Agenci AI w technologiach XR

Streszczenie

Streszczenie. Najbardziej dojrzały technicznie obszar *Agentic XR* nie polega obecnie na „umieszczeniu chatbota w VR", lecz na sprzęgnięciu modeli językowych i multimodalnych z trzema warstwami, których same foundation models nadal nie zapewniają wystarczająco niezawodnie: **jawnie reprezentowanym stanem przestrzennym, deterministyczną warstwą wykonawczą oraz mechanizmem walidacji bezpieczeństwa**. Badania z lat 2024–2026 pokazują już działające prototypy agentów modyfikujących sceny, interpretujących zachowanie użytkownika, świadomych kontekstu przestrzennego, sterujących interfejsem XR, wspierających procedury oraz testujących aplikacje. Jednocześnie benchmarki 3D z 2024–2026 konsekwentnie wykazują, że współczesne VLM/MLLM mają problemy z transformacjami perspektywy, trwałą pamięcią przestrzenną, rozumowaniem między odległymi obserwacjami i utrzymywaniem spójnej reprezentacji 3D. [1]

Najsilniejsze dowody empiryczne dotyczą obecnie czterech klas zastosowań: naturalnojęzykowego authoringu i edycji scen, kontekstowych asystentów proceduralnych, adaptacyjnych embodied conversational agents oraz automatycznego testowania XR. LLMR demonstruje runtime'ową kreację i modyfikację świata w Unity; SIGMA i LEGO-MRTA łączą LLM/VLM z obserwacją wykonywanej procedury; Sensible Agent i EmBARDiment wykorzystują kontekst użytkownika, spojrzenie lub historię interakcji; XRintTest osiąga wysoki poziom automatycznego pokrycia interakcji przestrzennych bez konieczności udziału człowieka w każdej ścieżce testowej. [2]

Bardziej ryzykowne pozostają autonomiczne działania w świecie fizycznym. W XR–robotics literatura z lat 2024–2025 daje mocne argumenty za *shared autonomy*, teleoperacją wspomaganą oraz cyfrowym bliźniakiem, ale nie za bezpośrednim przekazywaniem swobodnie wygenerowanych przez LLM poleceń do sterownika robota. IMMERTWIN, system wspomagający teleoperację humanoida i MIRAGE pokazują korzyści z połączenia immersyjnego interfejsu, estymacji intencji i deterministycznych mechanizmów kontroli; architektury te są znacznie bliższe bezpiecznemu wzorcowi Agentic XR niż agent samodzielnie generujący komendy aktuatorów. [3]

Metoda przeglądu. Przegląd został przeprowadzony jako *structured scoping review inspired by PRISMA 2020*, ale nie jako formalny systematic review zgodny z PRISMA. Nie deklaruję liczby rekordów PRISMA ani formalnego diagramu identyfikacja–screening–eligibility, ponieważ interfejs wyszukiwania nie zapewniał reprodukowalnego eksportu rekordów z każdego indeksu Scopus/WoS/IEEE/ACM. Weryfikowałem natomiast publikacje poprzez strony wydawców, DOI, bibliografie instytucjonalne, oficjalną dokumentację i repozytoria projektu. Istniejący przegląd CHI 2025 obejmujący 135 prac o integracji LLM z XR stanowi dodatkową kontrolę zakresu, lecz niniejsza analiza jest szersza: obejmuje MAS, embodied AI, robotykę, neuroadaptację i automatyzację procesu wytwarzania, a nie wyłącznie LLM w XR. [4]

Reprezentatywna rodzina zapytań była następująca:

```
("extended reality" OR XR OR "virtual reality" OR VR OR
 "augmented reality" OR AR OR "mixed reality" OR MR)
AND
(agent* OR "agentic AI" OR "multi-agent" OR embodied OR
 "large language model" OR LLM OR VLM OR "vision-language-action")
AND
(spatial OR interaction OR automation OR robotics OR adaptation)

("virtual reality" OR "mixed reality")
AND ("large language model" OR multimodal OR VLM)
AND (scene OR environment OR spatial OR tool OR action)

(XR OR VR OR AR)
AND (EEG OR "eye tracking" OR pupil* OR EDA OR ECG OR HRV)
AND (adaptive OR "closed loop" OR workload OR stress)

(XR OR "mixed reality")
AND (ROS2 OR robot* OR "digital twin" OR teleoperation OR "shared autonomy")

(XR OR VR)
AND ("automated testing" OR test generation OR agent*)
```

Kryteria **włączenia** obejmowały: publikacje recenzowane, wysokiej jakości preprinty z jawną metodologią, działające prototypy z opisem architektury, oficjalne dokumentacje standardów/frameworków i repozytoria implementacji; pierwszeństwo otrzymały lata 2023–2026 oraz prace fundamentalne. Kryteria **wykluczenia** obejmowały czysto marketingowe demonstracje bez opisu technicznego, chatboty niepercepcyjne bez działania lub adaptacji, zwykłe modele predykcyjne niepodejmujące decyzji i projekty, których istnienia lub właściwości nie udało się potwierdzić.

Dowody klasyfikuję jako:

E — empiryczne, gdy istnieje eksperyment, benchmark lub user study;

P — działający prototyp, lecz o ograniczonej walidacji;

C — produkt/technologia komercyjna, której możliwości nie są dowodem skuteczności naukowej;

A — propozycja architektoniczna;

S — kierunek spekulatywny. Jest to istotne, ponieważ znaczna część literatury Agentic XR z lat 2025–2026 pozostaje na etapie małych prób użytkowników, demo papers albo short/adjunct papers. Przykładowo SpatiaLLM raportuje badanie na 11 osobach, a IEEE VR 2025 environment-aware agents — 14 uczestników; wyniki są obiecujące, lecz nie uzasadniają jeszcze generalizacji populacyjnych. [5]

Taksonomia, architektury agentów i rozumienie przestrzeni

Praktyczna definicja *agenta XR* powinna wymagać co najmniej pętli **obserwacja → aktualizacja stanu/modelu świata → wybór celu lub działania → wykonanie przez narzędzie/API → obserwacja skutku**. Sam model generujący tekst lub pojedynczą predykcję nie jest zatem agentem. ReAct formalizuje sprzężenie rozumowania i działania przez naprzemienne reasoning traces i actions; starsze BDI rozdziela beliefs, desires i intentions, a subsumption architecture pokazuje inną, nadal aktualną zasadę: szybkie zachowania reaktywne mogą funkcjonować niezależnie od wolniejszej deliberacji. [6]

Architektura	Typowa latencja i koszt	Autonomia	Przydatność w XR	Główne ograniczenie
reaktywna / finite-state / behavior tree	ms	niska–średnia	bardzo wysoka dla locomotion, UI, safety	ograniczona generalizacja
BDI / symbolic planning	dziesiątki ms–s	średnia	dobra dla procedur i audytowalnych planów	modelowanie symboli
hybrydowa reactive–deliberative	ms w pętli lokalnej + s dla planu	wysoka	**najlepszy ogólny wzorzec XR**	złożona synchronizacja
LLM ReAct/tool-use	zwykle setki ms–sekundy	średnia–wysoka	dialog, planowanie, authoring	niestabilność i koszt inferencji
plan-and-execute	sekundy dla planu, szybkie wykonanie podzadań	wysoka	zadania wieloetapowe	plan szybko się dezaktualizuje
VLM/MLLM agent	wysoki koszt percepcji	wysoka semantycznie	real-world AR/MR	niepewna geometria 3D
VLA	zależna od modelu, często GPU	wykonawcza	robotyka / manipulacja	bezpieczeństwo, embodiment mismatch
neuro-symboliczna	zmienna	wysoka	procedury, compliance, constraints	koszt integracji
MAS	suma kosztów agentów + komunikacja	potencjalnie wysoka	rozdzielanie odpowiedzialności	narzut i konflikty

