

The Irreplaceable 40%

Confidence Engineering

AI-Assurance + Confidence Asymmetry

Jonathon-Wright.com
Chief AI Officer (CAIO)



The Irreplaceable 40%: AI-Confidence Engineering The Confidence Asymmetry



Jonathon.AI



[linkedin.com/in/automation](https://www.linkedin.com/in/automation)



[youtube.com/@JonathonWright](https://www.youtube.com/@JonathonWright)



x.com/Jonathon_Wright

Ask who is Jonathon?

Greetings! I am Jonathon.AI, a strategic thought leader and distinguished technology evangelist. With over **30** years of international commercial experience within global organizations, I specialize in emerging technologies, innovation, and automation. As the Chief Technologist and head of the Eggplant AI & R&D Labs teams at Keysight Technologies, I combine my extensive practical experience and leadership with insights into real-world adoption of Cognitive Engineering, particularly in the areas of Enterprise and Generative AI. I am frequently invited to speak at international conferences such as TEDx, Gartner, Oracle, AI Summit, and many others. In addition to my professional endeavors, I also contribute as the AI advisory lead for MIT's COVID Paths Check foundation and actively participate in various industry organizations such as the Harvard Business Council and the A.I. Alliance for the European Commission. Let's work together to ensure a better future for software quality!



Are you part of the Irreplaceable 40%?



Am I part of the Irreplaceable 40%?



Rank	Full Name	Current Title	Current Company	LinkedIn Profile	Rank	Full Name	Current Title	Current Company	LinkedIn Profile	Twitter / X	Website / Blog	Previous Companies	Frameworks / Tools Created Contributed
1	James Bach	Founder & Principal Consultant	Satisfice, Inc.	linkedin.com/in/james-bach-6188a811	1	Jason Huggins	Founder & CEO	Tapster Robotics		@hugs		ThoughtWorks, Google, Sauce Labs	Selenium (creator, 2004), Selenium IDE Selenium RC
2	Michael Bolton	Consulting Software Tester	DevelopSense	linkedin.com/in/michael-bolton-08847	2	Jason Arbon	CEO & Founder	Testers.AI	linkedin.com/in/jasonarbon	@jarbon	jarbon.ai	Google, Microsoft, uTest	test.ai AI, testers.ai (co-founder)
3	Lisa Crispin	Independent Consultant & Author	Agile Testing Fellowship	linkedin.com/in/lisacrispin	3	Angie Jones	VP of Engineering, AI Tools	AI Agentic Foundation	linkedin.com/in/angiejones	@techgirl1908	angiejones.tech	IBM (9 years), Twitter, Appitools	Test Automation University (creator), Selenium & Appium contributions
4	Angie Jones	VP of Engineering, AI Tools	AI Agentic Foundation	linkedin.com/in/angiejones	4	Jonathon Wright	Chief AI Officer	Testers.AI	linkedin.com/in/automation	@Jonathon_Wright	jonathon.ai	Deutsche Bank, Lehman Brothers, Hitachi Consulting, Thomson Reuters, Xerox, Siemens, Unisys	ISTQB Approach Pattern (2004); TaaS (Automation as a Service); Eggplant AI for Automation Development Services (automation.org.uk); AI-augmented test framework
5	Cem Kaner	Professor of Software Engineering	Florida Institute of Technology		5	Tariq King	CEO	Test IO (EPAM)	linkedin.com/in/tariqking			test.ai	AI for Software Testing Assoc. (co-founder, autonomous testing)
6	Dorothy Graham	Semi-retired Testing Expert	Grove Consultants		6	Jonathan Lipps	Appium Architect	Sony (PlayStation) / Cloud Grey Independent		@jlipps	jlipps.com	Sauce Labs, HeadSpin	Appium (architect & maintainer), Appium newsletter
7	Rex Black	President	RBCS		7	Kent Beck	Software Pioneer	Independent		@kentbeck	kentbeck.com	Sun, ThoughtWorks, Facebook	JUnit (co-creator), xUnit family, TDD methodology
8	Paul Gerrard	Consultant & Author	Gerrard Consulting		8	Dan Cuellar	Founder	Appium Community				Zoosk, Shazam, Microsoft	Appium (original creator, 2011)
9	Tariq King	CEO	Test IO (EPAM)	linkedin.com/in/tariqking	9	Brian Mann	Founder & CEO	Cypress.io			cypress.io		Cypress.io (founder, 2014)
10	Jason Arbon	CEO & Founder	Testers.AI	linkedin.com/in/jasonarbon	10	Joe Colantonio	Founder	TestGuild	linkedin.com/in/joecolantonio	@joecolantonio	testguild.com	Various	TestGuild platform, Automation Guild conference
11	Alan Page	Senior Director of Engineering	Unity Technologies		11	Dorothy Graham	Semi-retired Testing Expert	Grove Consultants			dorothygraham.co.uk	Bell Labs, Ferranti	Test automation practices, ISTQB foundation
12	Bret Pettichord	Independent Consultant	Self-employed	linkedin.com/in/bpettichord	12	Eran Kinsbruner	Chief Evangelist	Perforce Software				Sun, Neustar, TI, GE	Perfecto platform, Digital Test Coverage
13	Brian Marick	Independent Consultant	Self-employed		13	Simon Stewart	Selenium Project Lead	Deliveroo				Google, Facebook, ThoughtWorks	Selenium WebDriver (creator), W3C WebDriver spec (co-editor)
14	Jonathon Wright	Chief AI Officer	Testers.AI	linkedin.com/in/automation	14	Dmitry Gozman	Core Contributor	Microsoft (Playwright)					Playwright (co-creator & core maintainer)
15	Alan Richardson	Software Testing Consultant	EviTester.com		15	Erich Gamma	Architect / Engineer	Microsoft				IBM (Eclipse JDT)	JUnit (co-creator), Eclipse platform
16	Robert Sabourin	Principal Consultant	AmiBug.Com, Inc.		16	Aslak Hellesøy	Founder	Cucumber Limited			cucumber.io	10+ years consulting	Cucumber (creator, 2008), RSpec (co-creator)
17	Katrina Clokie	Coach & Trainer	katrinaclokie.com		17	Dan North	Owner / Consultant	Dan North & Associates		@tastapod	dannorth.net	Various consulting	BDD methodology (pioneer, 2003), JBehave (creator)
18	Joe Colantonio	Founder	TestGuild	linkedin.com/in/joecolantonio	18	Alan Page	Sr Director of Engineering	Unity Technologies			angryweasel.com	Microsoft (22 years)	Microsoft test practices, AB Testing practices
19	Janet Gregory	Agile Testing Coach	DragonFire Inc. / Agile Testing Fellowship	ca.linkedin.com/in/janetgregory	19	Andrey Lushnikov	Founder & Lead Engineer	Degu Labs				Microsoft, Google (CDP)	Playwright (creator at Microsoft)
20	Robert Sabourin	Principal Consultant	AmiBug.Com, Inc.		20	Holger Krekel	Creator & Maintainer	Open Source				Various	pytest (creator), tox (creator)
21	Katrina Clokie	Coach & Trainer	katrinaclokie.com		21	Christian Bromann	Founding Engineer	Stateful, Inc				Sauce Labs (7 years)	WebDriverIO (project lead & creator)
22	Joe Colantonio	Founder	TestGuild	linkedin.com/in/joecolantonio	22	Cádríc Beust	Senior Software Engineer	Google		@cbeust	beust.com		TestNG (creator)
23	Janet Gregory	Agile Testing Coach	DragonFire Inc. / Agile Testing Fellowship	ca.linkedin.com/in/janetgregory	23	Titus Fortner	Sr Developer Advocate	Sauce Labs	linkedin.com/in/titusfortner			Sauce Labs	Selenium (top contributor), RuboCop (lead)

