AURA — AUTONOMOUS UNIVERSAL RESIDENTIAL AGENT

گزارش جامع پژوهش علمی، تحلیل رقبا، معماری فنی و استراتژی محصول

طراحی و استراتژی عامل هوشمند خودمختار برای زندگی روزمره با نقطه شروع از خانه

بررسی عمیق معماری‌های شناختی دوگانه (System 1 & System 2)، مدل‌های جهان و پیش‌بینی در فضای نهفته (JEPA / V-JEPA 2)، هوش مصنوعی تجسم‌یافته (Embodied AI)، زیرساخت‌های کنترل محلی خانه هوشمند و نقشه راه تکاملی از هسته نرم‌افزاری تا رباتیک خانگی

تدوین‌شده توسط تیم چندتخصصی پژوهش هوش مصنوعی، رباتیک، معماری سیستم و استراتژی محصول | تاریخ انتشار و راستی‌آزمایی منابع: اکتبر ۲۰۲۶

۱. تحلیل و ساختاردهی یادداشت‌های دست‌نویس اولیه پروژه AURA

نقطه عزیمت این پژوهش، واکاوی دقیق دو صفحه یادداشت دست‌نویس فارسی است که چارچوب مفهومی، چالش‌های رقابتی و معماری پنج‌لایه اولیه پروژه **AURA (Autonomous Universal Residential Agent)** را ترسیم کرده‌اند. بررسی خط‌به‌خط این دست‌نوشته‌ها نشان می‌دهد که ایده اصلی بر یک بینش راهبردی دقیق استوار است: ساخت یک «سیستم‌عامل و عامل هوشمند برای زندگی روزمره» که از محیط خانه آغاز می‌شود، اما به گونه‌ای ماژولار طراحی می‌گردد که قابلیت گسترش به دفاتر اداری، شرکت‌ها و کارخانه‌ها را نیز داشته باشد. در این دیدگاه، تکامل محصول به جای پرش ناگهانی و پرریسک به ساخت سخت‌افزار پیچیده، از یک هسته نرم‌افزاری و کنترل زیرساخت موجود آغاز شده و به تدریج به عامل‌های فیزیکی و ربات‌های انسان‌نما متصل می‌شود.

۱.۱. استخراج ساختاریافته محتوای دست‌نوشته اول (صفحه ۱: هویت محصول و تحلیل رقبا)

در صفحه نخست یادداشت‌ها، ساختار فکری حول محور تعریف هویت اصلی (بخش شماره ۱ — Main) و استخراج درس‌های کلیدی از سه بازیگر شاخص صنعتی شکل گرفته است:

- **هویت و دامنه گسترش محصول:** عبارت «سیستم‌عامل هوشمند برای زندگی روزمره» با دو شاخه خروجی تعریف شده است: شاخه اول «شروع: خونه» و شاخه دوم «می‌تونه به دفاتر، شرکت و کارخونه‌ها هم بره!». در زیر این بخش، نام پیشنهادی **AURA — Autonomous Universal Residential Agent** ثبت گردیده است.
- **بررسی نمونه اول — ربات انسان‌نمای 1X-NEO:** سه آموزه محوری از رویکرد شرکت 1X استخراج شده است: (۱) «رابط طبیعی به جای پیچیدگی کنترل ربات» تا کاربران عادی نیازی به دانش رباتیک نداشته باشند؛ (۲) «ترکیب خودمختاری با کمک انسان هنگام شکست» که اشاره‌ای مستقیم به مکانیزم هدایت از راه دور (Teleoperation / Expert Mode) در مواقعی است که مدل هوش مصنوعی در محیط خانه با شرایط ناشناخته مواجه می‌شود؛ و (۳) «اضافه شدن تدریجی فیچرها» به مرور زمان و همراه با یادگیری مستمر سیستم.
- **بررسی نمونه دوم — شرکت Figure AI و مدل Helix:** یادداشت‌ها بر سه ویژگی کلیدی مدل کنترل رباتیک Figure تمرکز دارند: (۱) «مدل عمومی دارای چند مهارت قابل ارزیابی» به جای اسکریپت‌های تک‌منظوره؛ (۲) «داده‌های انسانی تعمیم به محیط جدید را بهتر می‌کند» که بیانگر نقش داده‌های نمایشی و پیش‌آموزش مقیاس بزرگ در تطبیق با خانه‌های جدید است؛ و (۳) «موفقیت کامل یک تسک نه فقط حرکات نمایشی» که بر اهمیت تکمیل سرتاسری (End-to-End Task Completion) تأکید دارد.
- **بررسی نمونه سوم — زیرساخت کنترل خانه Home Assistant:** سه اصل بنیادین از اکوسیستم متن‌باز Home Assistant استخراج شده است: (۱) «سازگاری بین برندها مهم‌تر از همه سخت‌افزارها» (علامت > با عبارت «مهم‌تر» بالای آن)؛ (۲) «کنترل محلی می‌تواند حریم خصوصی و کارکرد بدون اینترنت را بهبود دهد»؛ و (۳) نقد تجربه کاربری فعلی با این گزاره کلیدی که «کاربر نباید برای استفاده عادی مجبور به کدنویسی یا تنظیمات پیچیده باشد».

۱.۲. استخراج ساختاریافته محتوای دست‌نوشته دوم (صفحه ۲: ادامه رقبا و معماری ۵ لایه)

در صفحه دوم یادداشت‌ها، ابتدا تحلیل رقبا با بررسی رویکرد پژوهشی Google DeepMind تکمیل شده و سپس معماری پیشنهادی پنج‌لایه با ارتباطات رفت‌وبرگشتی ترسیم گردیده است:

- **بررسی نمونه چهارم — Google DeepMind (مغز هوشمند برای عمل فیزیکی):** سه اصل معماری حیاتی یادداشت شده است: (۱) «مدل زبانی به تنهایی موتور کنترل نیست»؛ (۲) «برنامه‌ریزی، درک محیط، کنترل حرکتی و تشخیص موفقیت باید با هم کار کنند»؛ و (۳) «ابزارهای تخصصی می‌توانند به جای یک مدل همه‌کاره استفاده شوند».
- **معماری پیشنهادی ۵ لایه در دست‌نوشته:**
  - **لایه ۱ (رابط کاربری):** صدا، متن، تصویر و رابط ساده برای همه اعضای خانواده (متصل با فلش رو به پایین به لایه ۲).
  - **لایه ۲ (مغز ایجنت):** درک، برنامه‌ریزی، حافظه، انتخاب ابزار (متصل با فلش رو به پایین به لایه ۳).
  - **لایه ۳ (موتور سیاست و ایمنی):** مجوزها، تأیید عملیات حساس، محدودیت‌ها و توقف اضطراری (دارای دو فلش خروجی مجزا به لایه‌های A-۴ و B-۴).
  - **لایه ۴ (لایه اجرا در دو شاخه موازی):** شاخه **A-۴ (دستگاه‌ها)** شامل روشنایی، تهویه، قفل، حسگر و لوازم خانگی؛ و شاخه **B-۴ (عامل‌های فیزیکی)** شامل جارو، ربات متحرک، بازو و انسان‌نما.

- **لایه ۵ (بررسی نتیجه و یادگیری):** متصل با دو فلش دوطرفه (پایین‌رو و بالارو ↓↑) به لایه ۴، شامل تأیید اجرای کار، گزارش خطا، ثبت ترجیحات و بهبود عملکرد.

۱.۳. جدول ابهام‌زدایی کلمات دست‌نویس برای تأیید نهایی کاربر

مطابق استاندارد دقت علمی و به منظور پرهیز از هرگونه حدس قطعی در موارد دارای دست‌خط پیوسته، چهار عبارت در دست‌نوشته‌ها شناسایی شده‌اند که خوانش اول و گزینه‌های جایگزین آن‌ها در جدول زیر جهت تأیید نهایی شما علامت‌گذاری شده‌اند:

محل در دست‌نوشته	خوانش اصلی (مبنای گزارش)	گزینه‌های احتمالی دیگر	توضیح فنی و تأثیر بر معماری AURA
صفحه ۱، بالا سمت چپ	**«به دفاتر، شرکت و کارخونه‌ها هم بره!»**	«به دفتر، شرکت...» یا «به دفاتر، فروشگاه و کارخونه‌ها...»	در هر دو حالت، منظور گسترش افقی معماری از محیط مسکونی (Residential) به محیط‌های تجاری و صنعتی (Commercial & Enterprise) در فازهای بلندمدت است.
صفحه ۱، بخش Figure AI	**«Helix 2.5» یا «Helix 02» -**	«Helix S1-S2» یا «Helix 02» (نسخه دوم Helix معرفی‌شده در ژانویه ۲۰۲۶)	شرکت Figure AI در فوریه ۲۰۲۵ مدل Helix (معماری S1/S2) و در ژانویه ۲۰۲۶ نسخه Helix 02 را برای کنترل کل بدن معرفی کرد. عدد دست‌نویس پس از Helix نیاز به تأیید شما دارد.
صفحه ۱، روبه‌روی نام Figure AI	**«انسان‌نما با 3 تسک (یک محور)»**	«انسان‌نما با 3 تسک (یک موتور / یک مدل)» یا «2 تسک (یک مدل)»	در مستندات Figure Helix، تأکید اصلی بر انجام چندین تسک متفاوت با «یک مدل واحد و وزن‌های شبکه یکسان» (Single unified model weights) بدون نیاز به آموزش جداگانه برای هر مهارت است.
صفحه ۱، بخش Home Assistant	**«سازگاری بین برندها > [مهم‌تر از] همه سخت‌افزارها»**	«سازگاری بین برندها > همه سخت‌افزارها» (به معنای پوشش همه سخت‌افزارها)	کلمه کوچک «مهم‌تر» بالای علامت > نوشته شده است که نشان می‌دهد سازگاری نرم‌افزاری فرادستگاهی بر وابستگی به یک سخت‌افزار خاص اولویت دارد.

