بالینی/عملیاتی با KPIهای مشخص

توسعه‌دهنده باید شاخص‌های کلیدی عملکرد (KPI) مدل را مشخص کند (مثل دقت، حساسیت، اختصاصیت، FPR، زمان پاسخ و...)، و برای هر شاخص فرمول محاسبه، آستانه قابل قبول، دوره پایش و منطق انتخاب بازه زمانی را ارائه دهد. نتایج پایش باید به‌صورت روندی (trend) رصد شود و هر نتیجه به نسخه مدل لینک شود.

استاندارد ۴-۳: کشف و مدیریت Model Drift (داده/مفهوم)

توسعه‌دهنده باید فرآیندی مستند برای پایش، کشف و مدیریت Drift ارائه دهد که شامل: تعریف انواع drift (Data/Concept)، تعیین توزیع پایه (Baseline)، انتخاب روش‌های آماری/تحلیلی معتبر برای تشخیص drift، تعیین آستانه‌های تصمیم‌گیری، تعیین دوره پایش و سیاست واکنش در صورت کشف drift (هشدار، محدودسازی خروجی، بازآموزی، بازنگری یا توقف کنترل‌شده) باشد. همچنین باید سوابق driftها، تحلیل ریشه‌ای و اقدامات اصلاحی مرتبط قابل رهگیری و آرشیو شوند.

استاندارد ۴-۴: مدیریت رخدادهای ناخواسته، عوارض و خطاهای بحرانی

توسعه‌دهنده باید برنامه‌ای مستند برای مدیریت ایمنی و ریسک پسابازار ارائه کند که شامل: تعریف انواع ریسک‌ها (ایمنی بیمار/خطاهای بحرانی/پیامدهای بالینی)، شاخص‌های هشدار زودهنگام (EWI)، روش جمع‌آوری و تحلیل رخدادها/خطاها/شکایات، مسیر گزارش‌دهی داخلی و در صورت نیاز گزارش‌دهی به مراجع، و CAPA باشد. همچنین پیوند رخداد با نسخه مدل و زمان وقوع باید حفظ شود.

استاندارد ۴-۵: پیوند نتایج مدل با نتایج بالینی/نتایج نهایی

توسعه‌دهنده باید فرآیندی مستند برای اتصال پیش‌بینی‌های مدل به نتایج واقعی/ground truth ارائه دهد تا عملکرد واقعی صرفاً «آماری» نباشد و به پیامد بالینی مرتبط شود. این فرآیند باید علاوه بر شاخص‌های کلاسیک، شامل پایش «تعامل بالینی» نیز باشد؛ مانند نرخ پذیرش/رد پیشنهاد مدل توسط پزشک، میزان هم‌راستایی خروجی مدل با تصمیم نهایی و مدیریت داده‌های ناقص در لینک‌کردن پیامدها.

استاندارد ۴-۶: تغییرات و کنترل نسخه

توسعه‌دهنده باید فرآیند مستند کنترل نسخه و مدیریت تغییرات مدل ارائه دهد؛ شامل تعریف نسخه‌ها، ثبت تغییرات هر نسخه (الگوریتم/داده/پیش‌پردازش/پارامترها)، تفکیک تغییرات کوچک و بزرگ، فرآیند بازبینی و تأیید پیش از انتشار، و حفظ ارتباط هر نسخه با داده آموزشی، عملکرد گذشته و نتایج PMS/PMCF. سوابق تغییرات باید آرشیو شود و در ممیزی قابل ارائه باشد.

استاندارد ۴-۷: حریم خصوصی، حفظ داده و انطباق قانونی

توسعه‌دهنده باید برنامه‌ای مستند برای حفاظت از حریم خصوصی و امنیت اطلاعات ارائه کند که شامل: سیاست‌های حفاظت داده‌های حساس، روش‌های رمزگذاری و pseudonymization، کنترل دسترسی محدود و قابل ردیابی، ذخیره‌سازی امن، نگهداری سوابق مجوزها و ممیزی‌ها، و انطباق با الزامات قانونی مرتبط باشد. در مواردی که گواهی داخلی برای برخی چارچوب‌های بین‌المللی وجود ندارد، ارائه مستندات فنی مبنی بر رعایت اصول و انطباق کفایت دارد. در این حیطه، تمرکز بر الزامات حفاظت داده در فرآیندهای پایش و گزارش‌دهی پس‌بازار است.