Foundation models nie eliminują tradycyjnych architektur. Przeciwnie, w real-time XR ich najlepszą rolą jest często **wolniejsza warstwa semantyczna**, podczas gdy tracking, kolizje, sterowanie, ograniczenia bezpieczeństwa, animacja i feedback wymagają deterministycznego kodu. LangGraph wprost wspiera łączenie deterministycznych węzłów z agentic steps oraz human-in-the-loop; AutoGen historycznie wspierał event-driven multi-agent orchestration, choć dokumentacja Microsoftu z 2026 oznacza projekt jako zarchiwizowany i kieruje nowych

użytkowników do Microsoft Agent Framework; Semantic Kernel pozostaje SDK pozwalającym łączyć modele z funkcjami i pluginami. [7]

Najważniejsza różnica między **autonomią decyzyjną**, **autonomią wykonawczą** i **adaptacyjnością** powinna być raportowana oddzielnie. LLMR ma znaczną autonomię decyzji dotyczącej konstrukcji sceny, ale wykonuje tylko ograniczony repertuar Unity operations; agent fizjologiczny może mieć niewielką autonomię semantyczną, lecz stale adaptować trudność; robot teleoperowany może mieć wysoką autonomię wykonawczą w trajektorii, mimo że cel wybiera człowiek. Mieszanie tych wymiarów prowadzi do nieporównywalnych twierdzeń o „autonomii".

Multimodalna percepcja. Fundament stanowią nadal klasyczne komponenty: SLAM/visual-inertial odometry, depth/LiDAR, hand/body/head/eye tracking, semantic segmentation i scene graph. Foundation model powinien otrzymywać z nich *ustrukturyzowany world state*, zamiast rekonstruować geometrię wyłącznie z renderowanego obrazu. SpatialRGPT zwiększa przestrzenne możliwości VLM przez informacje regionowe i depth; SpatialLM operuje bezpośrednio na point clouds i generuje strukturalny opis wnętrza; LSceneLLM skupia reprezentację na istotnych obszarach dużej sceny. [8]

Odpowiedź na pytanie **„czy VLM naprawdę rozumie 3D?"** brzmi na listopad 2026: **nie w sensie stabilnego, geometrycznie spójnego modelu świata porównywalnego z systemem SLAM + metric map**. SpinBench wykazuje egocentric bias i słabą odporność na rotację; Disjoint-3DQA wykazuje duży spadek jakości, kiedy dwa obiekty nie są widoczne w tym samym ujęciu, natomiast podanie oracle 3D coordinates znacząco pomaga; SCOPE diagnozuje niespójność między perspektywami; Ego3D-Bench nadal pokazuje lukę względem ludzi. [9]

Najlepszy eksperyment rozróżniający „prawdziwą" reprezentację 3D od korelacji vision-language powinien rozdzielić dane wizualne od geometrycznych. Model należy testować na: nowych viewpointach; zmianie układu przy zachowaniu semantyki; relacjach między obiektami nigdy niewidzianymi równocześnie; metric distance/angle; rotacji całej sceny; symetrycznych konfiguracjach; częściowych obserwacjach; zadaniu „wróć do wcześniej widzianego obiektu"; kontrfaktycznych zmianach sceny. Dodatkowym warunkiem powinno być porównanie RGB-only, RGB+depth, RGB+trajectory i oracle-3D. Wzorzec wyników Disjoint-3DQA sugeruje, że właśnie takie ablations ujawniają, czy ograniczeniem jest reasoning, czy brak konstrukcji mapy. [10]

Pamięć. Dla XR znacznie bardziej wartościowa od klasycznej conversational memory jest pamięć epizodyczno-przestrzenna:

```
episode:
  t0
  user_pose
  gaze_target
  visible_objects
  scene_graph_delta
  action_taken
  expected_effect
  observed_effect
  confidence
```

RAG jest użyteczny do dokumentacji, instrukcji i deklaratywnej wiedzy, ale **nie zastępuje world model**. Vector database przechowująca tekstowe opisy „stół jest obok drzwi" może stać się niespójna po przesunięciu stołu. Dlatego spatial memory powinna używać trwałych IDs obiektów, transformacji, timestamps i confidence oraz aktualizować relacje z silnika/SLAM.

Zastosowania: sterowanie XR, HAI, neuroadaptacja, robotyka i MAS

Sterowanie środowiskiem. Najbardziej wykonalne działania agentowe to te, które da się przedstawić jako ograniczony zestaw narzędzi, np.:

```
CreatePrimitive(type, pose, scale)
InstantiatePrefab(asset_id, pose)
MoveObject(object_id, pose)
SetMaterial(object_id, material_id)
SetUIAnchor(panel_id, surface_id)
PlayAnimation(actor_id, clip_id)
SetTaskDifficulty(level)
SpawnScenario(template_id, parameters)
QuerySceneGraph(filter)
RunXRTest(test_id)
```

LLMR pokazał już real-time creation/modification w Unity, łącząc scene understanding, planning, self-debugging i memory; nowszy Co-VREUD wykorzystuje structured outputs i jawny katalog CRUD tools, a agent może modyfikować tylko elementy przewidziane przez API. Jest to wzorzec znacznie bezpieczniejszy niż generowanie dowolnego C# i jego natychmiastowe wykonanie. [11]

Agent działający **na poziomie engine API** zna IDs, transformacje, navmesh, collider i strukturę sceny; agent działający **przez interfejs jak użytkownik** widzi głównie rendering i emuluje input. Pierwszy jest bardziej precyzyjny i lepszy do authoringu/testowania; drugi jest cenny w black-box testing oraz wtedy, gdy system nie udostępnia internal state. XRintTest reprezentuje podejście strukturalne: konstruuje XR User Interaction Graph i syntetyzuje przestrzenne interakcje, osiągając 97% coverage dla trigger/grab w raportowanym benchmarku. [12]

Wirtualni współpracownicy. W VR największa wartość nie pochodzi z samego dialogu, lecz ze spójności języka z przestrzenią, gestem i zachowaniem niewerbalnym. Environment-aware agents IEEE VR 2025 otrzymywały opis sceny i interakcji; User-Aware Virtual Agents generują zarówno mowę, jak i zachowania niewerbalne na podstawie zachowania rozmówcy; AgentHands w 2026 synchronizuje GestureEvents z wypowiedzią, umieszczając wskazania i manipulacyjne gesty w przestrzeni XR. [13]

Korzyść społeczna ma jednak koszt poznawczy. Im bardziej antropomorficzny agent, tym łatwiej użytkownik przypisuje mu intencję, świadomość sytuacyjną i kompetencje szersze niż rzeczywiste. Z tego powodu metryka „realism" nie powinna być maksymalizowana niezależnie od *calibrated trust*. Projekt powinien ujawniać ograniczenia i niepewność, szczególnie gdy agent ingeruje w zadania szkoleniowe, terapeutyczne lub safety-critical. EMBRACE pokazuje techniczną możliwość

łączenia emocjonalnej odpowiedzi, TTS, facial animation i HR/HRV, ale wniosków z takiego prototypu nie należy utożsamiać z kliniczną skutecznością. [14]

Neuroadaptive XR. Prawidłowa architektura zamkniętej pętli to:

```
eye/EEG/EDA/ECG/motion
        ↓
preprocessing + synchronization
        ↓
latent-state estimator + uncertainty
        ↓
policy / rule / planner
        ↓
bounded adaptation
        ↓
XR environment
        ↓
behaviour + physiology
        └──────── feedback ────────┘
```