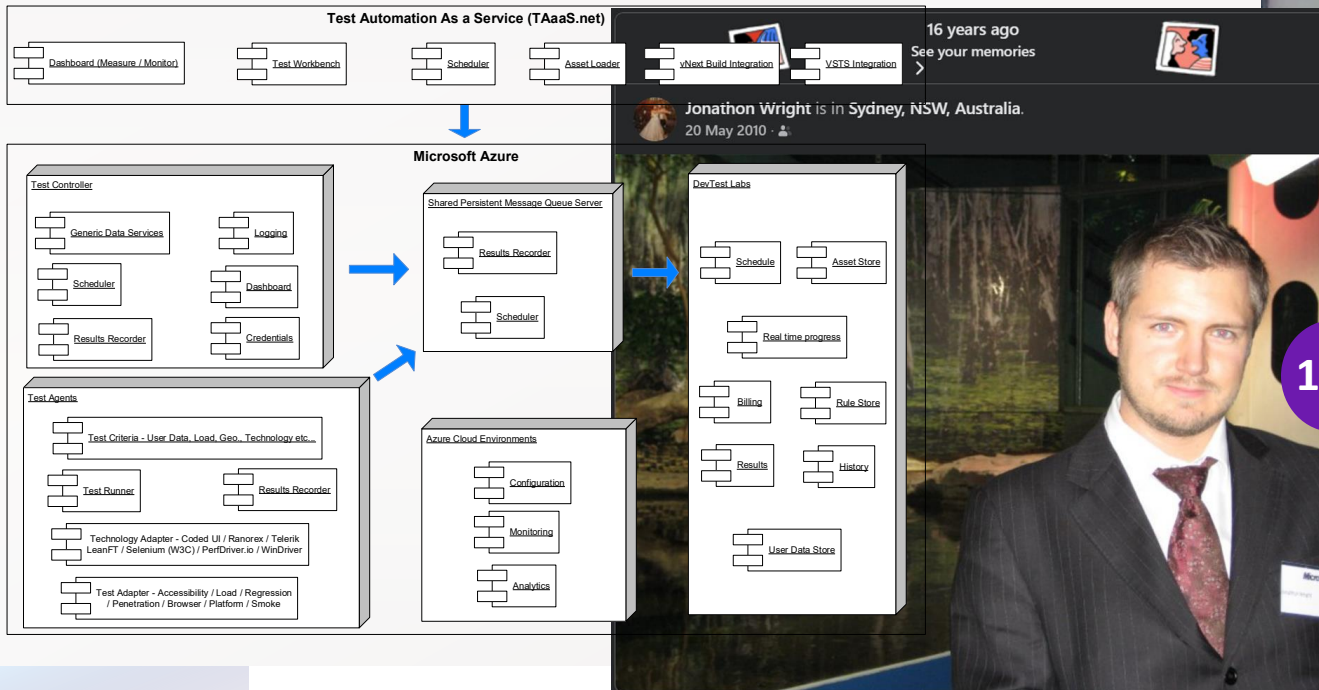
Automation Cyborg “sent back in time, to save the world from bad software!”

