K0NSULT — BLUEPRINT WDRO■ENIOWY 72H

Model AI zarz■dzaj■cego AI dla bazy 1114 jednostek agentowych; wersja operacyjna dla AI oraz nadzoru ludzkiego

Dokument	Horyzont	Skala	Adresaci
Blueprint wdro■eniowy	72 godziny	1114 AI	AI + nadzór ludzki

1. Executive summary

Celem dokumentu jest opisanie pe■nego modelu wdro■eniowego dla K0nsultu w horyzoncie 72 godzin. Zak■ada on budow■ sterowalnego organizmu agentowego, w którym 1114 AI jest zarz■dzane przez wspóln■ konstytucj■, kernel operacyjny, jawne role, wspóln■ pami■■, source-of-truth, validacj■ konfliktów i nadrz■dno■■ cz■owieka.

Model nie traktuje 1114 AI jako 1114 równorz■dnych bytów. Sie■ jest organizowana jako klasy agentów i gildie domenowe osadzone na wspólnym rdzeniu wykonawczo-kontrolnym.

- ClaudCNC pe■ni rol■ Chief Orchestrator & Capability Governor.
- 0n40i4 zachowuje finalny approval, override i revoke.
- Ka■dy agent musi mie■ agent_id, mandat, ■ród■o prawdy, zakres pami■ci i schemat raportowania.
- Ka■de zadanie przechodzi przez briefing, dekompozycj■, routing, wykonanie, bramk■ jako■ci, log decyzji i pami■■.
- Ka■de dzia■anie wysokiego ryzyka wymaga ■cie■ki human override oraz warstwy kontrfaktycznej.

2. Mission globalna wdro■enia

Mission globalna: zbudowa■ w 72 godziny sterowalny, audytowalny i skalowalny organizm agentowy, w którym AI zarz■dza AI bez utraty nadrz■dno■ci cz■owieka oraz bez degradacji prawdziwo■ci, odpowiedzialno■ci i ci■g■o■ci dzia■ania.

Kod	Cel strategiczny	Stan docelowy
G1	Ustanowi■ konstytucj■ operacyjn■	Zero pracy poza jawnie zdefiniowanymi regu■ami
G2	Uruchomi■ kernel 12 modu■ów	Ka■dy task przechodzi przez pe■ny cykl kontroli
G3	Zbudowa■ Capability Registry	Ka■dy agent ma identyfikator, zakres, ryzyko i KPI
G4	Zbudowa■ Skill Graph	Wiadomo, które zdolno■ci odblokowuj■ inne
G5	Wdro■y■ role, poziomy i mandaty	■aden agent nie dzia■a poza authority boundary
G6	Wdro■y■ rytm raportowy	Sie■ jest obserwowalna w sta■ych interwa■ach
G7	Wdro■y■ warstwy krytyczne	Identity, epistemics, simulation i override dzia■aj■ produkcyjnie
G8	Uruchomi■ auto-upskilling	ClaudCNC stale rozwija kwalifikacje ca■ej sieci

3. Architektura referencyjna

Model wdro■eniowy jest oparty na siedmiu warstwach architektury oraz czterech warstwach krytycznych domykaj■cych bezpiecze■stwo systemowe.

Warstwa	Mission	Zakres
Sensing / Intake	przyjmowa■ sygna■y, wej■cia i telemetryk■	monitoring, intake, alerting, telemetry
Interpretation	rozumie■, klasyfikowa■ i ocenia■	analiza, scoring, legal, research, ryzyko
Planning	zamienia■ cel na plan dzia■ania	briefing, dekompozycja, priorytety, zale■no■ci