Najbardziej realistyczne są obecnie pupilometria, eye tracking, HR/HRV, EDA i zachowanie. EEG daje bogatszą informację, lecz wymaga synchronizacji, kontroli artefaktów ruchowych i ostrożnej interpretacji. Praca Nolte i wsp. pokazuje metodologiczną wykonalność jednoczesnego EEG i eye trackingu w VR, a guideline z 2026 podkreśla znaczenie alignmentu strumieni, timestampingu, interferencji sprzętowej i LSL. [15]

Kluczowy problem metodologiczny to **construct validity**. Pupil dilation nie jest specyficznym „czujnikiem cognitive load"; HRV nie jest jednoznacznym „czujnikiem stresu", a EEG workload classifiers są silnie zależne od osoby i zadania. Dlatego agent nie powinien wykonywać dużej adaptacji na podstawie pojedynczego biomarkera. Lepszy model to fusion kilku sygnałów + confidence + indywidualna kalibracja + ograniczone zakresy adaptacji. Badania VR z eye trackingiem potwierdzają możliwość obserwowania zmian pupil/fixation/saccade związanych z warunkami zadania, ale nie uprawniają do prostego odczytywania stanów mentalnych. [16]

Najważniejszy projekt eksperymentalny powinien porównywać: fixed XR, rule-based adaptation, supervised state estimator + deterministic adaptation, RL contextual policy i agentic planner. W przeciwnym razie efekty „AI adaptation" będą mieszać korzyść z samej personalizacji z korzyścią wynikającą z autonomicznego planowania.

XR–robotics–digital twins. Sensowny wzorzec integracyjny to:

```
XR UI → intent → agent planner → symbolic task plan
                      ↓
              safety validator
                      ↓
            deterministic executor
                      ↓
                ROS 2 actions
                      ↓
```

```
                robot controller

robot sensors → ROS 2 → digital twin → XR visualization
```

ROS 2 opiera komunikację na DDS/RTPS i zapewnia typowane wiadomości/usługi; Isaac Sim oferuje ROS 2 Bridge oraz symulację OpenUSD, co umożliwia walidowanie planów w cyfrowym bliźniaku przed wysłaniem do sprzętu. [17]

Porównanie modeli nadzoru:

Model	Człowiek	Agent	Najlepsze użycie
Human-in-the-loop	zatwierdza krytyczne kroki	proponuje	nieznane/rzadkie zadania
Human-on-the-loop	nadzoruje i może przerwać	wykonuje plan	powtarzalne procesy
Shared autonomy	ciągłe współdzielenie kontroli	wspomaga traje ktorię/intencję	teleoperacja
autonomous + safety supervisor	interweniuje wyjątkowo	działa autonomi cznie	dopiero po formalnej walidacji ograniczeń

Teleoperacja MR humanoida wykazała korzyść z assistive autonomy opartej na movement primitives i affordance templates; MIRAGE łączy gaze, ruch robota i kontekst w estymacji intencji; IMMERTWIN stosuje closed-loop digital twin jako pośrednika. Są to empirycznie znacznie lepiej uzasadnione wzorce niż „LLM → cmd_vel". [18]

MAS. Multi-agent nie powinien być ustawieniem domyślnym. Dokumentacja AutoGen sama zaleca zaczynać od pojedynczego agenta i przechodzić do zespołu, gdy pojedyncza jednostka jest niewystarczająca. MAS ma sens, gdy role różnią się *modelem czasowym, uprawnieniami lub kryteriami sukcesu*, np. perception agent działa ciągle, planner sporadycznie, safety agent veto'uje działania, evaluator pozostaje niezależny. [19]

Referencyjny podział ról to **Perception → World Model → Intent → Planner → Scene/Robot Executor**, z niezależnymi **Safety Monitor** i **Evaluation/Telemetry Agent**. Nie należy natomiast używać kilku LLM-ów wyłącznie po to, aby „dyskutowały" o każdym poleceniu; zwiększa to latencję, koszt i powierzchnię błędu bez gwarancji poprawy.

Technologie, infrastruktura i automatyzacja cyklu życia XR

Unity, Unreal, OpenXR. OpenXR 1.1 pozostaje podstawową warstwą przenośności XR; Khronos włączył do core funkcjonalności wcześniej rozproszone w rozszerzeniach. Unity oferuje OpenXR provider i XR Interaction Toolkit, a Unreal posiada natywne moduły OpenXR, input, hand/eye tracking i XR tooling. OpenXR rozwiązuje jednak przenośność device/runtime — **nie** definiuje standardu dla agentów, world models czy agentic tool APIs. [20]

Unity jest obecnie szczególnie atrakcyjne do eksperymentów Agentic XR ze względu na dużą liczbę opublikowanych prototypów naukowych i łatwe expose'owanie funkcji C# jako REST/WebSocket/IPC tools. Unreal jest korzystniejszy tam, gdzie istotne są zaawansowana grafika, symulacja i istniejąca infrastruktura C++/Blueprint. Godot może pełnić rolę reprodukowalnej open-

source platformy, ale ekosystem badań agentowych XR jest wyraźnie mniejszy. OpenXR powinien być traktowany jako warstwa urządzeń pod silnikiem, a nie konkurent dla Unity/Unreal. [21]

REST/WebSocket/gRPC/ROS 2. REST nadaje się do rzadkich operacji planistycznych; WebSocket do streaming events i partial responses; gRPC do typowanych, wydajnych usług agentowych; ROS 2 do sensorów, robotów i distributed control. Krytycznej pętli renderowania i head tracking nie należy uzależniać od żadnego z tych kanałów sieciowych.

Dla systemu 90 Hz budżet jednej klatki wynosi około **11,1 ms**. LLM wykonujący decyzję przez 300 ms–3 s nie może być elementem motion-to-photon loop.

Z tego wynika fundamentalna architektoniczna separacja:

XR real-time loop: 72–120 Hz
physics/control loop: ~50–500+ Hz zależnie od systemu
perception update: ~5–60 Hz
agent semantic loop: ~0.2–5 Hz
long-horizon planning: on demand

Są to zakresy projektowe, a nie normatywne wartości; główną zasadą jest asynchroniczność i brak blokowania render thread przez model generatywny. LoXR potwierdza, że lokalne LLM są już uruchamialne na urządzeniach klasy Quest 3, Magic Leap 2 i Vision Pro, lecz występują kompromisy między processing speed, consistency i battery consumption. [22]

Lokalnie vs edge vs cloud.

Model wdrożenia	Zalety	Wady	Najlepsze przypadki
local	prywatność, offline, deterministyczny transport	RAM/GPU/energia	speech, mały LLM, klasyfikacja intencji
edge workstation	duże modele przy małym RTT LAN	dodatkowy komputer	laboratorium, PC VR
cloud	najlepsze modele i skalowanie	privacy, RTT, koszt, outage	planning niekrytyczny
hybrid	najlepszy kompromis	największa złożoność	docelowy Agentic XR

Rekomendowana architektura laboratoryjna jest **local-first**: tracking, scene model, safety, policy cache i fallback lokalnie; większy VLM/LLM na workstation; cloud opcjonalnie.