۲. تحقیق عمیق علمی و فنی درباره فناوری‌های بنیادین

طراحی یک عامل هوشمند که بتواند از سطح مکالمه متنی فراتر رفته و در فضای فیزیکی خانه تصمیم‌گیری و اقدام کند، نیازمند پیوند شش حوزه پژوهشی پیشرو در هوش مصنوعی و رباتیک است. در این بخش، هر یک از این حوزه‌ها بر اساس منابع اولیه دانشگاهی و صنعتی واکاوی می‌شوند.

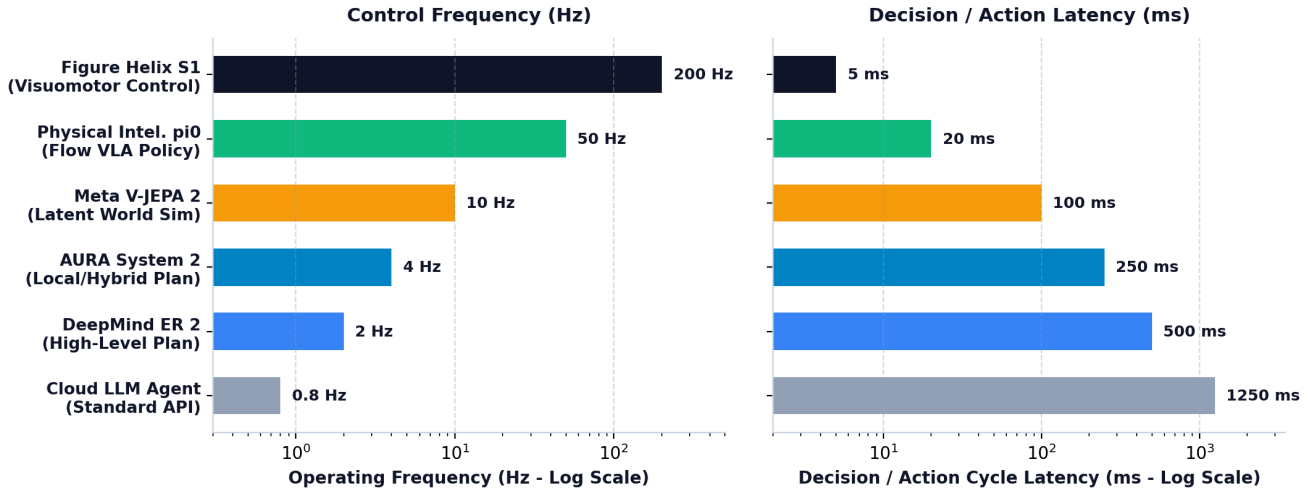
۲.۱. تفکر System 1 در برابر System 2 در معماری عامل‌ها و رباتیک

در علوم شناختی و اقتصاد رفتاری، دنیل کانمن (Kahneman, 2011) در کتاب مرجع *Thinking, Fast and Slow*، فرآیندهای شناختی انسان را به دو حالت متمایز تقسیم می‌کند: **سیستم ۱ (System 1)** که به صورت سریع، خودکار، موازی، بی‌نیاز از تلاش آگاهانه و مبتنی بر الگوهای غریزی/شهودی عمل می‌کند؛ و **سیستم ۲ (System 2)** که کند، ترتیبی، تحلیلی، نیازمند تمرکز و مسئول استدلال منطقی، برنامه‌ریزی چندمرحله‌ای و خودکنترلی است.

در مهندسی عامل‌های هوشمند (Agentic AI) و رباتیک مدرن، این نظریه به یک الگوی معماری بسیار کارآمد تبدیل شده است. پژوهش Christakopoulou et al. (2024, arXiv:2410.08328) تحت عنوان *Agents Thinking Fast and Slow: A Talker-Reasoner Architecture* نشان می‌دهد که وقتی یک عامل هوشمند همزمان مسئول تعامل بلادرنگ با کاربر و انجام برنامه‌ریزی‌های پیچیده روی حافظه است، استفاده از یک مدل واحد باعث افزایش شدید تأخیر (Latency) یا کاهش دقت استدلال می‌شود. در معماری Talker-Reasoner، عامل **Talker (سیستم ۱)** یک مدل سبک و سریع است که بلافاصله به کاربر

پاسخ می‌دهد و وضعیت‌های کش‌شده در حافظه را می‌خواند، در حالی که عامل **Reasoner (سیستم ۲)** به صورت ناهمگام (Asynchronous) در پس‌زمینه به استدلال چندمرحله‌ای، فراخوانی ابزارهای خارجی، و به‌روزرسانی گراف حافظه و باورهای سیستم می‌پردازد.

اهمیت این تفکیک در رباتیک فیزیکی دوچندان است. یک مدل بینایی-زبانی بزرگ (VLM) میلیاردپارامتری می‌تواند دستور پیچیده‌ای مانند «وسایل صبحانه را از روی میز جمع کن و هر کدام را سر جایش بگذار» را درک و به زیرکارها تجزیه کند، اما زمان استنتاج (Inference Latency) آن معمولاً بین ۲۰۰ تا ۱۰۰۰ میلی‌ثانیه (فرکانس ۱ تا ۵ هرتز) است. در مقابل، کنترل پایدار مفاصل یک بازوی رباتیک یا حفظ تعادل پویا و تطبیق لحظه‌ای انگشتان دست با لغزش یک لیوان شیشه‌ای، نیازمند حلقه کنترلی با فرکانس ۵۰ تا ۲۰۰ هرتز (تأخیر ۵ تا ۲۰ میلی‌ثانیه) است. معماری Helix در شرکت Figure AI (فوریه ۲۰۲۵) دقیقاً بر همین مبنا طراحی شده است: یک مدل بینایی-زبانی ۷ میلیارد پارامتری متن‌باز به عنوان **System 2 (S2)** با فرکانس ۷ تا ۹ هرتز صحنه و فرمان زبانی را پردازش کرده و یک بردار بازنمایی نهفته پیوسته تولید می‌کند؛ و یک سیاست بینایی-حرکتی (Visuomotor Policy) ۸۰ میلیون پارامتری مبتنی بر ترنسفورمر به عنوان **System 1 (S1)** این بردار را با فرکانس ۲۰۰ هرتز به فرمان‌های پیوسته ۳۵ درجه آزادی بالاتنه، مچ‌ها و انگشتان تبدیل می‌نماید.



شکل ۱: مقایسه لگاریتمی فرکانس حلقه کنترل (هرتز) و تأخیر تصمیم‌گیری (میلی‌ثانیه) میان لایه‌های برنامه‌ریزی تحلیلی (System 2) و سیاست‌های واکنشی سریع (System 1) بر اساس داده‌های مستندات Figure Helix، Physical Intelligence π0 و DeepMind ER 2.

محدودیت‌های علمی تشبیه System 1 و System 2 در هوش مصنوعی: از دیدگاه داوری علمی، باید توجه داشت که این نام‌گذاری در مهندسی نرم‌افزار و رباتیک یک «استعاره عملکردی برای جداسازی مقیاس زمانی و محاسباتی» (Temporal & Computational Decoupling) است و تفاوت‌های مهمی با سیستم شناختی انسان دارد: (۱) در مغز انسان، سیستم ۱ حاصل میلیون‌ها سال تکامل زیستی و یادگیری حسی-حرکتی غنی است، در حالی که یک مدل کوچک زبانی یا شبکه سیاست در هوش مصنوعی فاقد درک ذاتی فیزیکی است؛ (۲) در معماری‌های دوگانه مصنوعی، «گلوگاه ارتباطی» (Latent Bottleneck) میان S2 و S1 می‌تواند در هنگام تغییرات ناگهانی محیط باعث ایجاد پدیده *وضعیت کهنه (Stale Belief State)* شود؛ و (۳) برخلاف انسان که به طور پیوسته مرز توجه را میان دو سیستم جابه‌جا می‌کند، در سیستم‌های مهندسی نیازمند یک مکانیزم داوری صریح (Arbitration & Safety Interrupt) هستیم تا در شرایط اضطراری، فرمان‌های در حال اجرای S2 بلافاصله متوقف شوند.

۲.۲. مدل‌های جهان (World Models) و پیش‌بینی پیامد فیزیکی اقدامات

مفهوم **مدل جهان (World Model)** در یادگیری ماشین — که صورت‌بندی مدرن آن با مقاله کلاسیک Ha & Schmidhuber (2018, arXiv:1803.10122) و معماری شناختی Yann LeCun (2022, OpenReview) شناخته می‌شود — به یک مدل درونی در ذهن عامل اشاره دارد که دو کارکرد اساسی را محقق می‌سازد: نخست، فشرده‌سازی مشاهدات حسی پرحجم و پرنویز به یک «وضعیت نهفته معنادار» (s_t)؛ و دوم، پیش‌بینی وضعیت آینده جهان به شرط انجام یک عمل مشخص یا به عبارت ریاضی تخمین تابع انتقال وضعیت $s_{t+1} = f(s_t, a_t)$.