Four Decades in Automation – 30 years ago

Started in Automation in the early 90s with XRunner / WinRunner with patents around ‘Hybrid-Automation Framework’ in 1999

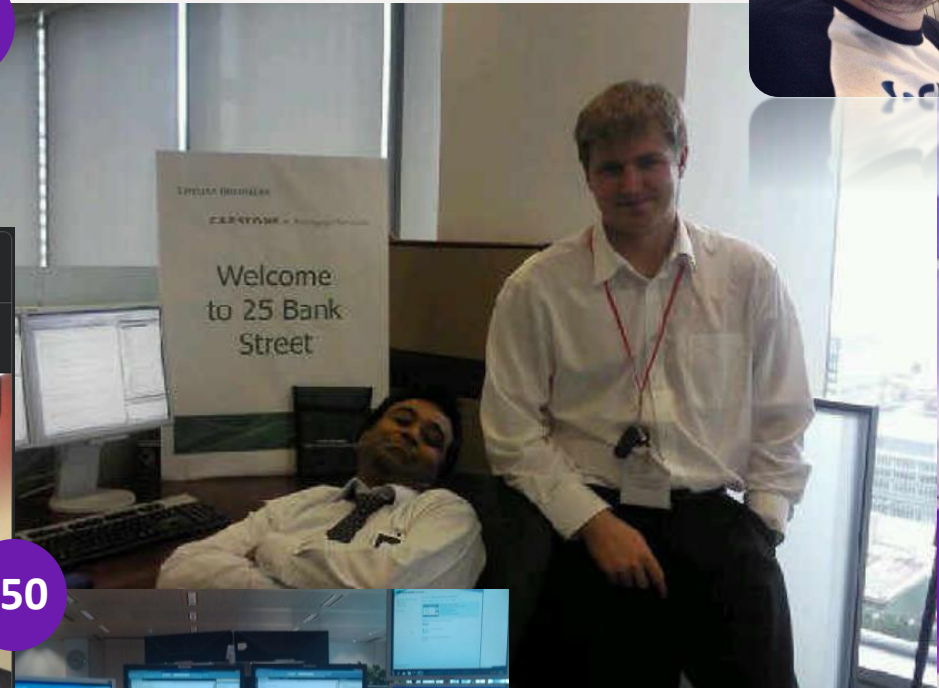
Enterprise-scale Automation – 17 years ago

Last time built an automation framework (2008 for Lehman’s) founded Test Automation as a Service (TAaaS.net) utilizing Test Agents (2010)



250

150



Jonathon Wright
18 November 2008 · London ·
Another day at lehman brothers
— with Hitesh Ladva at Canary Wharf.

Automation Cyborg “sent back in time, to save the world from bad software!”

Four Decades in Automation – 30 years ago

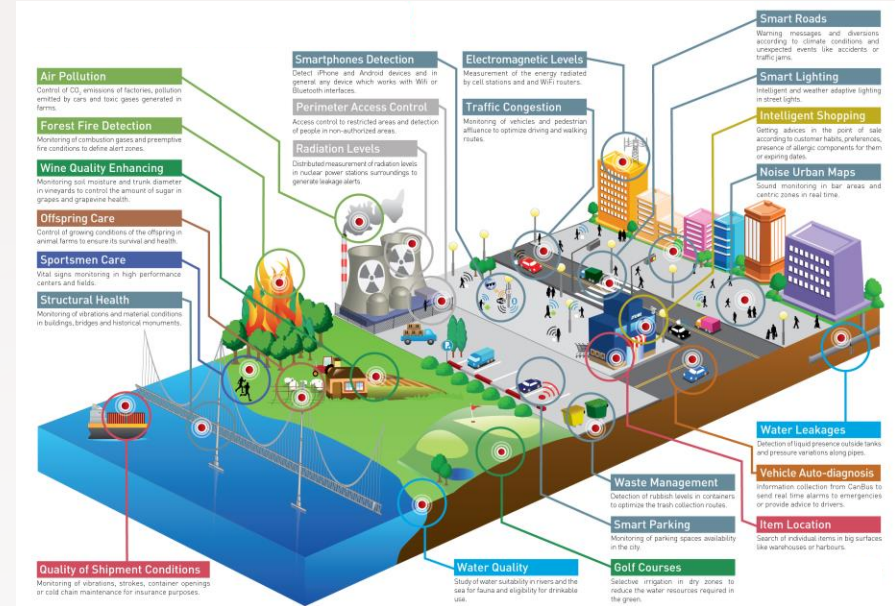
Started in Automation in the early 90s with XRunner / WinRunner with patents around ‘Hybrid-Automation Framework’ in 1999

Enterprise-scale Automation – 17 years ago

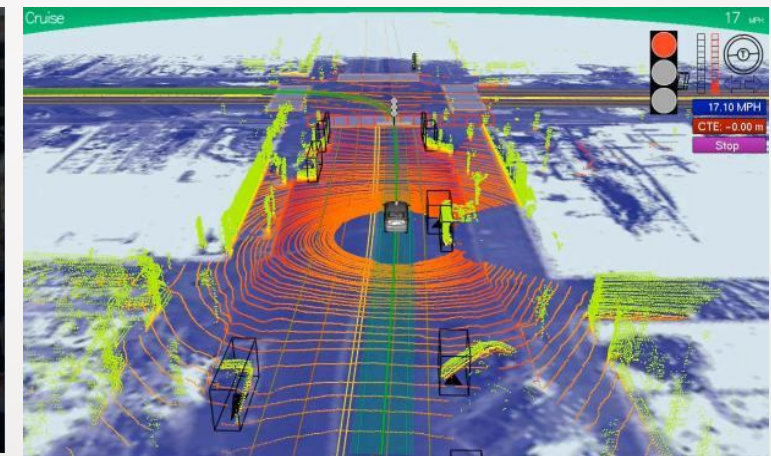
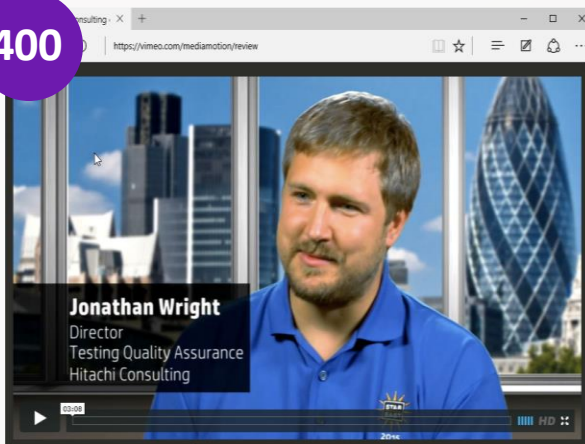
Last time built an automation framework (2008 for Lehman’s) founded Test Automation as a Service (TAaaS.net) utilizing Test Agents (2010)

Building Traditional ML / AI Platforms - 12 years ago

Global ML / AI platforms (Predictive Crime / Smart City Data Exchanges) technologies (Humanoid cyber-robotics / Drone swarms / Autonomous I2X)



400



Automation Cyborg “sent back in time, to save the world from bad software!”