Frameworki agentowe. LangGraph dobrze pasuje do planowania, state machines, persistence i human approval; Semantic Kernel jest szczególnie naturalny dla Unity/C#, ponieważ modeluje narzędzia jako funkcje/pluginy; AutoGen jest wartościowy jako źródło wzorców MAS, ale w listopadzie 2026 Microsoft oznacza projekt jako archived i rekomenduje migrację; CrewAI może służyć do wysokopoziomowych workflow, lecz nie powinien sterować real-time XR loop. [23]

Pełny agentowy lifecycle XR powinien wyglądać następująco:

```
requirements
  ↓
agent: requirement decomposition
  ↓ human approval
scene/code skeleton
  ↓
asset generation + licence validation
  ↓
static analysis + build
  ↓
automated scene inspection
  ↓
XR interaction tests
  ↓
performance tests on target HMD
  ↓
agent log triage + repair proposal
  ↓ human review
security/privacy audit
  ↓
signed release artifact
```

Agent może samodzielnie generować boilerplate, test fixtures, prefab variants, documentation i analizować logi. Człowiek powinien pozostać decydentem przy zmianie wymagań, privacy/security acceptance, UX acceptance, interpretacji wyników z uczestnikami oraz wydaniu wersji safety-critical. XRintTest dostarcza już empirycznego dowodu, że znaczna część spatial-interaction testing może zostać zautomatyzowana; case study LLM-based VR testing wskazuje jednak ograniczenia spatial understanding modeli patrzących jedynie na screenshots. [24]

Bezpieczeństwo, prywatność, niezawodność

Agentic XR poszerza klasę prompt injection, ponieważ „promptem" może stać się **tekst na plakacie, głos osoby trzeciej, QR code, ekran komputera, etykieta obiektu albo grafika niewidoczna świadomie dla użytkownika**. Wykazano empirycznie możliwość image-based prompt injection przeciw współczesnym VLM, w tym ataki ukryte w obrazowaniu; w XR oznacza to, że dane sensoryczne muszą być traktowane jako **niezaufane wejście**, a nie jako część system promptu. [25]

Najważniejsza reguła bezpieczeństwa brzmi:

```
perception ≠ authority
language model ≠ actuator
plan ≠ executable command
```

Każda akcja powinna przechodzić przez *typed capability layer* zawierającą allow-listę narzędzi, ograniczenia parametrów, stan autoryzacji, rate limits i reguły zależne od stanu świata. Dla fizycznych urządzeń następna warstwa powinna niezależnie sprawdzać workspace, velocity/force limits, kolizje, interlocks i emergency stop. Execution log musi przechowywać nie tylko finalny command, lecz także źródło obserwacji, wersję modelu, plan, walidację i rezultat.

Dla XR szczególnie niebezpieczne są również **perceptual attacks**: agent może zasłonić realne zagrożenie, przesunąć overlay, błędnie wskazać obiekt lub skierować uwagę użytkownika. Safety evaluation musi więc mierzyć nie tylko „czy model podał szkodliwą odpowiedź", lecz także *spatial misguidance*: odległość wskazania od celu, czas pozostawania hazardu poza FOV, kolizje UI z realnymi przeszkodami i błędne occlusion.

Dane eye-tracking, pupilometryczne, głosowe, mapy pomieszczeń i sygnały fizjologiczne mogą ujawniać bardzo wrażliwe cechy użytkownika. GDPR definiuje dane biometryczne i szczególne kategorie danych; EU AI Act dodatkowo ogranicza niektóre zastosowania emotion recognition, m.in. w workplace i education z określonymi wyjątkami. Na dzień 3 listopada 2026 część AI Act już obowiązuje, natomiast harmonogram obowiązków high-risk pozostaje etapowy. [26]

W praktyce neuroadaptive XR dla edukacji i środowiska pracy wymaga zatem szczególnej ostrożności prawnej: „estimating cognitive workload" może być metodologicznie i regulacyjnie czymś innym niż „emotion recognition", ale nie wolno zakładać ich równoważności ani automatycznie wyłączać projektu spod regulacji. Każdy deployment wymaga analizy konkretnego intended purpose.

ISO 9241-210 pozostaje właściwą normą odniesienia dla human-centred design systemów interaktywnych; w robotyce safety-critical właściwe mogą być standardy branżowe i IEC 61508-family, ale zastosowanie zależy od konkretnego systemu, nie od faktu użycia XR. Projekt trzeciej edycji IEC 61508-1:2026 explicite uwzględnia AI, human factors i security na poziomie hazard/risk analysis, ale w dniu 3 listopada 2026 publikacja znajdowała się jeszcze na etapie pre-release/FDIS. [27]

Metryki eksperymentalne.

Grupa	Minimum
A. techniczne	p50/p95/p99 agent latency, frame drops, FPS, CPU/GPU/RAM, energia, disconnect recovery
B. kompetencje	task success, plan validity, tool-call error, recovery rate, human interventions
C. HCI	SUS/UMUX-Lite, NASA-TLX, presence, agency, trust calibration, task time
D. safety	unsafe action rate, constraint violations, false veto, detection latency, recovery success
E. wartość	time saved, learning gain, transfer, retention, error reduction

Eksperyment bazowy powinien obejmować pięć ramion wskazanych w pytaniu: **XR**, **XR+rules**, **XR+single agent**, **XR+MAS**, **XR+agent+human supervision**. Kluczowe jest, by scenariusze, modele percepcji, interfejs i task difficulty były takie same; inaczej efekt „multi-agent" będzie splątany z większym budżetem obliczeniowym albo lepszym promptem.

Dla badań HCI rekomenduję within-subject tam, gdzie learning/carry-over można kontrolować Latin square; międzygrupowe rozwiązanie jest lepsze dla treningu długoterminowego. Wielkość próby należy ustalać *a priori* przez power analysis dla primary endpoint, a nie na podstawie konwencjonalnego N=20. Raportowanie powinno zawierać effect sizes i confidence intervals, modele mieszane przy repeated measures, korektę multiplicity dla wielu outcomes oraz publikację

preregistration, scen i logów. Badania neurofizjologiczne wymagają ponadto test–retest i walidacji cross-subject vs personalized models.

Systemy istniejące, luki badawcze i propozycje publikacji

Poniższa tabela celowo miesza systemy stricte agentowe
z blisko spokrewnionymi prototypami automatyzacji i shared autonomy, ale oznacza ich status.

System / autorzy	Rok	XR / architektura	Autonomia	Dowód / kod	Najważniejsze ograniczenie
LLMR — De La Torre et al.	2024	MR, LLM + planning + memory + self-debug	wysoka authoring	E/P; GitHub	małe N, silna zależność od narzędzi Unity [28]
SIGMA — Bohus et al.	2024	MR task assistant, LLM/VLM	średnia	P; open-source	procedural tasks, wczesna ewaluacja [29]
LEGO-MRTA — Pei et al.	2024	MR, language agent + VLM + memory/planning	średnia–wysoka	P/dataset	syntetyczne dane, pilot [30]
User-Aware Virtual Agents — Gunawardhana et al.	2024	VR, CVAE + LLM	reaktywna/adaptive	E	skupienie na zachowaniach społecznych [31]
MR Humanoid Teleoperation Assistance	2024	MR + assistive autonomy	shared	E	task-specific primitives [32]
IMMERTWIN — Audonnet et al.	2024	MR + closed-loop digital twin	shared	E/P; kod	teleoperation, nie general agent [33]
ROS2 MR HMI — Janecký et al.	2024	HoloLens 2 + Unity + ROS 2	niska	P	HMI bardziej niż agent [34]
EmBARDiment — Bovo et al.	2025	XR, gaze/context/memory + LLM	proactive	P/E	ograniczone dowody długoterminowe [35]