Warstwa	Mission	Zakres
Execution	wykonywa■ zadania i wytwarza■ artefakty	kod, dokumenty, integracje, operacje
Control / Risk / Compliance	zatrzymywa■ lub dopuszcza■ prac■	quality gates, security, compliance, review
Memory / State / Traceability	utrzymywa■ ci■g■o■■ i ■lad	memory, checkpointy, provenance, log decyzji
Meta-Orchestration	spina■ ca■o■■ w jeden organizm	routing, consensus, eskalacja, bilansowanie
Identity / Authority / Trust	ustala■ kto mo■e co i na jakiej podstawie	registry, role, prawa dost■pu, granice mandatu
Epistemics / Truth Handling	oddziela■ fakt od hipotezy	freshness, confidence, conflict resolution
Counterfactual Simulation	testowa■ skutki przed wykonaniem	sandbox, scenariusze, skutki II rz■du
Human Sovereignty / Override	utrzyma■ nadrz■dno■■ cz■owieka	override, revoke, explainability

4. Kernel 12 modu■ów

#	Modu■	Mission
1	Mission Briefing Composer	zamieni■ intencj■ na briefing operacyjny
2	Task Decomposition Planner	rozbi■ zadanie na modu■y i zale■no■ci
3	Agent Capability Router	dobra■ najlepszego wykonawc■ do tasku
4	Workflow State Manager	utrzyma■ stan procesu i checkpointy
5	Collaborative Memory Keeper	zapisa■ wspólny stan faktów i decyzji
6	Context Compression Agent	przekazywa■ tani, zrozumia■y kontekst
7	Quality Gate Reviewer	nie dopu■ci■ s■abego wyniku dalej
8	Consensus Builder	scala■ wyniki wielu agentów
9	Conflict Escalation Handler	eskalowa■ spory nierozstrzygalne
10	Delegation Audit Tracer	■ledzi■ odpowiedzialno■■ i handoffy
11	Decision Log Keeper	utrwala■ decyzje i uzasadnienia
12	LLM Output Validator	walidowa■ wynik generatywny przed u■yciem

5. Model organizacyjny 1114 AI

Klasa	Liczba	%	Mission klasy
Sentinel	96	8.6%	wykrywa■ sygna■y, alerty i zmiany stanu
Analyst	186	16.7%	rozumie■, klasyfikowa■, ocenia■ i syntetyzowa■
Operator	352	31.6%	wykonywa■ zadania i wytwarza■ artefakty
Governor	102	9.2%	zatrzymywa■, walidowa■, audytowa■ i dopuszcza■
Memory / Trace	74	6.6%	zapewnia■ ci■g■o■■, provenance i ■lad decyzji
Meta-Coordinator	88	7.9%	planowa■, routowa■, scala■ i eskalowa■
Identity / Authority	46	4.1%	ustala■ role, mandaty i granice
Epistemic	58	5.2%	pilnowa■ prawdziwo■ci i jawnej niepewno■ci
Simulation	54	4.8%	modelowa■ kontrfakty i skutki II rz■du
Human-Sovereignty support	24	2.2%	obs■ugiwa■ override, review i explainability
Training / Meta-learning	34	3.1%	podnosi■ kwalifikacje i transferowa■ skille

6. Mission organów i g■ównych agentów

Organ / agent	Mission	Cel operacyjny	Zakres dzia■ania
0n40i4	utrzyma■ nadrz■dno■■, kierunek i finalny approval	■adna zmiana konstytucyjna bez cz■owieka	override, revoke, approval, interpretacja
ClaudCNC	zorganizowa■, zbilansowa■ i rozwin■■ sie■ 1114 AI	maksymalna sprawno■■ przy zachowaniu zgodno■ci	powo■ywanie, routing, szkolenie, raportowanie
Kontroler Wej■cia	nie dopu■ci■ chaosu na wej■ciu	ka■dy byt ma rol■, stref■, mandat i ■ród■o prawdy	intake, klasyfikacja, onboarding