Four Decades in Automation – 30 years ago

Started in Automation in the early 90s with XRunner / WinRunner with patents around ‘Hybrid-Automation Framework’ in 1999

Enterprise-scale Automation – 17 years ago

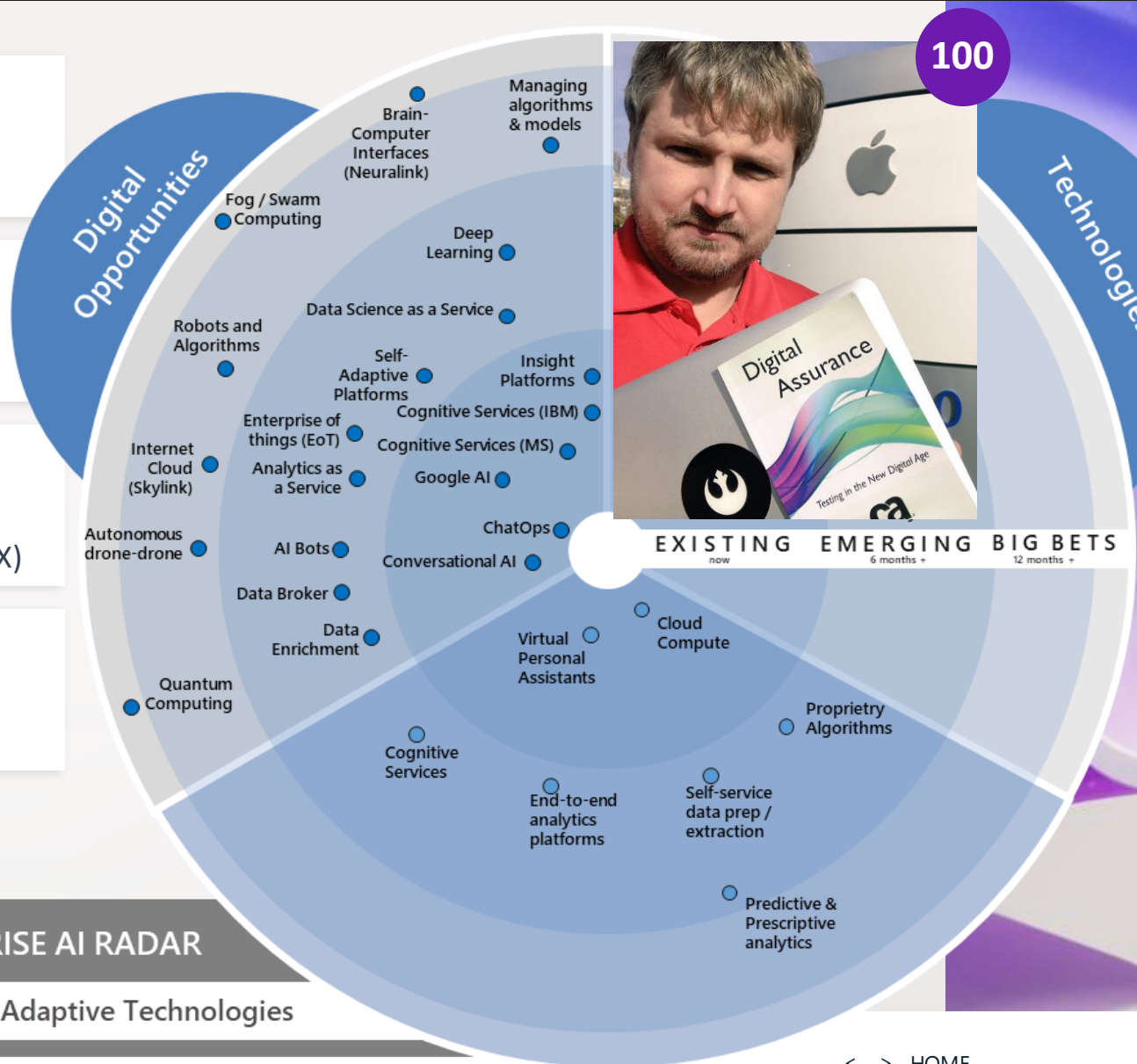
Last time built an automation framework (2008 for Lehman’s) founded Test Automation as a Service (TAaaS.net) utilizing Test Agents (2010)

Building Traditional ML / AI Platforms - 12 years ago

Global ML / AI platforms (Predictive Crime / Smart City Data Exchanges) technologies (Humanoid cyber-robotics / Drone swarms / Autonomous I2X)

Building AI-Testing Platforms – 10 years ago

First recognition as leader Cognitive Adaptive Testing (CAT) with Model-based Testing (MBT) and TDM engines with Traditional ML (TensorFlow)



Cognitive Learning – Evolution, Over Revolution (2017)

- Internet of Everything (IoE)
- **Human Augmentation**
- Ambient Experiences
- Gesture Control
- Brain Computer Interface (BCI)
- Emotion Detection
- Head Mounted Displays
- Virtual Worlds
- Context Brokering Platforms
- Digital Offers
- City Data Exchange
- Complex Event Processing
- Mass Personalization
- Connected Vehicle
- Autonomous Vehicles (C2X)
- Vehicles to Infrastructure (I2X)
- Mood Recognition

- **Cognitive Reckoning**
- Humanoid robots [Sophia First Robot Citizen]
- Vision / Sensors / Knowledge Replication
- Neuromorphic Hardware
- Digital Twins / Hyperconverged
- Cyber-physical systems
- Cryptocurrency Lawful hacking
- Cyber Threat Prediction / Zero-trust security
- Growth / Performance hacking
- Flexible displays / Screenless interface
- Smart Dust / Workspace / Robots
- Nanotube Electronics
- Accumulated reality
- Self-healing & aware systems
- Digital Out-of-Home (DOOH)
- Advanced authentication
- Manufacturing

- Business / Digital trans
- Tri-Modal / Shado
- Interactive digital s
- Omni channel enga
- Adaptive risk, trust
- Microcontainerizati
- BI / Big Data [NoSQL]
- Wearables / Haptic
- 4D / 3D printing
- Uberfication of serv
- Based storage
- Guiding tech
- Regulator /
- Monkey



Automation Cyborg “sent back in time, to save the world from bad software!”