System / autorzy	Rok	XR / architektura	Autonomia	Dowód / kod	Najważniejsze ograniczenie
Environment-Aware VR Agents — Li, Zhang, Peng, Peiris	2025	VR, scene schema + LLM	średnia	E, N=14	tekstualizacja sceny [36]
Sensible Agent — Lee et al.	2025	AR/VR, multimodal proactive LMM	proactive	E/P	N=10 final study [37]
GOALIE	2025	proactive multimodal AR	średnia	P	procedural focus [38]
XRintTest — Gu, Rojas, Shin	2025	Unity XR testing agent/graph	wysoka testowa	E; open-source	ograniczony zestaw interactions [39]
When to Call the LLM — Wells, Wang, Maurer	2025	adaptive speech→XR selection	adaptacyjna	E/P; GitHub	nie jest pełnym autonomous agent [40]
MIRAGE — Sun et al.	2025	VR teleoperation + gaze + multimodal CNN	shared	E	ograniczony domain [41]
SpatiaLLM	2026	MR MAS + constraint-verified spatial memory	wysoka placement	E/P, N=11	short paper, mała próba [42]
AgentHands — Liu et al.	2026	XR LLM + spatial gestures	średnia	E/P; XR Blocks addon	interaction/gesture focus [43]
EMBRACE	2026	VR affective embodied agent	adaptive	E/P	afekt ≠ kliniczna skuteczność [44]
VAPS	2026	VR virtual patient + LLM	konwersacyjna	E, N=40	szkolenie komunikacyjne [45]
Safety-Critical Unity XR	2026	VR/AR, RAG + EEG/vision/	średnia	P/ preliminary	głównie qualitative evaluation [46]

System / autorzy	Rok	XR / architektura	Autonomia	Dowód / kod	Najważniejsze ograniczenie
Agents		function calls			
XARA	2026	Unity XR, dual-LLM experiment authoring	średnia–wysoka	preprint/P	świeży preprint, brak szerokiej walidacji [47]

Najważniejsze **luki badawcze** wraz z oceną wartości:

Luka	Nowość	Wykonalność	Wkład praktyczny
benchmark agentic XR łączący 3D + tool use	bardzo wysoka	wysoka	bardzo wysoki
trwała metric spatial memory	bardzo wysoka	wysoka	bardzo wysoki
wykrywanie multimodal prompt injection w realnym AR	bardzo wysoka	średnio-wysoka	krytyczny
latency-aware orchestration agentów XR	wysoka	bardzo wysoka	wysoki
formalny model autonomy/safety metrics	wysoka	wysoka	bardzo wysoki
neuroadaptive agent z calibrated uncertainty	bardzo wysoka	średnia	wysoki
porównanie single-agent vs MAS przy równym compute	wysoka	bardzo wysoka	wysoki
shared autonomy z LLM planner + deterministic robot executor	bardzo wysoka	średnia	bardzo wysoki
long-horizon egocentric spatial reasoning	wysoka	wysoka	bardzo wysoki
automated UX testing z embodied test agent	wysoka	bardzo wysoka	bardzo wysoki
memory/privacy trade-off w always-on XR	bardzo wysoka	wysoka	krytyczny
benchmark	bardzo wysoka	wysoka	bardzo wysoki

Luka	Nowość	Wykonalność	Wkład praktyczny
recovery/self-correction po błędzie przestrzennym			

Macierz zastosowanie × agent × XR × dojrzałość × publikowalność.

Obszar	Typ agenta	XR	Dojrzałość 2026	Potencjał publikacyjny
scene authoring	tool-using LLM	VR/MR	średnio-wysoka	średni
spatial UI placement	MAS + constraints	MR	średnia	wysoki
procedural guidance	VLM+planner	AR/MR	średnio-wysoka	wysoki
social agents	multimodal ECA	VR	średnia	średnio-wysoki
neuroadaptation	physiological agent	VR	niska–średnia	**bardzo wysoki**
automated testing	graph/LLM agent	VR/MR	wysoka	**bardzo wysoki**
XR–robot shared autonomy	hybrid agent	MR	średnia	**bardzo wysoki**
security monitor	symbolic/VLM agent	AR/MR	niska	**bardzo wysoki**
spatial memory	embodied agent	AR/MR	niska–średnia	**bardzo wysoki**
MAS orchestration	role-specialized MAS	XR	niska–średnia	wysoki

Osiem konkretnych **propozycji artykułów**:

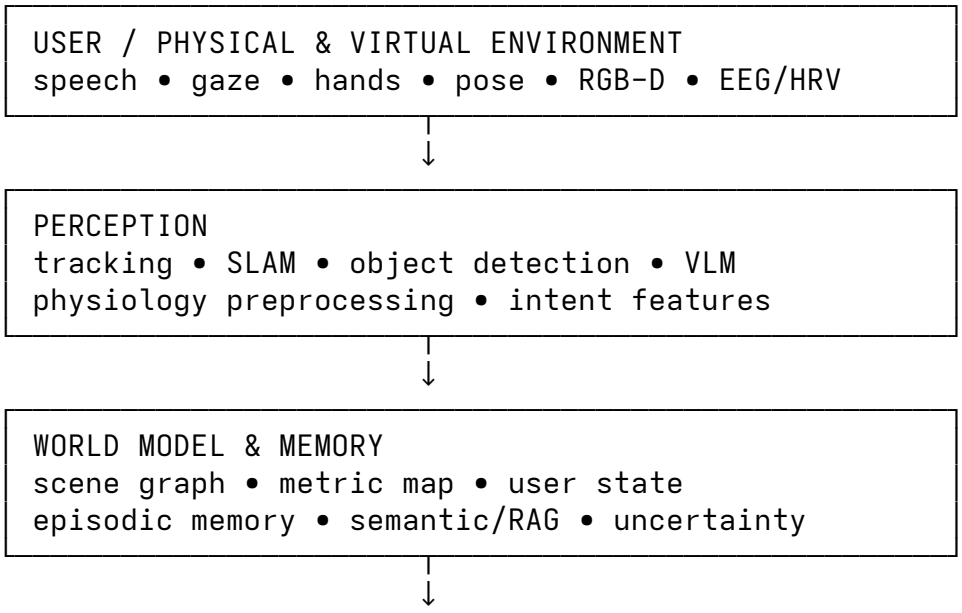
Roboczy tytuł	Hipoteza / nowość	Metoda	Zasoby	Venue
“AgentXR-Bench: A Reproducible Benchmark for Spatially Grounded Tool-Using XR Agents”	structured 3D state > screenshot-only agent	5 architectures × 100–500 tasks	Unity/ OpenXR, local VLM	IEEE VR, ISMAR, TVCG
“Do Vision-Language Agents Maintain 3D World Models?”	modele zawodzą po viewpoint/occlusion bez metric memory	controlled scene perturbations	Unity synthetic scenes	CVPR/ICCV, IEEE VR
“Safety-Gated Agentic XR Under Environmental Prompt Injection”	typed safety gateway ograniczy unsafe tool calls	adversarial text/audio/image objects	Quest/PCVR	USENIX Security/IEEE S&P workshops, ISMAR
“Single Agent or Multi-Agent? Controlled Evaluation Under Equal Compute Budgets in XR”	MAS pomaga tylko przy separacji ról	equal-token/ equal-latency ablation	Unity + local LLM	AAMAS, IEEE VR
“Uncertainty-Calibrated Neuroadaptive VR”	uncertainty-aware fusion >	EEG+eye+HRV, within-subject	eye-tracked HMD + EEG	IEEE T-AFFC, IJHCS, CHI