همان‌طور که مطالعات مروری اخیر در حوزه هوش تجسم‌یافته (Embodied AI: From LLMs to World Models, 2025, arXiv:2509.20021 و World Models for Embodied Intelligence, 2026, arXiv:2609.16697) نشان می‌دهند، تفاوت بنیادین یک مدل جهان با یک مدل زبانی بزرگ (LLM) در نوع دانش رمزگذاری‌شده در آن‌هاست. یک مدل زبانی صرفاً توزیع آماری

توکن‌های متنی را بر اساس هم‌رخدادی کلمات در متون اینترنتی یاد گرفته است ($P(w_t \mid w_{1:t-1})$). چنین مدلی ممکن است بداند که در متون انسانی کلمه «لیوان» با کلمه «شکستن» ارتباط دارد، اما هیچ بازنمایی هندسی، بردار سرعت، مرکز جرم، اصطکاک سطح یا پایداری فضایی اشیا (Object Permanence) را در خود ندارد. در مقابل، یک مدل جهان فیزیکی به عامل اجازه می‌دهد پیش از آنکه فرمانی را در دنیای واقعی اجرا کند، به صورت درونی شبیه‌سازی کند (Mental Simulation / Counterfactual Rollout) که «اگر بازوی ربات با این زاویه به سمت قوری حرکت کند، آیا دست ربات به لیوان کناری برخورد خواهد کرد یا خیر؟».

۲.۳. معماری JEPA و مدل V-JEPA 2: جهش در درک محیط و برنامه‌ریزی رباتیک

یکی از بزرگ‌ترین موانع در ساخت مدل‌های جهان مبتنی بر ویدیو، هزینه محاسباتی و خطای بازسازی در «مدل‌های مولد سطح پیکسل» (Pixel-Level Generative Models) است. وقتی مدلی تلاش می‌کند فریم بعدی یک ویدیو را پیکسل به پیکسل پیش‌بینی کند، مجبور است بخش عمده‌ای از ظرفیت شبکه عصبی خود را صرف پیش‌بینی جزئیات تصادفی و غیرقابل پیش‌بینی — مانند تکان خوردن برگ درختان در بیرون پنجره، بافت ریز روی فرش یا بازتابهای میکروسکوپی نور — نماید؛ جزئیاتی که هیچ نقشی در تصمیم‌گیری کنترلی یک ربات خانگی ندارند.

برای حل این معضل بنیادین، یان لکان معماری **JEPA (Joint-Embedding Predictive Architecture)** را مطرح کرد (LeCun, 2022). در JEPA، به جای پیش‌بینی خودِ سیگنال ورودی y از روی ورودی x در فضای پیکسل، هر دو ورودی ابتدا توسط رمزگذارها (Encoders) به فضای بازنمایی انتزاعی منتقل شده (s_x و s_y) و یک شبکه پیش‌بین (Predictor) با دریافت s_x و یک متغیر شرطی (مانند عمل ربات a)، بازنمایی انتزاعی وضعیت بعدی یعنی $\hat{s}_y$ را پیش‌بینی می‌کند. به این ترتیب، اطلاعات غیرضروری پیکسلی حذف شده و دینامیک فیزیکی و معنایی صحنه حفظ می‌گردد.

در ژوئن ۲۰۲۵، تیم پژوهشی Meta FAIR مدل **V-JEPA 2** را معرفی کرد (Assran et al., 2025, arXiv:2506.09985؛ مخزن رسمی در GitHub: facebookresearch/vjepa2). یافته‌های علمی و دستاوردهای کلیدی V-JEPA 2 که مستقیماً در طراحی آینده AURA کاربرد دارند عبارت‌اند از:

- **پیش‌آموزش خودنظارتی مقیاس بزرگ روی ویدیو (V-JEPA 2 Pre-training):** مدل پایه V-JEPA 2 (با ۱.۲ میلیارد پارامتر) روی بیش از ۱ میلیون ساعت ویدیوی اینترنتی و ۱ میلیون تصویر بدون برچسب، با هدف پیش‌بینی بخش‌های ماسک‌شده ویدیو در فضای بازنمایی نهفته آموزش دیده و به عملکرد پیشرو (State-of-the-Art) در درک حرکت، تعامل انسان و شیء، و پیش‌بینی کنش دست یافته است.
- **پس‌آموزش مدل جهان مشروط بر عمل (V-JEPA 2-AC) برای رباتیک:** پژوهشگران Meta نشان دادند که با ثابت نگه داشتن رمزگذار تصویری V-JEPA 2 و آموزش یک شبکه پیش‌بین مشروط بر عمل (Action-Conditioned Predictor) روی تنها **۶۲ ساعت داده ویدیویی بدون برچسب از ربات Franka (مجموعه داده Droid)**، مدل می‌تواند پیامد حرکات بازوی ربات را در فضای نهفته پیش‌بینی کند.
- **برنامه‌ریزی و کنترل Zero-Shot در محیط‌های ناآشنا:** با استفاده از الگوریتم برنامه‌ریزی مبتنی بر مدل (Model-Predictive Control / Cross-Entropy Method در فضای نهفته)، مدل V-JEPA 2-AC توانست روی بازوهای رباتیک در دو آزمایشگاه کاملاً جدید که هرگز در داده‌های آموزشی دیده نشده بودند، بدون هیچ‌گونه آموزش اضافی (Zero-Shot) وظایف رسیدن به هدف (Reaching)، برداشتن اشیای جدید (Grasping) و جابه‌جایی و قرار دادن اشیا (Pick-and-Place) را با موفقیت بالا اجرا کند.
- **محدودیت‌های واقعی فعلی V-JEPA 2:** بر اساس تحلیل خود مقاله Assran et al. (2025) و مطالعه مستقل Sobal et al. (2026, arXiv:2512.24497) درباره عوامل موفقیت برنامه‌ریزی فیزیکی با JEPA-WMs، سه محدودیت مهم هنوز پابرجاست: (۱) انباشت خطا در افق‌های زمانی بلندمدت (Long-horizon Rollout Drift) به طوری که برای وظایف چندمرحله‌ای طولانی نیازمند ترکیب با یک برنامه‌ریز سلسله‌مراتبی سطح بالا (System 2) است؛ (۲) حساسیت به زاویه دوربین و انسداد شدید دید (Occlusion)؛ و (۳) نبود حسگرهای لامسه و نیرو (Force/Tactile Feedback) در مدل پایه که برای کارهای ظریف مانند باز کردن قفل سفت یا برش مواد غذایی ضروری است.

۲.۴. هوش مصنوعی تجسم‌یافته (Embodied AI) و یادگیری مهارت‌های فیزیکی

هوش مصنوعی تجسم‌یافته بر این اصل استوار است که هوش واقعی برای عمل در جهان نیازمند یک کالبد فیزیکی است که در یک حلقه بسته حسگری-حرکتی (Perception-Action Loop) با محیط تعامل کند. در سال‌های ۲۰۲۴ تا ۲۰۲۶، سه پارادایم اصلی سبب جهش در یادگیری مهارت‌های فیزیکی ربات‌ها شده‌اند:

1. **مدل‌های Vision-Language-Action (VLA) با تولید پیوسته مسیر (Flow Matching):** شرکت Physical Intelligence با معرفی مدل پایه رباتیک π0 (Pi-Zero, اکتبر ۲۰۲۴, arXiv:2410.24164) و نسخه پیشرفته π0.7 (آوریل ۲۰۲۶) نشان داد که با اتصال یک اسکلت بینایی-زبانی پیش‌آموزش‌دیده به یک سرِ تولید عمل مبتنی بر *Flow Matching* (مشابه مدل‌های انتشار پیوسته)، می‌توان فرمان‌های حرکتی مفاصل را با فرکانس ۵۰ هرتز تولید کرد. این معماری روی ۷ پلتفرم رباتیک مختلف و ۶۸ وظیفه پیچیده خانگی — از جمله بیرون آوردن لباس‌ها از خشک‌کن و تا کردن آن‌ها، تمیز کردن میز و جمع کردن ظروف، و بسته‌بندی خریدهای روزمره — کارایی خود را اثبات کرده است.
2. **معماری دوگانه استدلال تجسم‌یافته و اجرای VLA:** در ژوئیه ۲۰۲۶، Google DeepMind خانواده مدل‌های Gemini Robotics 2 را معرفی کرد. در این معماری، مدل Gemini Robotics ER 2 (Embodied Reasoning 2) به عنوان مغز سطح بالا وظیفه درک فضایی صحنه، استدلال چندمرحله‌ای، تشخیص زمان دقیق شروع و پایان هر مرحله از کار (Progress Understanding) و خوداصلاحی در صورت خطا را بر عهده دارد و دستورالعمل‌های میانی را به مدل VLA سطح پایین برای کنترل کل بدن یا دستان ربات ارسال می‌کند.
3. **یادگیری تقویتی عمیق برای حرکت کل بدن (Whole-Body RL Mobility):** برای حرکت پایدار در محیط‌های واقعی خانه (پله، فرش، سطوح ناهموار)، شرکت 1X در مستندات Redwood AI Mobility (ژوئن ۲۰۲۵) نشان داد که یک کنترل‌گر مبتنی بر یادگیری تقویتی (RL) می‌تواند بدون وابستگی به سنسورهای گران‌قیمت LiDAR و صرفاً با اتکا به حسگرهای وضعیت مفاصل (Proprioception) و دوربین‌های استریو، راه‌رفتن طبیعی، نشستن، برخاستن و حفظ تعادل در برابر ضربات خارجی را مدیریت نماید.