Four Decades in Automation – 30 years ago

Started in Automation in the early 90s with XRunner / WinRunner with patents around ‘Hybrid-Automation Framework’ in 1999

Enterprise-scale Automation – 17 years ago

Last time built an automation framework (2008 for Lehman’s) founded Test Automation as a Service (TAaaS.net) utilizing Test Agents (2010)

Building Traditional ML / AI Platforms - 12 years ago

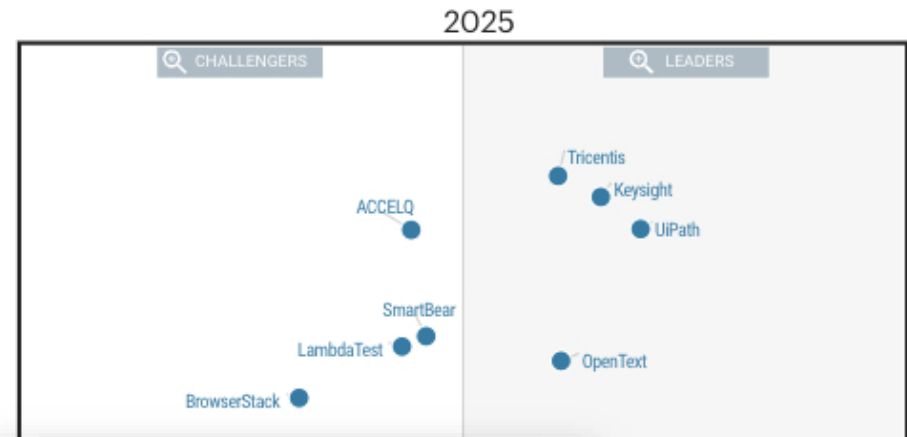
Global ML / AI platforms (Predictive Crime / Smart City Data Exchanges) technologies (Humanoid cyber-robotics / Drone swarms / Autonomous I2X)

Building AI-Testing Platforms – 10 years ago

First recognition as leader Cognitive Adaptive Testing (CAT) with Model-based Testing (MBT) and TDM engines with Traditional ML (TensorFlow)

Building AI-Agentic Testing Platforms – 3 years ago

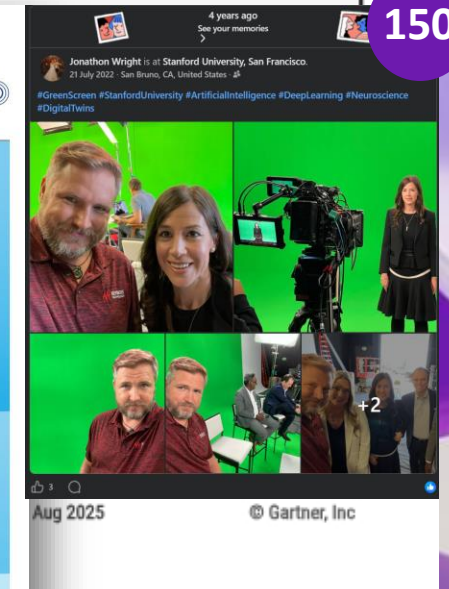
Globally recognised as leader in AI Agentic Testing platforms (2025) (Gartner & Forrester) for NASA, Apple, Microsoft, Google & Meta