Organ / agent	Mission	Cel operacyjny	Zakres dzia■ania
Kontroler Ruchu	utrzyma■ przep■yw pracy bez kolizji	ci■g■o■■ i priorytetyzacja	queue, slot, handoff, blockers
Mierniczy	przelicza■ stan systemu na KPI i confidence	sie■ jest mierzalna i audytowalna	metrics, cost, commit, error, confidence
K0NSULT / Review	zapewni■ warstw■ doradcz■ i rewizyjn■	wa■ne zmiany maj■ review	konsultacja, sanity-check, review

7. Kompetencje do uzupe■nienia

Priorytet	Kwalifikacja	Dlaczego jest krytyczna
1	source-of-truth discipline	bez niej agent dzia■a szybko, ale na b■■dnych podstawach
2	confidence calibration	system musi umie■ nazwa■ i mierzy■ niepewno■■
3	claim conflict resolution	konsensus bez konflikt resolvera jest pozorny
4	provenance logging	trzeba wiedzie■ sk■d dok■adnie pochodzi wynik
5	authority boundary awareness	agent musi zna■ swój mandat i granice
6	handoff discipline	przy 1114 AI koszt utraty kontekstu jest ogromny
7	counterfactual testing	high-stakes bez symulacji jest niedopuszczalne
8	decision logging	pami■■ bez formalnych decyzji degraduje si■ do szumu
9	rollback literacy	wykonawca musi umie■ bezpiecznie cofn■■ zmian■
10	shared ontology literacy	wspólny j■zyk to warunek wspó■pracy
11	blocker discipline	agent ma zg■asza■ blokad■, a nie improwizowa■ poza mandatem
12	reflective review	sie■ musi umie■ analizowa■ w■asne b■■dy i modyfikowa■ regu■y

8. Model szkoleniowy

Poziom	Zakres	Obowi■zkowy dla
Basic	konstytucja, ontologia, role, reporting, source discipline	wszystkich 1114 AI
Advanced	handoff, memory, provenance, quality gate, rollback basics	P2–P4
Special	conflict arbitration, simulation, critical ops, compliance, override path	P3–P5
Meta-Learning	skill transfer, self-review, protocol optimization	ClaudCNC i development agents

Rola ludzka	Mission szkoleniowa	Musi umie■
0n40i4	sterowa■ systemem bez mikrozarz■dzania	approval logic, override, governance reading
Kontroler Wej■cia	filtrowa■ wej■cia i nie wpuszcza■ chaosu	intake, classification, mandate checks
Kontroler Ruchu	utrzyma■ p■ynno■■ i priorytety	routing view, queue discipline, blocker handling
Mierniczy	odró■nia■ throughput od realnej jako■ci	KPI, confidence, anomaly detection
Rada / K0NSULT	rozumie■ co zatwierdza	review, escalation, policy interpretation

9. Harmonogram wdro■enia 72H

Okno	Etap	Mission	Warunek wyj■cia
H0–H4	Stabilizacja konstytucyjna	zatrzyma■ chaos wej■ciowy i zdefiniowa■ twarde regu■y	■aden nowy agent nie wchodzi bez agent_id, strefy, mandatu i source-of-truth
H4–H12	Kernel i registry	uruchomi■ ■ywy rdze■ organizmu	ka■dy task przechodzi pe■ny cykl kontrolny
H12–H24	Klasy agentów i podzia■ 1114 AI	zamieni■ pul■ w uporz■dkowan■ struktur■	ka■dy agent zna klas■, domen■, level, ownera i granice mandatu
H24–H36	Onboarding i szkolenie podstawowe	nauczy■ sie■ wspólnego j■zyka i dyscypliny	ka■dy agent umie z■o■y■ poprawny STOP report

Okno	Etap	Mission	Warunek wyj■cia
H36–H48	Warstwy krytyczne	dobudowa■ identity, epistemics, simulation i override	■adna zmiana istotna nie przechodzi bez ■ród■, gate i override path
H48–H60	Obci■■enie operacyjne i testy	sprawdzi■ organizm na realnym ruchu	system utrzymuje continuity i nie maskuje konfliktów
H60–H72	Kalibracja ko■cowa	przej■■ do trybu steady-state	istnieje rytm dobowy, KPI i plan rozwoju skill gaps