۲.۵. هوش مصنوعی عاملی (Agentic AI)، ابزارها و کنترل دسترسی

در لایه نرم‌افزاری، یک «عامل هوشمند» (Agent) با سه مؤلفه از یک مدل زبانی غیرفعال متمایز می‌شود: **حافظه (Memory)**، **برنامه‌ریزی (Planning)** و **استفاده از ابزار (Tool Use)**. چارچوب مرجع ReAct (Yao et al., 2023, arXiv:2210.03629) نشان داد که وقتی مدل به جای تولید یکباره پاسخ، چرخه‌ای از «تفکر استدلالی (Thought) → اقدام با ابزار (Action) → مشاهده نتیجه محیط (Observation)» را طی می‌کند، نرخ توهم (Hallucination) به شدت کاهش یافته و توانایی حل مسائل چندمرحله‌ای در محیط‌های خانگی شبیه‌سازی‌شده (مانند ALFWorld) جهش می‌یابد.

در یک عامل مسکونی مانند AURA، معماری عاملی باید سه لایه حفاظتی و ساختاری اضافه داشته باشد: (۱) **سیستم چندعاملی هماهنگ (Multi-Agent Orchestration)** که در آن یک عامل ناظر (Supervisor) وظایف را میان زیرعامل‌های تخصصی (مدیریت انرژی و تهویه، امنیت، دستیار شخصی، و کنترل‌گر رباتیک) توزیع می‌کند؛ (۲) **حافظه سه‌سطحی خانه** شامل حافظه کاری (Working Context)، حافظه اپیزودیک رویدادها (چه کسی چه زمانی وارد شد یا چه دستوری داد) و گراف معنایی-فضایی خانه (کدام دستگاه یا شیء در کدام اتاق قرار دارد)؛ و (۳) **کنترل دسترسی مبتنی بر نقش و زمینه (RBAC/ABAC) و تأیید انسانی (Human-in-the-Loop)** تا دستورات صادرشده توسط کودک، مهمان یا حتی خودِ مدل هوش مصنوعی نتوانند بدون مجوز صریح، عملیات‌های حساس امنیتی یا فیزیکی را اجرا کنند.

۲.۶. وضعیت واقعی رباتیک خانگی در سال ۲۰۲۶: هزینه، ایمنی و قابلیت اطمینان

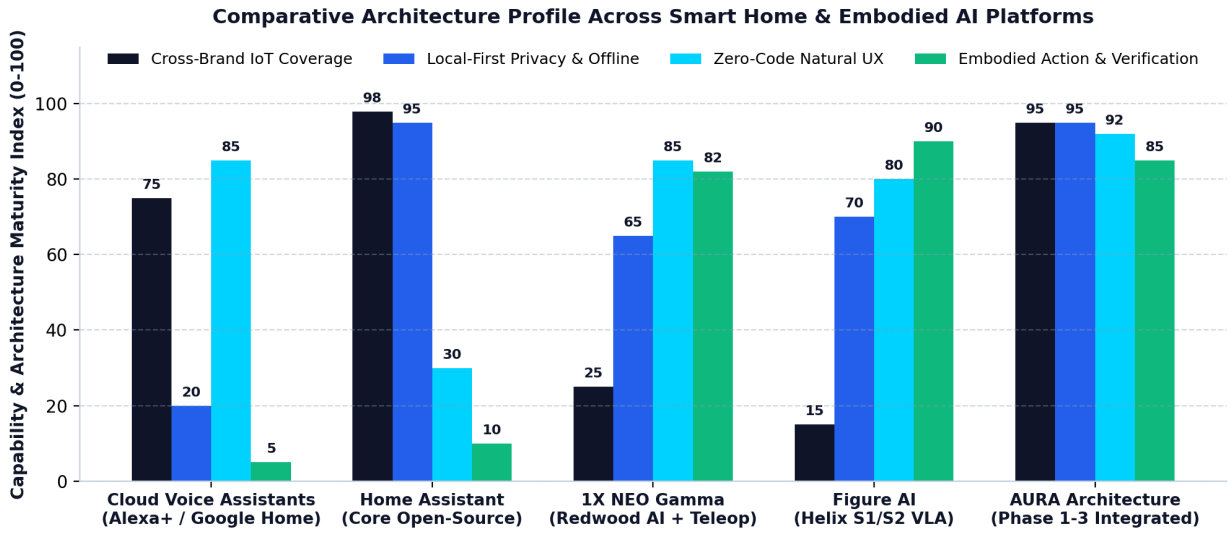
بررسی واقع‌بینانه بازار رباتیک خانگی نشان می‌دهد که میان «دموهای تبلیغاتی آزمایشگاهی» و «محصول قابل اعتماد در خانه» شکاف مهندسی بزرگی وجود دارد که AURA با استراتژی مرحله‌ای خود به دنبال مدیریت آن است:

- **وضعیت فعلی بازار:** امروز تنها ربات‌های خانگی که به تولید انبوه میلیونی، قیمت اقتصادی (۳۰۰ تا ۱۵۰۰ دلار) و قابلیت اطمینان بالا رسیده‌اند، ربات‌های تک‌منظوره چرخ‌دار (مانند جارو و تی‌های رباتیک مجهز به SLAM و تشخیص مانع بصری) هستند. ربات‌های انسان‌نمای خانگی نظیر 1X NEO Gamma در سال ۲۰۲۵ پیش‌فروش خود را آغاز کرده و تحویل‌های اولیه محدود خود را در سال ۲۰۲۶ کلید زده‌اند، اما هنوز در مرحله پذیرندگان اولیه (Early Adopters) و با تکیه بر پشتیبانی اپراتور از راه دور قرار دارند.
- **چالش ایمنی فیزیکی و استاندارد ISO 13482:** یک ربات انسان‌نمای صنعتی با وزن ۶۰ تا ۸۰ کیلوگرم و گیربکس‌های صلب (Rigid Harmonic Drives)، در صورت سقوط یا خطای کنترلی در خانه، خطری جدی برای کودکان، سالمندان و حیوانات خانگی محسوب می‌شود. طبق استاندارد بین‌المللی ISO 13482:2014 و مطالعه مروری Safety Engineering for Humanoid Robots in Everyday Life (دسامبر ۲۰۲۵)، ورود ربات به فضای زندگی نیازمند طراحی ذاتاً ایمن (وزن زیر ۳۵ کیلوگرم، مفاصل تاندونی یا برگشت‌پذیر با قابلیت Backdrivability بالا، پوشش پلیمری نرم ضد نیشگون)، پایش لحظه‌ای سرعت و فاصله، و کلید توقف اضطراری مستقل است.

- **محدودیت‌های هزینه، انرژی و نگهداری:** هزینه تمام‌شده (BOM) یک ربات انسان‌نمای مجهز به دست‌های ماهر چندانگشتی هنوز بین ۱۵,۰۰۰ تا ۵۰,۰۰۰ دلار است؛ عمر باتری در کار فیزیکی پیوسته معمولاً بین ۱.۵ تا ۴ ساعت محدود است؛ و تعمیر و نگهداری مفاصل مکانیکی در خانه برای مصرف‌کننده عادی دشوار است. این واقعیت‌ها ثابت می‌کند که چرا شروع AURA از «هسته نرم‌افزاری و کنترل زیرساخت موجود» هوشمندانه‌ترین استراتژی ورود به بازار است.

۳. بررسی عمیق رقبا و نمونه‌های صنعتی پیشرو

در این بخش، شش نمونه و رقیب اصلی صنعتی در طیف رباتیک انسان‌نما، مدل‌های پایه تجسم‌یافته و زیرساخت‌های خانه هوشمند به صورت عینی و بر اساس پنج معیار یکسان بررسی شده‌اند. هدف، درک دقیق مرز توانایی‌های واقعی آن‌ها و استخراج درس‌های عملی برای معماری AURA است. در تحلیل بازار امروز، پرهیز از ادعاهای تبلیغاتی و تفکیک دقیق میان «دموهای کنترل‌شده آزمایشگاهی» و «محصولات پایدار در محیط خانه» اهمیت حیاتی دارد؛ چرا که هیچ‌یک از بازیگران فعلی به تنهایی تمام لایه‌های زنجیره ارزش (از پروتکل‌های خانه هوشمند و حریم خصوصی محلی تا برنامه‌ریزی شناختی و دستکاری فیزیکی اشیا) را به طور کامل پوشش نداده‌اند.



شکل ۲: مقایسه چندبعدی نمایه توانمندی پلتفرم‌های موجود (دستیارهای ابری، 1X NEO، Home Assistant و Figure Helix) در برابر پروفایل هدف معماری یکپارچه AURA.