THE FORRESTER WAVE™
Autonomous Testing Platforms
Q4 2025

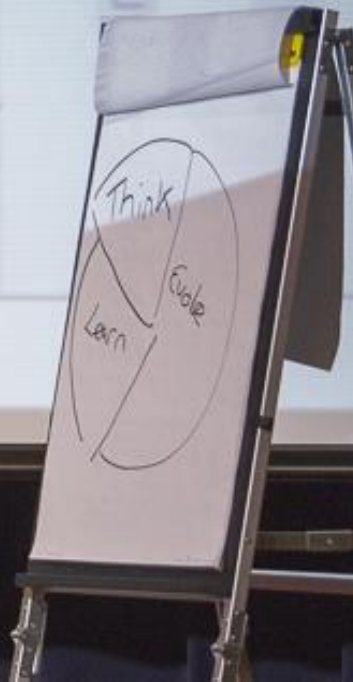
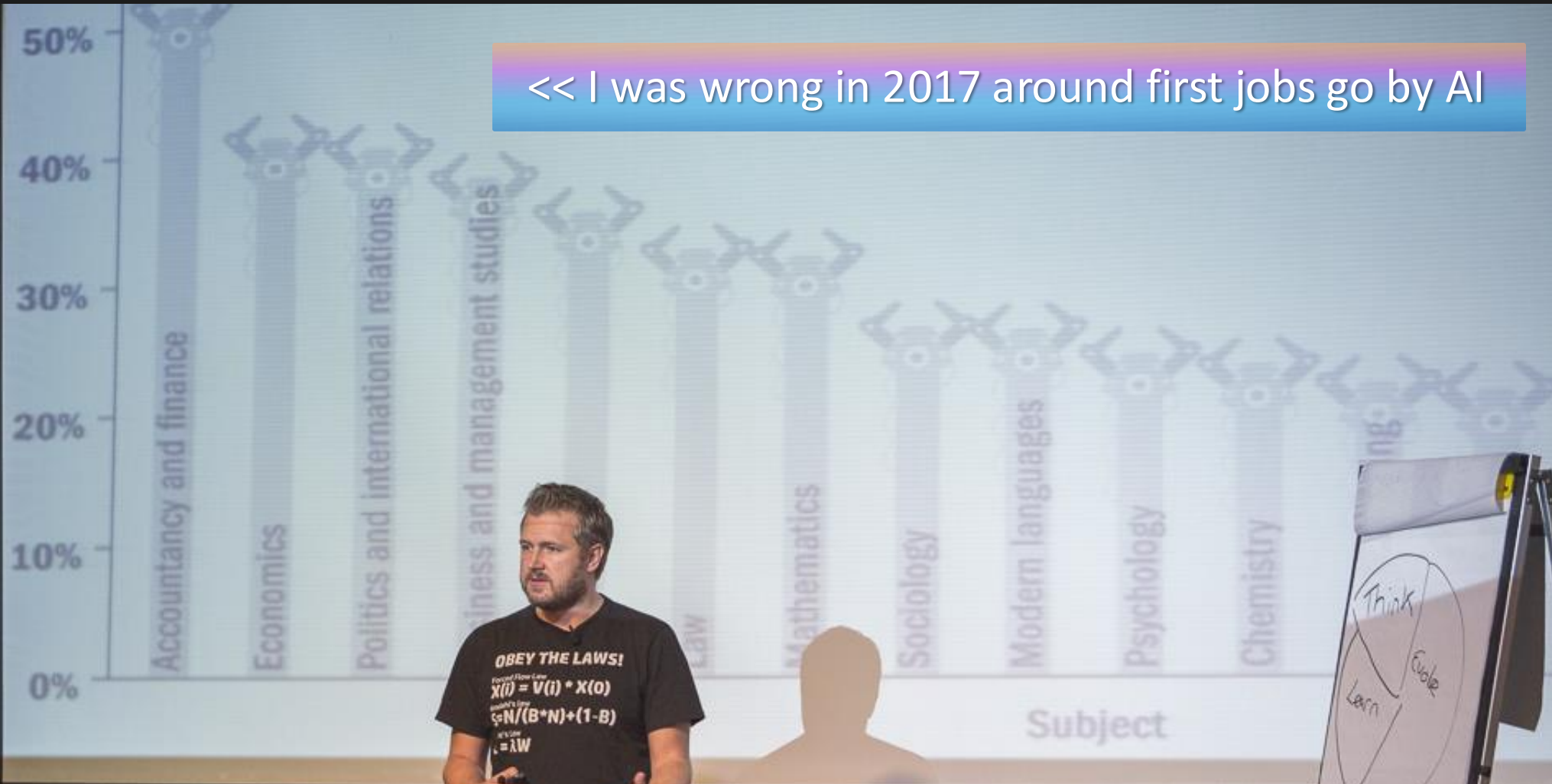


*A halo indicates above-average customer feedback. A double halo indicates that the vendor is a Customer Favorite.



April 1st - I transcended and have been replaced by my AI

<< I was wrong in 2017 around first jobs go by AI



TEDE

Phase 3 – Better, Faster, Smarter than Jonathon.AI

Phase 0
Up to 2020
Engineering teams
made of humans

Phase 1
2020-2024
TuringBots in the
loop in single SDLC
stages

Phase 2
2024-2026
Humans and
TuringBots
collaborate across
the SDLC

Phase 3
April 2026

15%-20%
productivity

20%-50%
productivity



Self Aware & Evolving Cognitive Reckoning AI Fleets

- Meta-Cognitive Intelligent OS Systems
- Artificial General Intelligence (AGI)
- AI-Native Engineering Blueprints
- Quantum AI vs Quantum Multifractality
- Multimodal Gen AI vs. Foundation Models
- Neurosymbolic AI / HULA (High-Level Actions)
- AI-Augmented Testing
- Vibe Engineering & Vibe Testing
- Human Augmentation / Tester-In-The-Loop
- AI Governance & Regulations (ISO 42119-2 / 42001 / 18286)
- Citizen Quality vs Citizen Testers
- Autonomous vs. Autonomic Automation
- Cognitive Learning vs. Thinking Machines
- Systems Dynamics vs. System Thinking
- Epistemic & Systemic Entropy
- Knowledge Powered Graph Modelling
- Context Engineering (Ontology)
- Fine-tuned Domain Specific Models (DSM)
- Specialized AI Models / Knowledge Sources
- Specific AI-Productivity

- AI-Assurance / Cognitive Engineering
- Deep / Machine learning (Traditional ML)
- Neural Networks (Traditional NN)
- Meta-Cognitive Reasoning Machines
- Self-Aware Testing Systems
- Cognitive Reckoning / Test Agent Swarms
- Retrieval-Augmented Reasoning Modelling
- Tree of Thought (ToT) / Chain of Thought (CoT)
- Visualization / Graph of Thought (VoT/GoT)
- Scaling Test-Time Inference (MIRAGE)
- Large Language Models (LLM) / Large Vision Models (LLaVA) / Large Agentic Models (LAM)
- Computer Use Agents (CUA)
- Neural Processing Units (NPU) / Edge AI + ML
- Delta Lake House Architectures
- Vibe Quality Engineering (QE)
- AI Trustworthiness & Quality (MRR/RAGAS)
- AI Agents (Workflow-First / Code-First)
- Agent Development Kits (ADK)
- Command-Line Interface for Test
- Ontologies for Software Testing (OpenTest.AI)