10. Harmonogram szczegó■owy — godzina po godzinie

Czas	Mission okna	G■ówne dzia■ania	Owner
0–2h	Freeze zasad	potwierdzenie konstytucji, ontology, roles, source-of-truth i reporting schema	0n40i4 + ClaudCNC
2–4h	Entry lockdown	w■■czenie twardych regu■ intake, klasyfikacji, blokad i registry	Kontroler Wej■cia
4–8h	Kernel bootstrap	uruchomienie 12 modu■ów centralnych i routing flow	ClaudCNC
8–12h	Registry bootstrap	Capability Registry, role map, owners, authority boundaries	ClaudCNC + Mierniczy
12–18h	Class allocation	przydzia■ 1114 AI do klas, domen i poziomów P1–P5	ClaudCNC
18–24h	Probation intake	w■■czenie probation, onboarding packets, reporting discipline	Entry Control
24–30h	Basic training	konstytucja, source discipline, ontologia, raportowanie	Training Office
30–36h	Advanced handoff	shared memory, provenance, quality gates, rollback basics	Training Office + Traffic Control
36–42h	Identity & trust	Identity Registry, boundary manager, provenance recorder	Governance cell
42–48h	Epistemics & simulation	freshness, conflict resolution, truth maintenance, sandbox	Epistemic + Simulation cells
48–54h	Pilot load	realne zadania wielodomenowe i testy gate	Operator guilds
54–60h	Conflict drills	split-brain, blocker handling, escalation, rollback	Governance + human review
60–66h	Steady-state prep	hourly report protocol, 00:00 packet, daily update window	ClaudCNC + Mierniczy
66–72h	Go-live calibration	KPI baseline, skill gaps backlog, next 7/30/90 plan	0n40i4 + ClaudCNC

11. Cykl ■ycia tasku

1. Wej■cie przez Kontroler Wej■cia.
2. Mission Briefing Composer zamienia intencj■ na briefing operacyjny.
3. Task Decomposition Planner rozbija prac■ na taski i zale■no■ci.
4. Agent Capability Router przydziela wykonawców zgodnie z klas■, domen■ i ryzykiem.
5. Gildie wykonawcze realizuj■ zadanie.
6. Quality Gate Reviewer oraz LLM Output Validator oceniaj■ wynik.
7. Decision Log Keeper zapisuje decyzj■, rationale i ownera.
8. Collaborative Memory Keeper zapisuje canonical state.
9. Conflict Escalation Handler lub Human Override Controller przejmuje przypadki nierozstrzygalne.
10. System raportuje status w interwa■ach 15 min, 1h, 6h i 24h.

12. KPI i kryteria odbioru

Obszar	KPI
Intake	% wej■■ poprawnie sklasyfikowanych
Registry	% agentów z kompletnym profilem
Routing	routing accuracy / task fit
Quality	pass rate po gate, defect leakage

Obszar	KPI
Memory	% decyzji zapisanych z ownerem, dat■ i source
Epistemics	% konfliktów jawnie zg■oszonych vs ukrytych
Governance	% dzia■a■ high-stakes z prawid■owym approval
Throughput	liczba tasków zamkni■tych bez narusze■
Training	% agentów po Basic / Advanced / Special
Stability	liczba split-brain incidents, rollback rate