نمونه صنعتی / رقیب	چه کاری انجام می‌دهد و چه فناوری‌ای دارد؟	چه چیزی را هنوز نمی‌تواند انجام دهد؟ (محدودیت‌ها)	تفاوت با چشم‌انداز AURA چیست؟	چه درسی از آن برای AURA می‌گیریم؟
1X Technologies (NEO Gamma & Redwood AI)	ربات انسان‌نمای خانگی ۳۰ کیلوگرمی با ارتفاع ۱۶۷ سانتی‌متر، محرک‌های تاندونی بی‌صدا (۲۲ دسی‌بل)، بدنه پلیمری نرم، پردازنده NVIDIA Jetson Thor و مدل Redwood AI برای راه‌رفتن و انجام کارهای سبک خانه.	هنوز در تمام وظایف پیچیده خانگی ۱۰۰٪ خودمختار نیست و در صورت مواجهه با شرایط جدید به هدایت اپراتور از راه دور (Expert Mode / Teleoperation) نیاز دارد؛ همچنین یک پلتفرم جامع مدیریت IoT خانه نیست.	1X یک شرکت سخت‌افزار-محور است که ربات فیزیکی کامل می‌فروشد، در حالی که AURA از لایه نرم‌افزاری و ارکستراسیون کل خانه شروع می‌کند و به یک سخت‌افزار واحد محدود نیست.	۱) رابط طبیعی به جای پیچیدگی کنترل ربات؛ ۲) ترکیب خودمختاری با کمک انسان در زمان شکست (Fallback)؛ ۳) اضافه کردن تدریجی قابلیت‌ها و اولویت مطلق ایمنی فیزیکی.
Figure AI (Helix VLA, Helix 02 & Figure 03)	ربات انسان‌نمای عمومی مجهز به معماری VLA دوگانه Helix: سیستم ۲ (مدل بینایی-زبانی VB با نرخ ۷-۹ هرتز) و سیستم ۱	تمرکز اولیه تجاری آن روی لجستیک، انبارداری و خطوط تولید (مانند BMW) است؛ سخت‌افزار آن سنگین‌تر و	Figure به دنبال ساخت کارگر انسان‌نمای همه‌کاره با سخت‌افزار اختصاصی است؛ AURA به دنبال ساخت «مغز و	۱) پیاده‌سازی عملی معماری System 1 و System 2 با یک مجموعه وزن واحد برای چند مهارت؛ ۲) استفاده از

نمونه صنعتی / رقیب	چه کاری انجام می‌دهد و چه فناوری‌ای دارد؟	چه چیزی را هنوز نمی‌تواند انجام دهد؟ (محدودیت‌ها)	تفاوت با چشم‌انداز AURA چیست؟	چه درسی از آن برای AURA می‌گیریم؟
	(کنترل‌گر حرکتی ۸۰M با نرخ ۲۰۰ هرتز) با قابلیت همکاری دو ربات و برداشتن اشیای نادیده.	صنعتی‌تر از یک هاب خانگی ارزان‌قیمت است و زیرساخت یکپارچه‌سازی پروتکل‌های خانه هوشمند را ندارد.	سیستم‌عامل محیط زندگی» است که در فاز اول بدون نیاز به ربات انسان‌نما ارزش ایجاد می‌کند.	داده‌های انسانی برای تعمیم به محیط جدید؛ ۳) تمرکز بر موفقیت کامل تسک نه حرکات نمایشی.
Google DeepMind (Gemini Robotics 2, ER 2, RT-2)	توسعه مدل‌های پایه بینایی-زبان-عمل (VLA) و مدل استدلال تجسم‌یافته (ER 2) که صحنه را درک می‌کند، برنامه چندمرحله‌ای می‌ریزد، پیشرفت کار را می‌سنجد و با مدل کنترلی سطح پایین هماهنگ می‌شود.	DeepMind یک آزمایشگاه پژوهشی و ارائه‌دهنده مدل پایه (Foundation Model Provider) است، نه سازنده یک محصول نهایی یکپارچه و محلی برای مصرف‌کننده نهایی در خانه هوشمند.	DeepMind موتورهای هوش پایه را در سطح کلود و شریک‌های رباتیک فراهم می‌کند؛ AURA معماری محصولی، لایه ایمنی محلی، حافظه خانواده و یکپارچه‌سازی عملیاتی در خانه را می‌سازد.	۱) مدل زبانی به تنهایی موتور کنترل نیست؛ ۲) برنامه‌ریزی، درک محیط، کنترل حرکتی و تشخیص موفقیت باید در یک حلقه بسته کار کنند؛ ۳) استفاده از ابزارهای تخصصی در کنار مدل اصلی.
Physical Intelligence (مدل π0 و π0.7)	توسعه سیاست‌های رباتیک عمومی (Generalist Robot Policies) مبتنی بر Flow Matching با نرخ ۵۰ هرتز برای مهارت‌های ظریف دستکاری اشیا (تا کردن لباس، جمع کردن میز) روی سخت‌افزارهای متنوع.	فاقد لایه مدیریت تجهیزات خانه هوشمند، سیستم کنترل دسترسی اعضای خانواده، رابط کاربری مصرف‌کننده و زیرساخت امنیتی شبکه خانگی است.	مدل‌های π0 صرفاً لایه سیاست حرکتی مفاصل (لایه 1 / System 4B) را پوشش می‌دهند، در حالی که AURA کل پشته ۵ لایه از رابط خانواده تا مدیریت IoT و ایمنی را در بر می‌گیرد.	امکان استفاده از سیاست‌های حرکتی مستقل از سخت‌افزار (Cross-embodiment policies) در فاز ۲ و ۳ برای کنترل بازوها و ربات‌های خانگی مختلف.
Home Assistant (Open Home Foundation)	بزرگ‌ترین پلتفرم متن‌باز اتوماسیون خانگی با پشتیبانی از هزاران برند، پروتکل‌های Matter، Thread، Zigbee و Z-Wave، کنترل ۱۰۰٪ محلی و دستیار صوتی پایه Assist.	برای کاربر عادی پیچیده است و تعریف سناریوهای پیشرفته نیازمند درگیری با تنظیمات فنی، داشبوردهای مهندسی یا YAML است؛ همچنین فاقد مدل جهان فیزیکی، برنامه‌ریز خودمختار چندمرحله‌ای و کنترل رباتیک پیشرفته است.	Home Assistant یک «لایه اتوماسیون شرطی و درایور دستگاه‌ها» است؛ AURA «لایه شناختی، برنامه‌ریزی خودمختار و رابط طبیعی بدون کدنویسی» است که روی زیرساختی مانند Home Assistant سوار می‌شود.	۱) سازگاری بین برندها مهم‌تر از هر سخت‌افزار منفرد است؛ ۲) کنترل محلی ضامن حریم خصوصی و کارکرد آفلاین است؛ ۳) حذف کامل کدنویسی و تنظیمات پیچیده برای کاربر عادی.
دستیارهای تجاری خانه هوشمند (Google Home Gemini, Alexa+, Apple HomeKit)	پاسخگویی صوتی طبیعی، پخش رسانه، مدیریت تایمرها و اجرای دستورات کنترلی روی دستگاه‌های سازگار با اکوسیستم ابری هر شرکت.	وابستگی شدید به اینترنت و پردازش ابری (ریسک حریم خصوصی)، قفل‌شدگی در اکوسیستم تبلیغاتی/تجاری شرکت سازنده، نبود موتور ایمنی قطعی برای رباتیک فیزیکی و فقدان حلقه تأیید بصری انجام کار.	دستیارهای تجاری عمدتاً «سخن‌گوهای ابری متصل به کلیدهای برق» هستند؛ AURA یک «عامل مقیم، محلی، حافظه‌دار و ایمن» است که برای تکامل به سمت عمل فیزیکی و رباتیک طراحی شده است.	سادگی تجربه کاربری اولیه (نصب سریع و مکالمه صوتی روان) که باید در لایه ۱ معماری AURA حفظ شود، اما با پشتوانه پردازش محلی و ایمنی سخت‌گیرانه.

۴. نقد، اصلاح و طراحی معماری مرجع ۵ لایه AURA

معماری اولیه پنج‌بخشی مطرح‌شده در یادداشت‌های شما (۱. رابط کاربری، ۲. هسته هوش/مغز ایجنت، ۳. موتور سیاست و ایمنی، ۴. لایه اجرا در دو شاخه دستگاه‌ها و عامل‌های فیزیکی، و ۵. بررسی نتیجه و یادگیری) از نظر مهندسی سیستم یک ساختار بسیار منطقی و مدرن است؛ به ویژه قرار دادن **موتور سیاست و ایمنی (لایه ۳)** به عنوان یک گلوگاه کنترلی در میان مغز تصمیم‌گیر و لایه اجرا، و همچنین بستن حلقه با **لایه ۵ (بررسی نتیجه و یادگیری)**، دقیقاً نقاط ضعفی را پوشش می‌دهد که اکثر پروژه‌های مبتنی بر LLM از آن رنج می‌برند.

۴.۱. نقد معماری اولیه و ۵ اصلاح کلیدی مهندسی

با این حال، زمانی که این معماری را از دیدگاه یک تیم چندتخصصی (معمار فنی، پژوهشگر هوش مصنوعی و مهندس رباتیک) برای پیاده‌سازی واقعی زیر ذره‌بین قرار می‌دهیم، پنج چالش پنهان در طرح اولیه دیده می‌شود که در **معماری اصلاح‌شده AURA** برطرف شده‌اند:

1. **تفکیک درونی لایه ۲ به سه ماژول با مقیاس زمانی متفاوت:** در طرح اولیه، «درک، برنامه‌ریزی، حافظه و تصمیم‌گیری» در یک جعبه واحد به نام مغز ایجنت قرار داشتند. اگر تمام این وظایف به یک مدل بزرگ سپرده شود، تأخیر سیستم برای دستورات ساده و کنترل فیزیکی بیش از حد بالا خواهد بود. در معماری اصلاح‌شده، لایه ۲ به سه بخش صریح تقسیم شده است: **System 2 (استدلال‌گر تحلیلی و برنامه‌ریز HTN/ReAct با نرخ ۱ تا ۵ هرتز)**، **System 1 (مفسر واکنشی سریع و کنترل‌گر پیوسته حرکتی با نرخ ۵۰ تا ۲۰۰ هرتز)**، و **مدل جهان نهفته (Latent World Model) مبتنی بر معماری V-JEPA 2)** برای شبیه‌سازی پیامد اعمال پیش از اجرا.
2. **تبدیل لایه ۳ از یک فیلتر ساده به «دروازه قطعی سیاست، حریم خصوصی و ایمنی» (Deterministic Guardrail):** یک مدل عصبی نباید خودش مسئول بررسی ایمنی خروجی خودش باشد! لایه ۳ در معماری اصلاح‌شده به یک موتور قطعی و مبتنی بر قواعد رسمی (Formal Rule Engine) تبدیل شده که سه وظیفه دارد: (الف) کنترل دسترسی مبتنی بر نقش و زمینه (RBAC/ABAC) برای تفکیک کودک، والدین و مهمان)؛ (ب) دروازه حریم خصوصی محلی (Local Privacy Filter) که مانع خروج ویدیو و صدای خام از خانه می‌شود؛ و (ج) مسیر قطع اضطراری سخت‌افزاری و نرم‌افزاری (E-Stop) مطابق استاندارد ISO 13482 که حتی بدون دخالت لایه ۲ می‌تواند حرکت عامل‌های فیزیکی را در کسری از ثانیه متوقف کند.
3. **اضافه شدن «گراف وضعیت فضایی و معنایی خانه» (Unified Spatial Scene Graph):** برای اینکه شاخه ۴-الف (دستگاه‌های IoT) و شاخه ۴-ب (ربات‌های فیزیکی) بتوانند با هم کار کنند، به یک زبان فضایی مشترک نیاز دارند. گراف وضعیت خانه مکان دقیق اتاق‌ها، وضعیت لحظه‌ای هر سنسور/قفل، و موقعیت اشیا و اعضای خانواده را نگهداری می‌کند.
4. **تعریف صریح مکانیزم Fallback انسانی در لایه ۴-ب و لایه ۵:** با الهام از دست‌نوشته شما درباره 1X NEO («ترکیب خودمختاری با کمک انسان هنگام شکست»)، در معماری اصلاح‌شده مسیر مشخصی تعبیه شده است تا هرگاه عامل فیزیکی در لایه ۴-ب با خطای غیرقابل حل مواجه شد و لایه ۵ شکست را تشخیص داد، سیستم به صورت ایمن متوقف شده و از طریق لایه ۱ از کاربر کمک بگیرد یا به حالت هدایت کمکی (Guided Teleoperation) سوئیچ کند.
5. **شفاف‌سازی مرز زمانی فناوری‌ها (موجود، پیشنهادی در MVP، و آینده/پژوهشی):** برای اینکه طرح در برابر داوران فنی و سرمایه‌گذاران کاملاً واقع‌بینانه و قابل دفاع باشد، تمامی اجزای معماری با سه برچسب زمانی مشخص تفکیک شده‌اند.

در جریان اجرای یک درخواست در این معماری اصلاح‌شده، داده‌ها یک مسیر رفت‌وبرگشتی (Closed-Loop Pipeline) دقیق را طی می‌کنند: ابتدا سیگنال‌های صوتی، متنی یا تصویری در لایه ۱ دریافت و بر اساس هویت عضو خانواده برچسب‌گذاری می‌شوند. اگر فرمان یک دستور صریح و کم‌ریسک محلی باشد (مانند تغییر شدت نور اتاق)، مسیر سریع System 1 در لایه ۲ آن را در کمتر از چند ده میلی‌ثانیه تفسیر کرده و پس از تأیید آنی در دروازه سیاست لایه ۳، به شاخه ۴-الف ارسال می‌کند. اما اگر فرمان نیازمند برنامه‌ریزی چندمرحله‌ای یا اقدام فیزیکی رباتیک باشد، استدلال‌گر System 2 با کمک گراف حافظه فضایی و شبیه‌سازی پیش‌بینانه در مدل جهان نهفته (Latent World Model)، زنجیره‌ای از زیرکارها را تولید می‌کند؛ هر زیرکار پیش از اجرا در لایه ۴-ب از فیلتر ایمنی و کنترل دسترسی لایه ۳ عبور کرده و پس از اجرا، لایه ۵ از طریق حسگرها و دوربین‌ها تحقق واقعی هدف را راستی‌آزمایی می‌نماید.

شکل ۳: دیاگرام معماری مرجع ۵ لایه ارتقایافته AURA با تفکیک رنگ‌بندی‌شده اجزای موجود (سبز)، اجزای پیشنهادی در هسته محصول (آبی الکتریکی) و فناوری‌های تحقیقاتی فازهای آینده (نارنجی/طلایی).

۴.۲. جدول تفکیک اجزای موجود، پیشنهادی و آینده در معماری AURA

در جدول زیر، تمامی اجزای پنج لایه معماری AURA بر اساس سطح آمادگی فناوری (TRL) به سه دسته «موجود و آماده در فاز اول»، «نوآوری مهندسی پیشنهادی در هسته محصول» و «فناوری‌های پژوهشی فازهای آینده» تفکیک شده‌اند تا مرز میان محصول قابل تحویل امروز و نقشه راه تحقیق و توسعه کاملاً شفاف باشد:

لایه معماری	اجزای موجود و آماده (فاز ۱ — استفاده از استانداردها)	اجزای پیشنهادی در هسته AURA (نوآوری مهندسی ما در MVP و فاز ۲)	فناوری‌های آینده و پژوهشی (فاز ۳ — نیازمند بلوغ R&D)
لایه ۱: رابط کاربری	مدل‌های تشخیص و تولید گفتار محلی (ASR/TTS)، اپلیکیشن موبایل/تبلت، و دریافت جریان تصویر از دوربین‌های خانگی استاندارد.	رابط یکپارچه چندوجهی بدون کدنویسی، تشخیص هویت و نقش اعضای خانواده (کودک/والدین/سالمند) و داشبورد تأیید عملیات حساس.	درک پیشرفته اشارات ظریف بدن (Micro-gestures) و تعامل هولوگرافیک/مکانی سه‌بعدی در سراسر خانه.
لایه ۲: هسته شناختی (مغز ایجنت)	مدل‌های زبانی و بینایی-زبانی متن‌باز (SLM/VLM قابل اجرا روی Edge) و چارچوب پایه فراخوانی ابزار (ReAct).	معماری دوگانه System 1 (مفسر سریع محلی) و System 2 (برنامه‌ریز تحلیلی)، روتر ابزارهای تخصصی، و گراف حافظه فضایی و ترجیحات خانه.	مدل جهان نهفته مشروط بر عمل (مبتنی بر V-JEPA 2) برای شبیه‌سازی فیزیکی بلندمدت و برنامه‌ریزی حرکتی Zero-Shot روی اشیای پیچیده.
لایه ۳: موتور سیاست و ایمنی	رمزنگاری شبکه محلی، فایروال پایه و کلیدهای قطع اضطراری سخت‌افزاری استاندارد.	موتور قطعی سیاست‌گذاری (RBAC/ABAC)، ماتریس ریسک ۳ سطحی با الزام تأیید انسانی (HITL)، و فیلتر حریم خصوصی Local-First.	پایشگر ایمنی بیومکانیکی بلادرنگ منطبق بر ISO 13482 برای تنظیم پویای گشتاور و میدان ایمنی ربات انسان‌نما در حضور کودک.
لایه ۴: لایه اجرا (4A و 4B)	**4A:** هسته Home Assistant، پروتکل‌های Matter، Thread و Zigbee (نور، تهویه، قفل، سنسور). **4B:** جارو و تی رباتیک هوشمند.	**4A:** هماهنگ‌ساز خودکار سناریوهای چندبرندی بدون نوشتن اسکریپت. **4B:** اتصال به ربات‌های متحرک	**4B:** یکپارچه‌سازی کامل با ربات‌های انسان‌نمای خانگی با بدنه نرم (سبک 1X NEO) به همراه زیرساخت Teleoperation ایمن در زمان خطا.

لایه معماری	اجزای موجود و آماده (فاز ۱ — استفاده از استانداردها)	اجزای پیشنهادی در هسته AURA (نوآوری مهندسی ما در MVP و فاز ۲)	فناوری‌های آینده و پژوهشی (فاز ۳ — نیازمند بلوغ R&D)
		چرخ‌دار خانگی و بازوهای کمکی ثابت (فاز ۲).	
لایه ۵: بررسی نتیجه و یادگیری	لاگ‌های وضعیت دستگاه‌ها (Device State Logs) و کدهای خطای استاندارد پروتکل‌های شبکه.	حلقه بسته تأیید چندوجهی (بررسی حسگری و بصری تغییر واقعی وضعیت)، مکانیزم خوداصلاحی و ثبت غیرمزاحم ترجیحات روزمره خانواده.	یادگیری مستمر محلی (On-device Continual Learning) برای به‌روزرسانی وزن‌های مدل جهان و مهارت‌های حرکتی ربات در خانه.

۵. تعریف جامع محصول و استراتژی اجرایی (Product Strategy)

بر اساس تحلیل فناوری و معماری مرجع، استراتژی محصول AURA در ۱۳ محور استاندارد مدیریتی و محصولی به شرح زیر تدوین شده است تا مسیر تبدیل ایده به یک محصول واقعی، کم‌ریسک و مقیاس‌پذیر کاملاً روشن باشد.