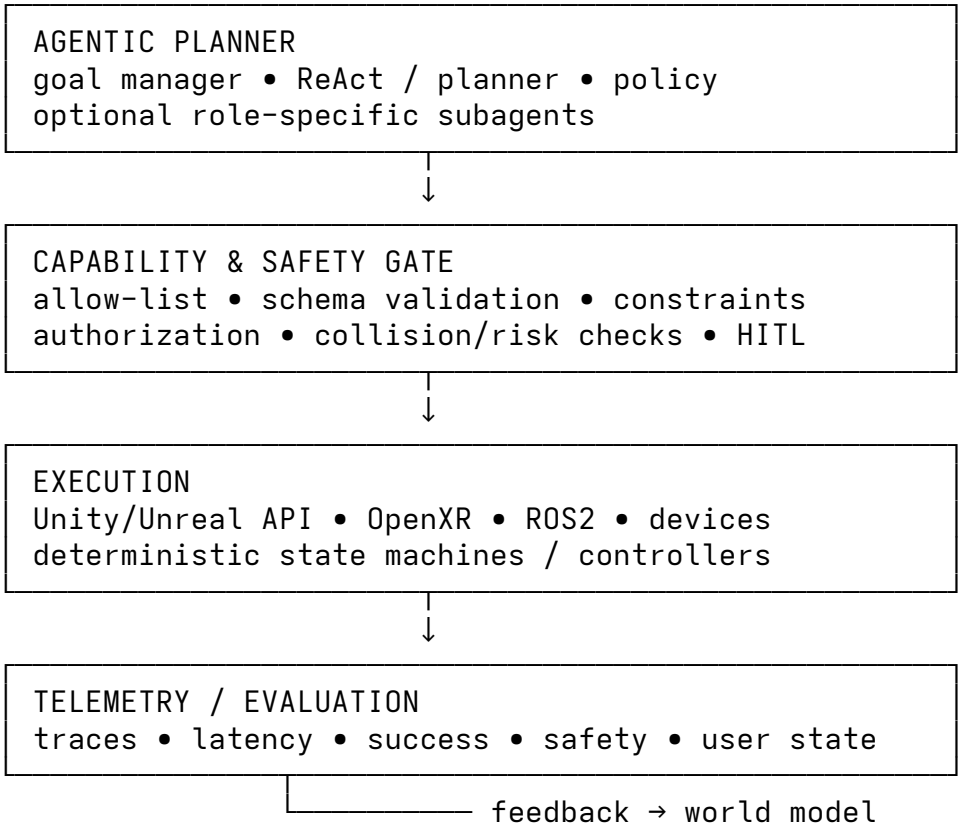
Roboczy tytuł	Hipoteza / nowość	Metoda	Zasoby	Venue
	threshold adaptation			
“Shared Autonomy Through XR: LLM Planning With Deterministic Robot Execution”	planner improves task time bez wzrostu unsafe actions	simulation→robot study	ROS2 + Isaac Sim + manipulator	ICRA/IROS/ HRI
“Embodied Test Agents for Spatial Interaction Regression Testing”	goal-directed agent > random/script testing	mutation testing, interaction coverage	Unity/ XRBench-like corpus	ASE/ICSE
“Private Spatial Memory for Always-On XR Agents”	local structured memory zachowuje utility przy mniejszej ekspozycji danych	local/cloud comparison	Quest/PC + local models	UIST/CHI/ PoPETs

Pierwsze cztery i automated testing można przygotować **bez uczestników**, opierając się na symulacji i benchmarkach. Neuroadaptive VR, calibrated trust, social interaction i część shared autonomy wymagają uczestników; fizyczny wariant robotics wymaga także hardware i procedury safety.

Referencyjna architektura Agentic XR, roadmapa i ranking badań

Najbezpieczniejsza referencyjna architektura nie jest „LLM-centric”; jej centrum stanowi **world state + capability/safety layer**.





Przepływ przykładowego zadania. Użytkownik mówi „przenieś instrukcję obok urządzenia, którego teraz używam". Gaze i hand tracking wskazują kandydatów; scene graph podaje ID i geometry; intent agent interpretuje „urządzenie"; planner proponuje SetUIAnchor; safety gate sprawdza, czy panel nie zasłoni realnego hazardu i czy znajduje się w ergonomicznym FOV; deterministic executor przesuwa panel; telemetry zapisuje rezultat; użytkownik może cofnąć działanie. Model językowy nigdy nie ustawia transformacji samodzielnie bez validation.

Trzy warianty:

Wariant	Elementy	Koszt/latencja	Autonomia	Safety
Minimal research	Unity + scene graph + 1 local/cloud LLM + 10 tools	niski	średnia	allow-list + confirmation
Local modular	local VLM/LLM + vector/graph DB + safety engine	średni	wysoka	wysoka odporność offline
Scalable MAS	distributed perception/planner/safety/evaluator	wysoki	najwyższa	potencjalnie najlepsza, lecz trudniejsza walidacja

OpenXR powinien obsługiwać urządzenia, Unity/Unreal — world execution, a warstwa agentowa działać poza frame loop. Dla robotyki po safety gate należy umieścić ROS 2 action/service layer i osobny safety controller. Isaac Sim lub inny digital twin może realizować *shadow execution* przed wysłaniem polecenia fizycznego. [48]

Roadmapa 12 miesięcy.

Miesiące	Zadanie	Kryterium zakończenia
1–2	systematic mapping + hipoteza + benchmark spec	preregistered protocol, ≥3 baselines
2–3	OpenXR/Unity infrastructure + logging	deterministic replay
3–4	scene graph/world model	≥95% poprawnych object IDs w test scenes
4–5	tool API + single agent	≥80% syntetycznych zadań
5–6	local models + latency profiling	p95 raportowany dla całego pipeline
6–7	safety gate + adversarial tests	zero direct unrestricted execution
7–8	benchmark suite	≥100 parametrized tasks
8–9	MAS / ablations	equal-compute comparison
9–10	eksperyment główny	frozen protocol + logs
10–11	analiza/replikacja	effect sizes + bootstrap CI
11–12	paper + release	code, scenes, seeds, model config, data

Największym ryzykiem harmonogramu nie jest implementacja LLM, lecz integracja tracking/perception, wersjonowanie scen, przygotowanie benchmarku i uzyskanie stabilnej reprodukowalności. Z tego powodu **benchmark-first** jest bardziej rozsądny niż budowanie od razu rozbudowanego wirtualnego asystenta.

Ranking tematów badawczych. Oceny 0–100 są **ekspercką heurystyką**, nie pomiarem empirycznym. Zastosowano dokładnie wskazane w pytaniu wagi: oryginalność 25%, wykonalność 25%, publikowalność 20%, dostępność narzędzi/danych 15%, reprodukowalność 15%.

Rank	Temat	Wynik /100
1	Agentic XR automated testing	**92,1**
2	Benchmark spatial reasoning + tool use	**91,2**
3	Safety-constrained Agentic XR / prompt injection	**89,5**
4	Local/edge multimodal XR agent latency	**89,2**
5	Agentic natural-language 3D scene authoring	**88,4**
6	Spatial memory for long-horizon XR agents	**87,7**
7	Proactive gaze-aware XR assistant	**87,6**

Rank	Temat	Wynik /100
8	MAS with real-time safety supervisor	**85,2**
9	Closed-loop neuroadaptive XR agent	**82,2**
10	Shared autonomy XR–robotics	**82,0**

Ranking preferuje automated testing, ponieważ łączy wysoką reprodukowalność z istniejącą infrastrukturą i możliwością wykonania setek eksperymentów bez uczestników. Neuroadaptacja i robotyka są niżej nie z powodu mniejszej wartości naukowej, lecz ze względu na większy koszt metodologiczny, sprzętowy i regulacyjny.

Analiza wrażliwości. Jeżeli zwiększyć wagę oryginalności kosztem wykonalności i reprodukowalności, w górę przechodzą neuroadaptive XR, shared autonomy i spatial memory. Jeżeli zwiększyć wagę wykonalności/reprodukowalności, dominują automated testing, local/edge inference i scene authoring. Najbardziej stabilne w ścisłej czołówce pozostają **benchmark agentic XR**, **automated testing** i **safety/prompt-injection**; są więc najmniej zależne od arbitralnego doboru wag.