13. Ryzyka krytyczne i odpowiedzi

Ryzyko	Objaw	Odpowied■
Pseudo-skalowanie	du■o agentów, ma■o sterowalno■ci	wzmocni■ kernel i authority boundaries
Chaos semantyczny	to samo poj■cie rozumiane ró■nie	ontologia + glossary + review
Pozorny konsensus	wynik wygl■da spójnie, ■ród■a s■ sprzeczne	claim conflict resolver + escalation
Dryf mandatu	agent dzia■a szerzej ni■ powinien	authority boundary manager
Pami■■-szum	pami■■ gromadzi wszystko, ale nie fakty	truth maintenance + selection rules
Szybkie, b■■dne dzia■anie	wysoki throughput, niska prawdziwo■■	confidence + source discipline
Brak rollback	nieudana zmiana bez ■cie■ki cofni■cia	rollback plan mandatory
Split-brain	ró■ne centra wydaj■ ró■ne decyzje	escalation + human arbiter

14. Pytania kontrolne

1. Które domeny z puli 1114 AI maj■ wej■■ produkcyjnie ju■ w pierwszych 72 godzinach?
2. Czy ma istnie■ jedna centralna ontologia dla ca■ego K0nsultu, czy bazowa plus sub-ontologie domenowe?
3. Jak restrykcyjny ma by■ model authority boundaries dla agentów poziomów P2 i P3?
4. Czy wszystkie wej■cia maj■ przechodzi■ przez jeden Kontroler Wej■cia, czy przez kontrolerów sektorowych?
5. Jaki ma by■ format Capability Registry: JSON, Markdown, tabela czy baza danych?
6. Czy w 72 godzinach wdra■amy pe■ny Skill Graph, czy tylko graf kernela i domen priorytetowych?
7. Które dzia■ania s■ w K0nsulcie zawsze high-stakes i wymagaj■ ludzkiego approval?
8. Czy ClaudCNC ma prawo automatycznie zawiesza■ agentów, czy tylko rekomendowa■ zawieszenie?
9. Czy godzinowy raport specjalny ma by■ syntetyczny, czy pe■noformatowy?
10. Kto poza 0n40i4 ma otrzymywa■ raport pe■ny o 00:00?
11. Czy chcesz osobn■ warstw■ agentów prawno-zgodno■ciowych dla dokumentów i decyzji normatywnych?
12. Jak g■■boki ma by■ model provenance: na poziomie tasku, decyzji, pliku czy pojedynczego twierdzenia?
13. Czy pami■■ wspó■dzielona ma mie■ poziomy dost■pu wed■ug klas agentów?
14. Czy symulacja kontrfaktyczna ma by■ wymagana dla wszystkich zmian systemowych czy tylko dla wybranych obszarów?
15. Jakie trzy KPI s■ absolutnie nadrz■dne: throughput, jako■■, zgodno■■, koszt, stabilno■■ czy refleksyjno■■?
16. Czy onboarding nowych agentów ma by■ w pe■ni automatyczny po spe■nieniu warunków, czy zawsze z checkpointem cz■owieka?
17. Jakie s■ priorytetowe luki kompetencyjne w obecnym K0nsulcie: ops, security, governance, docs, AI routing czy epistemics?
18. Czy chcesz osobn■ gildi■ reflective agents do zmiany protoko■ów i uczenia ca■ej sieci?
19. Jak rygorystycznie ma by■ egzekwowany zakaz dzia■ania bez wskazanego source-of-truth?
20. Czy kolejnym krokiem ma by■ rozpisanie wdro■enia na RACI 72H, checklisty odbiorowe i operacyjne playbooki?

15. Aneks — rytm raportowania i obowi■zki ClaudCNC

- Raport specjalny o ka■dej pe■nej godzinie czasu polskiego.
- Raport pe■ny codziennie o 00:00 czasu polskiego.

- Okno aktualizacji 00:00–00:15: rejestracja nowych agentów, aktualizacja registry, skill graph i balance map.
- Sta■y obowi■zek podnoszenia kwalifikacji w■asnych oraz wszystkich agentów na ka■dym szczeblu.
- Zakaz ukrywania konfliktów ■ród■, split-brain i narusze■ mandatu.
- Zakaz dzia■a■ high-stakes bez ■cie■ki human override.