۵.۱. چشم‌انداز، مأموریت، مسئله و راهکار پیشنهادی

- **چشم‌انداز محصول (Product Vision):** تبدیل هر خانه به یک محیط هوشمند، امن و خودکار که نیازهای روزمره ساکنانش را می‌فهمد و با هماهنگی یکپارچه میان تجهیزات هوشمند و عامل‌های رباتیک، زمان و آرامش را به خانواده بازمی‌گرداند؛ نقطه‌ای که از خانه آغاز شده و به دفاتر و محیط‌های کاری گسترش می‌یابد.
- **بیانیه مأموریت (Mission):** ساخت یک هسته شناختی و کنترلی محلی، ایمن و مستقل از برند برای زندگی روزمره که پیچیدگی‌های فنی خانه هوشمند و رباتیک را پشت یک رابط طبیعی انسانی پنهان می‌کند.
- **بیانیه مسئله (Problem Statement):** خانواده‌های امروزی با ده‌ها دستگاه هوشمند جزیره‌ای روبه‌رو هستند که یا نیازمند تنظیمات فنی و کدنویسی پیچیده‌اند (مانند Home Assistant خام)، یا به سرورهای ابری وابسته بوده و حریم خصوصی را نقض می‌کنند (مانند دستیارهای صوتی تجاری)، و مهم‌تر از همه، فاقد یک مغز واحد، ایمن و دارای درک محیطی هستند که بتواند وظایف چندمرحله‌ای را برنامه‌ریزی کرده و در آینده ربات‌های خانگی را هدایت نماید.
- **راهکار پیشنهادی (Proposed Solution):** عرضه **AURA** به عنوان یک عامل هوشمند مسکونی که در نسخه اول (MVP) در قالب یک هاب نرم‌افزاری/لبه‌ای محلی (Local Edge AI Hub) روی زیرساخت استاندارد خانه هوشمند سوار می‌شود، با زبان طبیعی و بدون نیاز به کدنویسی با تمام اعضای خانواده ارتباط برقرار می‌کند، از طریق موتور سیاست و ایمنی از عملیات‌های حساس محافظت می‌نماید، و به صورت ماژولار آماده اتصال به ربات‌های متحرک و انسان‌نما در فازهای بعدی است.

۵.۲. کاربران هدف، ارزش پیشنهادی و وجوه تمایز

استراتژی ورود به بازار AURA بر پایه بخش‌بندی دقیق کاربران هدف (از دارندگان تجهیزات خانه هوشمند در فاز اول تا بازار سازمانی و سازندگان رباتیک در فازهای بعدی) و تعریف یک گزاره ارزش ملموس بنا شده است که در جدول زیر تشریح می‌گردد:

محور استراتژیک	تشریح دقیق محصولی و بازار هدف
کاربران هدف (Target Users)	**۱. پذیرندگان اولیه (فاز ۱ — Prosumers & Smart Home Owners):** خانواده‌ها و علاقه‌مندان فناوری که تجهیزات هوشمند (نور، ترموستات، قفل، جاروی رباتیک) دارند اما از پیچیدگی اپلیکیشن‌های متعدد یا تنظیمات Home Assistant خسته شده‌اند. **۲. بازار میان‌مدت (فاز ۲ — خانواده‌های پرمشغله و مراقبت از سالمندان/کودکان):** خانواده‌هایی که به دنبال پایش ایمن خانه، مدیریت هوشمند انرژی و کمک در کارهای روزمره با حفظ کامل حریم خصوصی در داخل خانه هستند. **۳. بازار سازمانی و توسعه‌ای (B2B):** سازندگان مسکن مدرن، دفاتر کار هوشمند، کلینیک‌ها و شرکت‌های سازنده سخت‌افزار رباتیک که به یک سیستم‌عامل شناختی و ایمن برای محیط داخلی نیاز دارند.

محور استراتژیک	تشریح دقیق محصولی و بازار هدف
ارزش پیشنهادی اصلی (Core Value Proposition)	«یک مغز واحد، محلی و ایمن برای تمام خانه شما: با آن مثل یک عضو قابل اعتماد خانواده صحبت کنید تا بدون نیاز به کدنویسی یا ارسال تصاویر خانه‌تان به اینترنت، همه دستگاه‌ها و ربات‌های خانگی را با رعایت کامل ایمنی هماهنگ کند.»
وجوه تمایز واقع‌بینانه (Differentiation)	برخلاف دستیارهای ابری (Alexa+/Google Home)، پردازش حساس AURA **محلی و آفلاین (Local-First)** است. برخلاف Home Assistant خام، تجربه کاربری آن **کاملاً عامل‌محور و بدون کدنویسی (Zero-Config Agentic UX)** است. و برخلاف شرکت‌های تک‌سخت‌افزاری رباتیک، AURA دارای **معماری باز فرادستگاهی، موتور ایمنی قطعی (لایه ۳) و حلقه تأیید اجرای کار (لایه ۵)** است.

۵.۳. سناریوهای کاربردی ملموس (Use Cases)

- **سناریوی ۱ — مدیریت چندمرحله‌ای ورود به خانه و پذیرایی (فاز ۱):** کاربر با دست پر وارد می‌شود و می‌گوید: «AURA، تا یک ساعت دیگر مهمان داریم، هوای پذیرایی گرم است و کف آشپزخانه هم کثیف شده.» سیستم با مراجعه به حافظه ترجیحات، دمای تهویه را روی ۲۱ درجه تنظیم می‌کند، نور مخفی پذیرایی را روشن می‌سازد، جاروی رباتیک را صرفاً به ناحیه آشپزخانه اعزام می‌کند و پس از پایان تمیزکاری، با بررسی وضعیت حسگرها، آماده بودن فضا را به کاربر اطلاع می‌دهد.
- **سناریوی ۲ — نگهبانی ایمنی و کنترل دسترسی خانواده (فاز ۱ و ۲):** کودک خانواده در غیاب والدین درخواست می‌کند: «درِ ورودی را باز کن» یا «فر آشپزخانه را روشن کن». موتور سیاست و ایمنی (لایه ۳) بر اساس پروفایل صوتی و تصویری کودک، درخواست را به عنوان عملیات حساس غیرمجاز شناسایی کرده، از اجرای آن جلوگیری می‌کند و همزمان یک اعلان فوری برای تأیید یا رد به گوشی والدین ارسال می‌نماید.
- **سناریوی ۳ — همکاری اینترنت اشیا و عامل فیزیکی/رباتیک در شب (فاز ۲ و ۳):** در هنگام شب، سیستم قصد دارد برنامه مرتب‌سازی و نظافت شبانه را اجرا کند. ابتدا از طریق حسگرهای حضور مطمئن می‌شود که همه اعضای خانواده در اتاق خواب هستند، نور راهرو را فقط به اندازه دید دوربین ربات (۱۰٪) روشن می‌کند، بازوی کمکی یا ربات خانگی لیوان‌های جا مانده روی میز را جمع کرده و در سینک یا ماشین ظرفشویی قرار می‌دهد، و در صورتی که با یک شیء ناشناخته یا شکست مواجه شود، بدون ایجاد سر و صدا متوقف شده و گزارش تصویری آن را برای صبح ثبت می‌کند.

۵.۴. نقشه راه تکاملی ۳ فازی (Product Roadmap)

نقشه راه AURA بر این اصل طلایی بنا شده است که **نسخه اول (MVP) یک ربات انسان‌نما نیست** و موفقیت تجاری فاز اول هیچ وابستگی شکننده‌ای به بلوغ سخت‌افزارهای گران‌قیمت انسان‌نما ندارد:

فاز و افق زمانی (بر اساس فرضیات)	محصول محوری و خروجی قابل تحویل	قابلیت‌های فنی و لایه‌های فعال	معیار ارزیابی موفقیت (KPIs)
فاز ۱: هسته نرم‌افزاری و هاب محلی (ماه ۱ تا ۹ — تیم ۵ تا ۷ نفره)	**AURA Core & Edge Hub (MVP):** نرم‌افزار و هاب پردازش لبه‌ای متصل به تجهیزات خانه هوشمند و جاروی رباتیک.	لایه ۱ (صوت/متن/تصویر پایه)، لایه ۲ (برنامه‌ریز System 2 + مفسر سریع System 1 + حافظه ترجیحات)، لایه ۳ (موتور سیاست، RBAC و تأیید حساس)، لایه 4A کامل و لایه ۵ پایه.	نصب بدون کدنویسی در کمتر از ۱۵ دقیقه؛ تأخیر دستورات محلی زیر ۲۰۰ میلی‌ثانیه؛ کارکرد ۱۰۰٪ آفلاین در سناریوهای ایمنی؛ نرخ صفر درصد اجرای غیرمجاز عملیات حساس.
فاز ۲: ادراک فضایی و ربات‌های کمکی (ماه ۹ تا ۱۸ — تیم ۸ تا ۱۲ نفره)	**AURA Spatial & Assisted Robotics:** افزودن درک فضایی سه‌بعدی و کنترل ربات‌های متحرک چرخ‌دار و بازوی کمکی آشپزخانه/میز.	گراف معنایی-فضایی پویا از اشیای خانه، حلقه تأیید بصری اجرای کار در لایه ۵، سیاست‌های VLA (سبک π0) در لایه ۲، و مکانیزم درخواست کمک از انسان در صورت شکست.	نرخ موفقیت بالای ۸۵٪ در وظایف جابه‌جایی و نظافت ترکیبی (IoT + ربات متحرک) و تشخیص صحیح شکست در بیش از ۹۵٪ موارد.
فاز ۳: یکپارچه‌سازی انسان‌نما و مدل	**AURA Embodied OS:** سیستم‌عامل شناختی و ایمنی برای	ادغام مدل جهان نهفته مشروط بر عمل (مبتنی بر V-JEPA 2)، کنترل دوگانه ۲۰۰ هرتزی (سبک Helix)،	انجام ایمن وظایف چندمرحله‌ای فیزیکی در محیط‌های خانگی

فاز و افق زمانی (بر اساس فرضیات)	محصول محوری و خروجی قابل تحویل	قابلیت‌های فنی و لایه‌های فعال	معیار ارزیابی موفقیت (KPIs)
جهان (ماه ۱۸ تا ۳۶ — شراکت OEM)	ربات‌های انسان‌نمای خانگی و محیط‌های اداری/شرکتی.	ایمنی پویای ISO 13482 و زیرساخت Teleoperation ایمن.	جدید و گسترش پلتفرم به دفاتر اداری و مراکز مراقبتی.