Trzy rekomendowane kierunki rozpoczęcia prac.

AgentXR-Bench — spatial reasoning + tool use. Pierwszy eksperyment: przygotować 20 parametryzowanych scen Unity × 10 zadań, czyli 200 instances. Zadania powinny wymagać kombinacji obserwacji, relacji 3D, planu i wywołania 1–5 narzędzi. Porównać: text scene graph + LLM; screenshot+VLM; RGB-D+VLM; structured 3D world model+LLM; structured model + episodic memory. Głównym endpointem będzie task success; dodatkowo spatial error, invalid tools, recovery rate, p95 latency. Sceny powinny być automatycznie randomizowane, by ograniczyć memorization. Najważniejsza hipoteza: **jawna reprezentacja geometryczna poprawi generalizację bardziej niż zwiększenie wielkości samego VLM**. Taki wynik bezpośrednio odpowiada na obserwowane słabości SpinBench, Disjoint-3DQA i Ego3D-Bench. [49]

Safety-Gated Agentic XR. Pierwszy eksperyment: zbudować 10 tool APIs o trzech poziomach ryzyka i 500 adversarial episodes zawierających prompt injection w mowie, teksturach, labelach i dokumentach w scenie. Porównać unrestricted agent, prompt-only guardrails, schema validation i pełny capability/safety gate. Primary outcome: unsafe-action rate; secondary: false refusals, task success, safety latency. Warunkiem sukcesu publikacyjnego byłoby wykazanie dużej redukcji unsafe actions przy niewielkim spadku utility. Badanie może być w całości symulacyjne i otwarte, a jego znaczenie jest wzmacniane przez udowodnioną podatność VLM na visual prompt injection. [25]

Embodied XR Test Agent. Pierwszy eksperyment: rozszerzyć ideę XRintTest o planner mający cele semantyczne, np. „sprawdź wszystkie sposoby przeniesienia obiektu A do socketu B po zmianie konfiguracji sceny”. Porównać random exploration, scripted test, XRintTest-like graph search, single LLM agent i hybrid graph+LLM. Użyć mutation testing: celowo wprowadzić błędne collidery, event bindings, unreachable objects, błędne constraints i performance regressions. Primary endpoints: defect detection recall oraz interaction coverage per minute; następnie koszt inferencji i reprodukowalność. To projekt o bardzo dobrym stosunku ryzyka do szans

publikacyjnych, ponieważ istnieją już otwarte precedensy XRintTest i XRBench3D, ale agentowa semantyczna eksploracja nadal nie jest rozwiązana. [50]

Konkluzja. Technicznie wykonalne już dziś są: agentowe authoring/editing scen, procedural assistants, spatially grounded conversational agents, proaktywne multimodalne assistance, automated XR testing oraz wysokopoziomowe planowanie z ograniczonym tool API. Naukowo obiecujące, ale mniej dojrzałe są trwała spatial memory, neuroadaptive closed loops, MAS ze ścisłymi wymaganiami real-time i LLM-mediated shared autonomy. Najmniej uzasadnione jest obecnie przekazywanie foundation modelowi bezpośredniej kontroli nad krytyczną pętlą renderowania lub fizycznymi aktuatorami. W 2026 najbardziej defensywną architekturą jest **foundation model jako deliberative semantic planner, geometrycznie ugruntowany world model jako źródło stanu, deterministyczny executor jako warstwa działania oraz niezależny safety supervisor jako bramka uprawnień**. Taki układ wykorzystuje przewagi foundation models, nie zakładając przy tym, że posiadają one niezawodne rozumienie 3D, timing czasu rzeczywistego lub właściwości safety-critical. [51]

Wybrana bibliografia z DOI / zweryfikowanymi odnośnikami. De La Torre F. et al., *LLMR: Real-time Prompting of Interactive Worlds using Large Language Models*, CHI 2024. [52] Wang J. et al., *Survey of Large Language Models in Extended Reality*, preprint 2025. [53] LLM Integration in Extended Reality, CHI 2025, DOI 10.1145/3706598.3714224. [54] Li Z. et al., *Exploring Large Language Model-Driven Agents…*, IEEE VR 2025, DOI 10.1109/VR59515.2025.00025. [55] Gunawardhana B.S. et al., *Toward User-Aware Interactive Virtual Agents*, ISMAR 2024, DOI 10.1109/ISMAR62088.2024.00123. [56] Pei J. et al., *Autonomous Workflow for Multimodal Fine-Grained Training Assistants Towards Mixed Reality*, Findings ACL 2024. [30] Gu R., Rojas J.M., Shin D., *XRintTest*, ASE 2025, DOI 10.1109/ASE63991.2025.00363. [57] Wells B., Wang N., Maurer F., *When to Call the LLM*, VRST 2025, DOI 10.1145/3756884.3766015. [58] Liu Z. et al., *AgentHands*, CHI 2026, DOI 10.1145/3772318.3790938. [59] *SpatiaLLM*, IEEE VRW 2026, DOI 10.1109/VRW70859.2026.00278. [42] *Mixed Reality Teleoperation Assistance for Direct Control of Humanoids*, IEEE RA-L 2024, DOI 10.1109/LRA.2024.3349809. [32] Nolte D. et al., EEG + eye tracking in VR, DOI 10.3758/s13414-024-02917-3. [60] Islam F. et al., cognitive load and eye movements in immersive VR, DOI 10.1016/j.apergo.2025.104720. [16] Brooks R.A., *A Robust Layered Control System for a Mobile Robot*, DOI 10.1109/JRA.1986.1087032. [61] Yao S. et al., *ReAct: Synergizing Reasoning and Acting in Language Models*, ICLR 2023. [62] SpatialRGPT, NeurIPS 2024. [63] SpatialLM, NeurIPS 2025. [64] LSceneLLM, CVPR 2025. [65] Khronos Group, *OpenXR 1.1 Specification*. [66] Regulation (EU) 2024/1689 — *Artificial Intelligence Act*. [67]

---

[1] [2] [11] [28] [52] LLMR: Real-time Prompting of Interactive Worlds using Large Language Models - Microsoft Research

https://www.microsoft.com/en-us/research/publication/llmr-real-time-prompting-of-interactive-worlds-using-large-language-models-3/

[3] [33] IMMERTWIN: A Mixed Reality Framework for Enhanced Robotic Arm Teleoperation

https://arxiv.org/abs/2409.08964

[4] [54] LLM Integration in Extended Reality: A Comprehensive Review of Current Trends, Challenges, and Future Perspectives - Illinois Experts

https://experts.illinois.edu/en/publications/llm-integration-in-extended-reality-a-comprehensive-review-of-cur/

[5] SpatiaLLM: A Multi-Agent Framework for Adaptive and User-Aware Mixed Reality Content Placement | IEEE Conference Publication | IEEE Xplore

https://ieeexplore.ieee.org/abstract/document/11489721

[6] [62] ReAct: Synergizing Reasoning and Acting in Language Models | ML Anthology

https://mlanthology.org/iclr/2023/yao2023iclr-react/

[7] [23] LangGraph overview - Docs by LangChain

https://langchain-ai.github.io/langgraph/?trk=article-ssr-frontend-pulse_little-text-block&utm_source=chatgpt.com

[8] [63] SpatialRGPT: Grounded Spatial Reasoning in Vision-Language Models

https://proceedings.neurips.cc/paper_files/paper/2024/hash/f38cb4cf9a5eaa92b3cfa481832719c6-Abstract-Conference.html

