AI Robotics

Megtestesült AI és a funkcionális biztonság: mikor mondhatjuk, hogy egy robot mellett biztonságban vagyunk?

2026. 09. 07. 14:25 | Méta Open Lab | HUN · 20 perc

Dr. Paniti Imre
kutató, tudományos főmunkatárs · HUN-REN SZTAKI

Dr. Paniti Imre, a HUN-REN SZTAKI kutatója, keretrendszert hoz létre a megtestesült mesterséges intelligencia és robotika funkcionális biztonsága számára.

MIRŐL VOLT SZÓ

A prezentáció az embodied intelligence fejlődését vizsgálja, a Turing-tesztig visszavezetve Rodney Brooks fogalmaira. A funkcionális biztonságot úgy határozzák meg, mint az elérhető eszközök használatát az emberek, rendszerek és gépek egymástól való védelmére elfogadható kockázati küszöbön belül. Például a biztonsági szabványok szükségesek, mert lehetetlen minden lehetséges kockázati forgatókönyvet szimulálni vagy kiküszöbölni összetett környezetekben.

A jelenlegi szabályozási keretek, mint például az EU új mesterséges intelligencia-törvénye és a különböző ISO szabványok, képezik az ipari és együttműködő robotok biztonságának alapját. Ezek a szabványok tartalmaznak konkrét műszaki előírásokat a fájdalomküszöbök és ütközési erők mérésére, hogy biztosítsák az emberi biztonságot az interakció során. Például a együttműködő robotokra vonatkozó műszaki előírás 2025-ben válik a szabvány hivatalos részévé. Ezek az elemek együttesen jelzik az AI-integrált rendszerek szabványosított, ellenőrizhető biztonsági protokolljai felé való átmenetet.

KIEMELT MEGÁLLAPÍTÁSOK

- A funkcionális biztonság a maradék kockázatok kezelésével foglalkozik, amelyeket még kiterjedt szimulációval sem lehet teljes mértékben kiküszöbölni.
- A 2025-ös ipari robot szabványok frissítése tartalmazni fog konkrét követelményeket a kollaboratív robotokra a fizikai sérülések megelőzése érdekében.
- A biztonsági auditoknak ellenőrizniük kell, hogy az AI rendszerek, legyen szó klasszikus architektúrákról vagy nagy nyelvi modellekről, akkor is biztonságosak maradjanak, ha az AI komponens meghibásodik.

ÜZLETI HATÁS

A biztonságkritikus rendszerek piacra dobását tervező vállalatoknak egyensúlyt kell teremteniük a védett „fekete doboz” adatok és a szabályozói átláthatósági követelmények között. A Nvidia által kifejlesztett biztonsági operációs rendszerek szabványosított felületet biztosítanak a fejlesztők számára a kompatibilis, biztonsági előírásoknak megfelelő komponensek létrehozásához. Ezek a rendszerek lehetővé teszik külső érzékelők és aktív motorvezérlés integrálását, például olyan vészleállítási funkciókat, amelyek egy humanoid robotot biztonságosan összeomlanak, ahelyett, hogy elesne.

KÖVETKEZTETÉS

Dr. Paniti Imre azt javasolja, hogy a robotvezérlés legyen teljesen átlátható, ideális esetben a tervezett mozgások peer-to-peer kommunikációja révén, még azok végrehajtása előtt. Ez az átláthatóság elengedhetetlen a biztonság szubjektív érzésének növeléséhez az ember-robot interakciókban, különösen a háztartási robotok esetében. A végső cél annak biztosítása, hogy minden biztonságkritikus mesterséges intelligencia rendszert szigorú tesztelésnek és ellenőrzésnek vessenek alá piacra kerülés előtt, hogy megelőzzék a baleseteket.

AI Robotics

Embodied AI and Functional Safety: When Can We Say We Are Safe Around a Robot?

2026. 09. 07. 14:25 | Méta Open Lab | HUN · 20 perc

Dr. Imre Paniti
Researcher,Senior Research Associate · HUN-REN SZTAKI

Dr. Imre Paniti, a Researcher at HUN-REN SZTAKI, establishes a framework for functional safety in embodied AI and robotics to ensure human protection during machine interaction.

WHAT WAS DISCUSSED

Dr. Imre Paniti explains that functional safety aims to protect humans, systems, and machines from one another within an acceptable risk threshold. Because it is impossible to simulate or account for every possible scenario, safety relies on a cost model where known risks are eliminated while residual risks are managed through standards. For example, the speaker notes that even with reinforcement learning in simulations, some risks remain that must be tolerated based on industry-specific limits.

The presentation details the evolution of safety standards, moving from general directives like the EU's new GP directive to specific technical specifications for industrial and collaborative robots. Dr. Imre Paniti highlights that these standards now include measurable pain thresholds for human-robot interaction and transparency requirements for AI-generated content. These elements collectively signify a transition toward a regulated, auditable workflow where AI systems must pass rigorous safety checks before entering the market.

KEY TAKEAWAYS

- Functional safety is defined by protecting entities from each other up to a predefined level of acceptable risk.
- Technical specifications for collaborative robots are evolving to include specific physical metrics, such as pain thresholds and collision speeds.
- A mandatory audit workflow is necessary to ensure that even if an AI component fails, the overall system remains safe.

BUSINESS IMPACT

The integration of safety standards into the AI and robotics market creates a structured environment for companies to deploy safety-critical systems. By following established technical specifications, such as those involving NVIDIA's safety operating systems, developers can provide compatible interfaces and certified components. This allows businesses to navigate regulatory hurdles, avoid heavy fines, and build consumer trust by ensuring that robots in manufacturing or healthcare do not cause harm.

CONCLUSION

Dr. Imre Paniti concludes by emphasizing the necessity of rigorous testing to prevent accidents involving children or vulnerable populations. The speaker recommends that robot control becomes fully transparent, utilizing peer-to-peer communication to broadcast planned movements before they occur. This transparency is presented as a vital step for increasing the subjective sense of safety in human-robot interaction, particularly for domestic robots.