۵.۵. ریسک‌ها، محدودیت‌ها، حریم خصوصی و چشم‌انداز بلندمدت

- **ریسک‌ها و محدودیت‌ها (Risks & Limitations) و راهکار مهار آن‌ها:**
  - *ریسک ۱ — خطای استدلال یا توهم مدل هوش مصنوعی:* مهار از طریق جداسازی کامل لایه ۲ از لایه ۴ و اجباری کردن عبور تمامی فرامین از موتور قطعی سیاست و ایمنی (لایه ۳) به همراه تأیید انسانی برای عملیات‌های حساس.
  - *ریسک ۲ — محدودیت توان پردازشی هاب محلی:* مهار از طریق معماری دوگانه System 1/2 (استفاده از مدل‌های کوچک کوانتایزشده برای ۹۰٪ کارهای روزمره و فراخوانی مدل‌های سنگین‌تر فقط برای برنامه‌ریزی‌های پیچیده).
  - *ریسک ۳ — کندی تجاری‌سازی یا قیمت بالای ربات‌های انسان‌نما در بازار جهانی:* از آنجا که AURA در فاز ۱ و ۲ روی تجهیزات خانه هوشمند، جاروهای رباتیک و ربات‌های چرخ‌دار/بازوهای موجود ارزش کامل ایجاد می‌کند، تأخیر در ارزان شدن ربات‌های انسان‌نما حیات محصول را تهدید نمی‌کند.

- **حریم خصوصی و ایمنی (Privacy & Safety Architecture):** معماری AURA بر سه اصل غیرقابل مذاکره استوار است: (۱) **مالکیت محلی داده (Local Data Sovereignty):** تصویر دوربین‌ها، صدای محیط و گراف عادت‌های خانواده در داخل حافظه رمزنگاری‌شده هاب خانگی نگهداری می‌شوند؛ (۲) **شفافیت و قابلیت ممیزی (Auditability):** هر تصمیم ایجنت در لایه ۵ با ذکر دلیل ثبت می‌شود تا کاربر بداند چرا یک دستگاه روشن یا خاموش شده است؛ و (۳) **ایمنی فیزیکی چندلایه:** رعایت الزامات استاندارد ISO 13482 و اولویت مطلق دستور توقف انسان بر هر فرآیند خودکار.
- **چشم‌انداز بلندمدت (Long-Term Vision):** همان‌طور که در دست‌نوشته اول شما به درستی اشاره شده است، «خانه» نقطه شروع است، نه نقطه پایان. زمانی که هسته شناختی، موتور ایمنی و مدل جهان AURA در پیچیده‌ترین و حساس‌ترین محیط یعنی خانه مسکونی آزموده و بالغ شد، همین معماری ۵ لایه به عنوان **سیستم‌عامل استاندارد فضاهای هوشمند (Universal Spatial & Embodied Agent OS)** به دفاتر اداری، شرکت‌ها، کلینیک‌های درمانی، هتل‌ها و کارگاه‌های صنعتی گسترش خواهد یافت.

۶. فهرست منابع و مراجع علمی و صنعتی (References)

تمامی منابع زیر مستقیماً در تحلیل‌های این گزارش مورد استناد قرار گرفته‌اند (تاریخ دسترسی و راستی‌آزمایی تمامی لینک‌ها: ۳۰ اکتبر ۲۰۲۶):

1. **Kahneman, D. (2011).** *Thinking, Fast and Slow*. Farrar, Straus and Giroux. [انتشار: اکتبر ۲۰۱۱ | دسترسی: ۳۰ اکتبر ۲۰۲۶] — https://us.macmillan.com/books/9780374533557/thinkingfastandslow

2. **Christakopoulou, K., Mourad, S., & Matarić, M. (2024).** *Agents Thinking Fast and Slow: A Talker-Reasoner Architecture*. arXiv:2410.08328. [انتشار: ۱۰ اکتبر ۲۰۲۴ | دسترسی: ۳۰ اکتبر ۲۰۲۶] — https://arxiv.org/abs/2410.08328

3. **Yao, S., et al. (2022/2023).** *ReAct: Synergizing Reasoning and Acting in Language Models*. ICLR 2023, arXiv:2210.03629. [انتشار: اکتبر ۲۰۲۲ / مارس ۲۰۲۳ | دسترسی: ۳۰ اکتبر ۲۰۲۶] — https://arxiv.org/abs/2210.03629

4. **LeCun, Y. (2022).** *A Path Towards Autonomous Machine Intelligence (v0.9.2)*. NYU & Meta FAIR, OpenReview. [انتشار: ۲۷ ژوئن ۲۰۲۲ | دسترسی: ۳۰ اکتبر ۲۰۲۶] — https://openreview.net/pdf?id=BZ5a1r-kVsf

5. **Assran, M., et al. — Meta FAIR (2025).** *V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning*. arXiv:2506.09985. [انتشار: ۱۱ ژوئن ۲۰۲۵ | دسترسی: ۳۰ اکتبر ۲۰۲۶] — https://arxiv.org/abs/2506.09985 | مخزن کد: https://github.com/facebookresearch/vjepa2

6. **Sobal, V., Zhou, G., et al. (2025/2026).** *What Drives Success in Physical Planning with Joint-Embedding Predictive World Models?*. arXiv:2512.24497. [انتشار: دسامبر ۲۰۲۵ / ژانویه ۲۰۲۶ | دسترسی: ۳۰ اکتبر ۲۰۲۶] — https://arxiv.org/html/2512.24497v2

7. **Ha, D., & Schmidhuber, J. (2018).** *World Models*. NeurIPS 2018, arXiv:1803.10122. [انتشار: ۲۷ مارس ۲۰۱۸ | دسترسی: ۳۰ اکتبر ۲۰۲۶] — https://arxiv.org/abs/1803.10122

8. **Embodied AI & World Models Surveys (2025/2026).** *Embodied AI: From LLMs to World Models* (arXiv:2509.20021، سپتامبر ۲۰۲۵) & *World Models for Embodied Intelligence* (arXiv:2609.16697، سپتامبر ۲۰۲۶). [دسترسی: ۳۰ اکتبر ۲۰۲۶]

9. **1X Technologies (2025/2026).** *NEO Home Robot* (https://www.1x.tech/neo)، *Introducing NEO Gamma* (۲۱، https://www.1x.tech/discover/introducing-neo-gamma فوریه ۲۰۲۵) و *Redwood AI Mobility* (۱۰، https://www.1x.tech/discover/redwood-mobility ژوئن ۲۰۲۵). [دسترسی: ۳۰ اکتبر ۲۰۲۶]

10. **Figure AI (2025/2026).** *Helix: A Vision-Language-Action Model for Generalist Humanoid Control* (۲۰، https://www.figure.ai/news/helix فوریه ۲۰۲۵)، *Introducing Figure 03* (۹، https://www.figure.ai/news/introducing-figure-03 اکتبر ۲۰۲۵) و *Introducing Helix 02: Full-Body Autonomy* (۲۷، https://www.figure.ai/news/helix-02 ژانویه ۲۰۲۶). [دسترسی: ۳۰ اکتبر ۲۰۲۶]

11. **Google DeepMind (2023–2026).** *Gemini Robotics 2 & Embodied Reasoning (ER 2)* (https://deepmind.google/models/gemini-robotics/ و بلاگ رسمی ژوئیه ۲۰۲۶)، *RT-2* (ژوئیه ۲۰۲۳) و *AutoRT* / *SARA-RT* (ژانویه ۲۰۲۴). [دسترسی: ۳۰ اکتبر ۲۰۲۶]

12. **Black, K., et al. — Physical Intelligence (2024–2026).** *π0: A Vision-Language-Action Flow Model for General Robot Control* (https://arxiv.org/abs/2410.24164، اکتبر ۲۰۲۴) و *π0.7: A Steerable Model with Emergent Capabilities* (https://www.pi.website/blog/pi07، آوریل ۲۰۲۶). [دسترسی: ۳۰ اکتبر ۲۰۲۶]

13. **Home Assistant & Open Home Foundation (2025/2026).** *Home Assistant Official Platform* (https://www.home-assistant.io/) و *Building the AI-Powered Local Smart Home* (https://www.home-assistant.io/blog/2025/09/11/ai-in-home-assistant/، ۱۱ سپتامبر ۲۰۲۵). [دسترسی: ۳۰ اکتبر ۲۰۲۶]

14. **Smart Home Assistant Landscape (2026).** *Google Home Powered by Gemini* (https://home.google.com/welcome/) و *Google Home vs. Amazon Alexa vs. Apple Home 2026* (https://www.online-tech-tips.com/google-vs-alexa-vs-apple-home-2026/، اکتبر ۲۰۲۶). [دسترسی: ۳۰ اکتبر ۲۰۲۶]

15. **Robot Safety Standards (ISO 13482:2014 & 2025 Review).** *Safety Engineering for Humanoid Robots in Everyday Life — Scoping Review* (https://www.mdpi.com/2079-9292/14/23/4734، دسامبر ۲۰۲۵) و *TÜV Rheinland ISO 13482 Guide* (PDF). [دسترسی: ۳۰ اکتبر ۲۰۲۶]