استاندارد ۴-۸: تعامل با کاربران نهایی و مدیریت بازخورد

توسعه‌دهنده باید فرآیندی مستند برای جمع‌آوری، ثبت، تحلیل و اقدام بر اساس بازخورد کاربران ارائه کند (نظرسنجی، فرم گزارش خطا، کانال پشتیبانی و...)، و نشان دهد بازخورد چگونه به بهبود مدل/رابط کاربری تبدیل می‌شود. همچنین شاخص‌های مرتبط با رضایت و قابلیت استفاده (Usability metrics) باید تعریف و به‌صورت دوره‌ای گزارش شود و سوابق تغییرات ناشی از بازخورد قابل پیگیری باشد.

حیطه ۵) اخلاق و امنیت داده (Ethics, Data Protection & Security)

تضمین استقرار بالینی ایمن، اخلاق‌محور و قابل اعتماد محصولات هوش مصنوعی در سلامت از طریق رعایت الزامات حفاظت داده، امنیت سایبری، پایش پیامدهای ناعادلانه و سازوکارهای نظارت انسانی، پاسخگویی و شفافیت امکان‌پذیر است. در این حیطه تأکید می‌شود که الزامات اخلاقی و امنیتی صرفاً مربوط به مرحله پیش از صدور گواهی نیستند و باید در طول چرخه عمر محصول استمرار داشته باشند.

استاندارد ۵-۱: حفاظت از داده و انطباق حقوقی

پردازش داده‌های سلامت در محصول باید بر مبنای حقوقی/قراردادی مشخص انجام شود و الزامات حریم خصوصی رعایت گردد. توسعه‌دهنده موظف است چرخه حیات داده (جمع‌آوری، ذخیره‌سازی، پردازش، اشتراک/انتقال و حذف) را شفاف و قابل ممیزی مستند کند و در موارد لازم، ارزیابی اثرات حریم خصوصی (DPIA) یا سند معادل را ارائه دهد.

استاندارد ۵-۲: امنیت سایبری و تاب‌آوری عملیاتی

محصول باید دارای الزامات امنیت سایبری پایه و قابل اتکا باشد تا از داده، مدل و سرویس در برابر تهدیدات رایج محافظت کند و در صورت بروز اختلال، تداوم خدمت و رفتار ایمن تضمین شود. توسعه‌دهنده باید حداقل کنترل‌های امنیتی و سازوکارهای پایداری سرویس را مستند و قابل ممیزی ارائه نماید.

استاندارد ۵-۳: انصاف الگوریتمی و پیامدهای ناعادلانه

توسعه‌دهنده باید نشان دهد پس از استقرار، پیامدهای واقعی مدل بر گروه‌های مختلف کاربران/بیماران پایش می‌شود و سازوکار شناسایی، گزارش و اصلاح پیامدهای ناعادلانه وجود دارد. این الزامات ناظر به مرحله استقرار است و باید در پایش پس‌بازار به‌صورت مستمر پیگیری شود.

استاندارد ۵-۴: نظارت انسانی، پاسخگویی و شفافیت

استقرار محصول باید همراه با امکان نظارت مؤثر انسان، تصمیم‌گیری آگاهانه و سازوکار پاسخگویی باشد. نقش انسان در فرایند (ناظر/تأییدکننده)، امکان override تصمیم مدل، مسئولیت‌ها و مسیر رسیدگی به خطا/اعتراض باید شفاف و مستند باشد و محدودیت‌ها و شرایط استفاده ایمن به کاربر اعلام گردد.