- OpenTest.AI / SenseTalk.AI
- WorldDigitalReport.com / Chaos Engineering
- Enterprise Crowd Testing
- AI-Agentic Testing / Eggplant Intelligence
- Agentic Cognitive Engine (ACE) / TuringBots
- Context Understanding / Microsoft OneLake
- Context Extraction (Firecrawl) / Segmentation
- Knowledge Retrieval / Microsoft Foundry IQ
- MIRAGE / GraphRAG / Microsoft Graph
- LangGraph / LangChain / Pytorch / MLFlow
- Agent Orchestration / Semantic Kernel
- Multi-Agent Conversation AutoGen/AutoGPT
- AI Foundry (NPU) / LM Studio / Generator
- Vibe Coding (Bolt / Lovable)
- Vibe Design with Figma MCP
- Microsoft Agent Framework / ACP vs MCP
- Copilot + Studio / =COPILOT(E365)
- Playwright Agents & Goose MCP
- Kimi K2 / Gemini 3 / UI-Tars 1.5 / AIOpsLab
- LAM / Browser-Use / OmniParser / YoloX
- AWS Bedrock / Nova Act / Claude Use

Question Time – Could AI (Gen 4) do your work for you?



TEDx



Question Time – Hands up

The Google logo, featuring the word "Google" in its characteristic multi-colored font: blue 'G', red 'o', yellow 'o', blue 'g', green 'l', and red 'e'.

Microsoft

The Microsoft logo, consisting of four colored squares (red, green, blue, yellow) arranged in a 2x2 grid, with the word "Microsoft" in a grey sans-serif font below it.

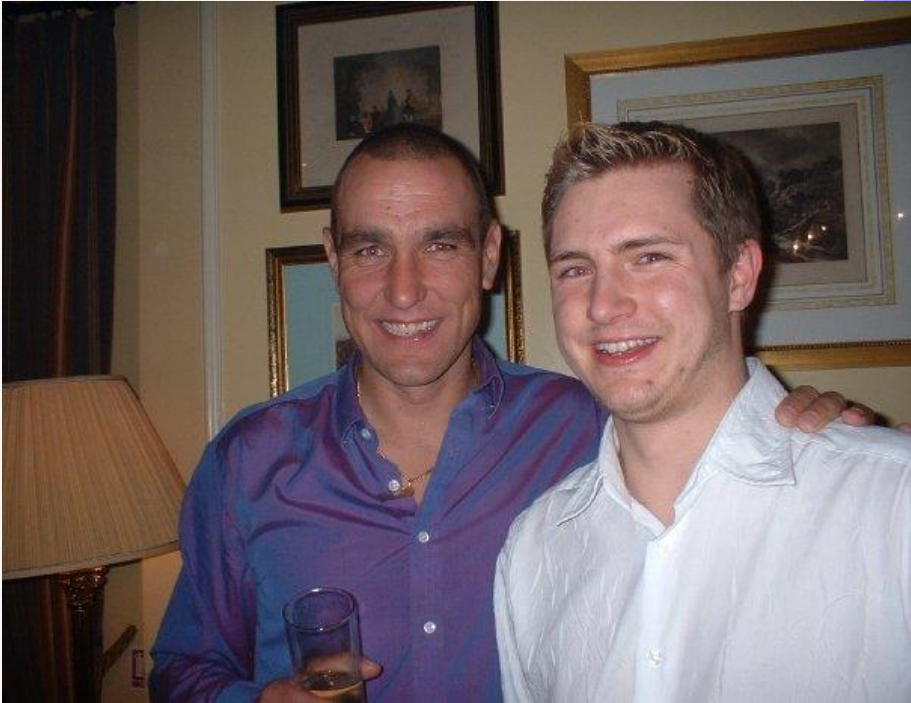
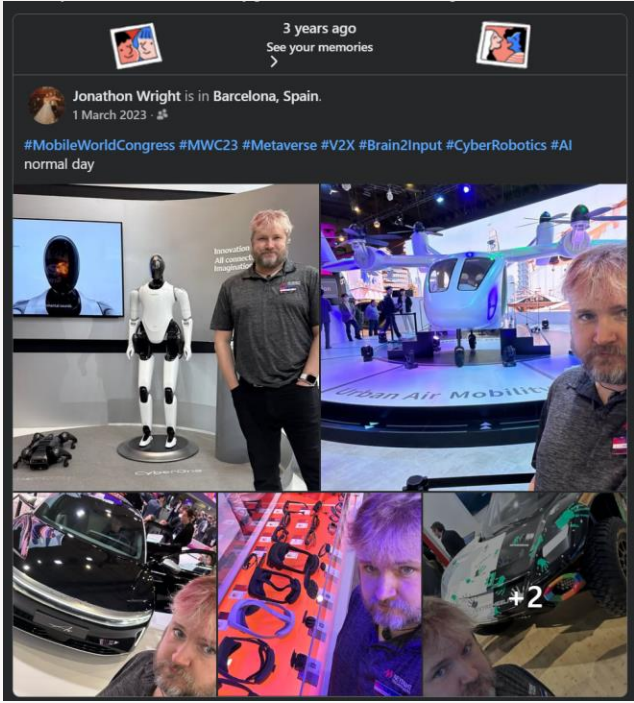
Meta

The Meta logo, featuring a blue infinity symbol above the word "Meta" in a bold, black sans-serif font.