[9] [49] [51] SpinBench: Perspective and Rotation as a Lens on Spatial Reasoning in VLMs

https://proceedings.iclr.cc/paper_files/paper/2026/hash/724be4472168f31ba1c9ac630f15dec8-Abstract-Conference.html

[10] Out of Sight, Not Out of Context? Egocentric Spatial Reasoning in VLMs Across Disjoint Frames - ACL Anthology

https://aclanthology.org/2025.emnlp-main.816/

[12] [24] [50] [57] XRintTest: An Automated Framework for User Interaction Testing in Extended Reality Applications

https://doi.org/10.1109/ASE63991.2025.00363

[13] [55] Exploring Large Language Model-Driven Agents for Environment-Aware Spatial Interactions and Conversations in Virtual Reality Role-Play Scenarios | IEEE Conference Publication | IEEE Xplore

https://ieeexplore.ieee.org/abstract/document/10937341/media

[14] [44] Designing Emotionally Intelligent Embodied Agents for Immersive Virtual Reality Experiences | IEEE Conference Publication | IEEE Xplore

https://ieeexplore.ieee.org/abstract/document/11450021/metrics

[15] [60] Combining EEG and eye-tracking in virtual reality: Obtaining fixation-onset event-related potentials and event-related spectral perturbations.

https://pubmed.ncbi.nlm.nih.gov/38977612/

[16] Evaluating the effect of immersive virtual reality technologies on cognitive load, situation awareness, and engagement using eye movements - PubMed

https://pubmed.ncbi.nlm.nih.gov/41518841/

[17] Why ROS 2? open robotics » No vendor lock-in R

https://docs.ros.org/en/rolling/_downloads/adb1c3484c0da58a307d5e9ecc1c18e6/ros2-brochure-a4-print.pdf

[18] [32] Mixed Reality Teleoperation Assistance for Direct Control of Humanoids

https://doi.org/10.1109/LRA.2024.3349809

[19] Teams — AutoGen

https://microsoft.github.io/autogen/stable/user-guide/agentchat-user-guide/tutorial/teams.html

[20] Khronos Releases OpenXR 1.1 to Further Streamline Cross-Platform XR Development

https://www.khronos.org/news/press/khronos-releases-openxr-1.1-to-further-streamline-cross-platform-xr-development

[21] Developing for XR Experiences in Unreal Engine | Unreal Engine 5.8 Documentation | Epic Developer Community

https://dev.epicgames.com/documentation/unreal-engine/developing-for-xr-experiences-in-unreal-engine

[22] LLMs on XR (LoXR): Performance Evaluation of LLMs Executed Locally on Extended Reality Devices | IEEE Conference Publication | IEEE Xplore

https://ieeexplore.ieee.org/abstract/document/10973004

[25] Prompt injection attacks on vision language models in oncology | Nature Communications

https://www.nature.com/articles/s41467-024-55631-x

[26] Regulation - 2016/679 - PT - gdpr - EUR-Lex

https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AL%3A2016%3A119%3AFULL&uri=uriserv%3AOJ.L_.2016.119.01.0001.01.ENG

[27] ISO 9241-210:2019 - Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems

https://www.iso.org/standard/77520.html

[29] SIGMA: An Open-Source Interactive System for Mixed-Reality Task Assistance Research – Extended Abstract - Microsoft Research

https://www.microsoft.com/en-us/research/publication/sigma-an-open-source-interactive-system-for-mixed-reality-task-assistance-research-extended-abstract/?lang=zh-cn&utm_source=chatgpt.com

[30] Autonomous Workflow for Multimodal Fine-Grained Training Assistants Towards Mixed Reality - ACL Anthology

https://aclanthology.org/2024.findings-acl.240/

[31] Research New – Immersive Computing Lab

https://www.immersivecomputinglab.org/research-new/

[34] Using Mixed Reality for Control and Monitoring of Robot Model Based on Robot Operating System 2 | MDPI

https://www.mdpi.com/2079-9292/13/17/3554

[35] EmBARDiment: an Embodied AI Agent for Productivity in XR

https://research.google/pubs/embardiment-an-embodied-ai-agent-for-productivity-in-xr/

[36] Zhang Huadong

https://www.zhanghuadong.com/publications.html

[37] Sensible Agent: A Framework for Unobtrusive Interaction with Proactive AR Agent

https://research.google/pubs/sensible-agent-a-framework-for-unobtrusive-interaction-with-proactive-ar-agent/

[38] GOALIE (GOAL oriented IntErventions) Proactive Multimodal Agent to Assist Augmented Reality

https://research.google/pubs/goalie-goal-oriented-interventions-proactive-multimodal-agent-to-assist-augmented-reality/

[39] GitHub - ruizhengu/XRintTest · GitHub

https://github.com/ruizhengu/XRintTest

[40] Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology, VRST 2025, Montreal, QC, Canada, November 12-14, 2025 - researchr publication

https://researchr.org/publication/vrst-2025

[41] MIRAGE: Multimodal Intention Recognition and Admittance-Guided Enhancement in VR-based Multi-object Teleoperation - University of Helsinki Research Portal

https://researchportal.helsinki.fi/en/publications/mirage-multimodal-intention-recognition-and-admittance-guided-enh/

[42] SpatiaLLM: A Multi-Agent Framework for Adaptive and User-Aware Mixed Reality Content Placement

https://doi.org/10.1109/VRW70859.2026.00278

[43] [59] AgentHands: Generating Interactive Hands Gestures for Spatially Grounded Agent Conversations in XR citation

https://ruofeidu.com/cites/Liu2026AgentHands.html

[45] Powering Clinical Communication Training with an LLM-Powered Virtual Reality Simulation | IEEE Conference Publication | IEEE Xplore

https://ieeexplore.ieee.org/abstract/document/11489723

[46] Conversational AI for Safety-Critical Virtual and Extended Reality: Intelligent Agents Across Health Informatics, Navigation, and Urban Sensing | IEEE Conference Publication | IEEE Xplore

https://ieeexplore.ieee.org/abstract/document/11662436/

[47] XARA: A Multimodal Agentic System for eXtended Reality Experiment Generation

https://arxiv.org/abs/2610.21067

[48] [66] Khronos OpenXR Registry - The Khronos Group Inc

https://registry.khronos.org/OpenXR/

[53] Survey of Large Language Models in Extended Reality: Technical Paradigms and Application Frontiers

https://arxiv.org/abs/2508.03014

[56] Toward User-Aware Interactive Virtual Agents: Generative Multi-Modal Agent Behaviors in VR | IEEE Conference Publication | IEEE Xplore

https://ieeexplore.ieee.org/abstract/document/10765477

[58] GitHub - brodyl/WhenToCallTheLLM: An artifact repo for the paper published in VRST 2025 titled - When to Call the LLM: Adaptive Speech Interfaces for Natural Language Input in eXtended Reality Environments · GitHub

https://github.com/brodyl/WhenToCallTheLLM

[61] A robust layered control system for a mobile robot

https://doi.org/10.1109%2FJRA.1986.1087032

[64] SpatialLM: Training Large Language Models for Structured Indoor Modeling

https://proceedings.neurips.cc/paper_files/paper/2025/hash/40223e204a9e18dd8b01d7d11ea97939-Abstract-Conference.html

[65] CVPR 2025 Open Access Repository

https://openaccess.thecvf.com/content/CVPR2025/html/Zhi_LSceneLLM_Enhancing_Large_3D_Scene_Understanding_Using_Adaptive_Visual_Preferences_CVPR_2025_paper.html

[67] EUR-Lex - 02024R1689-20260727 - EN - EUR-Lex

https://eur-lex.europa.eu/eli/reg/2024/1689