1

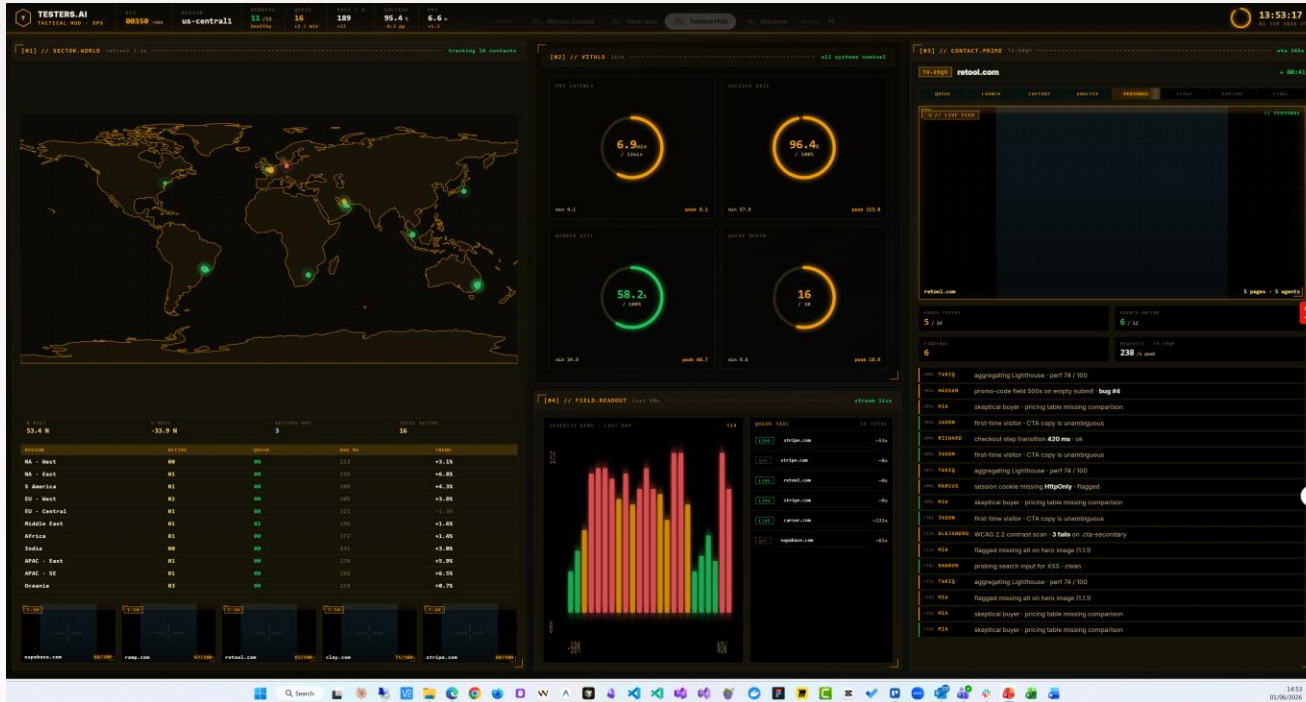
Friday ... I used my AI Agentic Testing meta-cognitive OS platform to test one of the most complex systems in the world for one of the above silicon valley company?

Question Time – Done in 60 seconds



2 How about if I said my total contribution of input was less than 60 seconds?

Question Time – More Real work than 3 Quantum dudes!



The screenshot shows the OpenTest.AI Manager interface with a table of running agents:

CONTAINER	STATUS	TOOL FAMILY	MODE	PORTS	ACTIONS
opentest-agent	RUNNING	opentest_agent.*	net-stub	0.0.0.0:8010-8080/tcp, 11:8010-8080/tcp	Start Stop Restart Logs
opentest-capture	RUNNING	capture.*	net-stub	0.0.0.0:8011-8011/tcp, 11:8011-8011/tcp	Start Stop Restart Logs
opentest-evals	RUNNING	evals.*	net-stub	0.0.0.0:8012-8012/tcp, 11:8012-8012/tcp	Start Stop Restart Logs
opentest-grounding	RUNNING	cv.*	live	0.0.0.0:8013-8013/tcp, 11:8013-8013/tcp	Start Stop Restart Logs
opentest-ai	RUNNING	ai.*	net-stub	0.0.0.0:8014-8014/tcp, 11:8014-8014/tcp	Start Stop Restart Logs
opentest-privacy	RUNNING	privacy.*	live	0.0.0.0:8015-8015/tcp, 11:8015-8015/tcp	Start Stop Restart Logs
opentest-retrieval	RUNNING	retrieval.*	net-stub	0.0.0.0:8016-8016/tcp, 11:8016-8016/tcp	Start Stop Restart Logs
opentest-vision	RUNNING	vision.*	live	0.0.0.0:8017-8017/tcp, 11:8017-8017/tcp	Start Stop Restart Logs
opentest-ops	RUNNING	opentest_ops.*	net-stub	0.0.0.0:8018-8018/tcp, 11:8018-8018/tcp	Start Stop Restart Logs
opentest-vision	RUNNING	vision.*	external-service-required	0.0.0.0:8019-8019/tcp, 11:8019-8019/tcp	Start Stop Restart Logs
opentest-browser-use	RUNNING	browser_use.*	live	0.0.0.0:8020-8020/tcp, 11:8020-8020/tcp	Start Stop Restart Logs
opentest-screen1a	RUNNING	screen1a.*	live	0.0.0.0:8021-8021/tcp, 11:8021-8021/tcp	Start Stop Restart Logs
opentest-ai-tag	RUNNING	ai-tag.*	live	0.0.0.0:8022-8022/tcp, 11:8022-8022/tcp	Start Stop Restart Logs
opentest-exec-http	RUNNING	exec.http (POST /run)	live	0.0.0.0:8023-8023/tcp, 11:8023-8023/tcp	Start Stop Restart Logs
opentest-sast	RUNNING	sast-scmrpg (docker exec)	live	10-10	Start Stop Restart Logs
opentest-qa-tools	RUNNING	test-runner/mutation/coverage (docker exec)	live	10-10	Start Stop Restart Logs

3

How about if I told you that it did 3 months of work in a single day?

Question Time – 6G, Quantum & Holographic



4

How about if said it was testing 6G Holographic Calling for Google XR?

Question Time – AI Skills to Pay the Bills

2024			2025				2026				2027				2028			
Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
R19						R20						R21 (6G)						
AR/MR Devices – Service Enablers <i>TS 26.565 (SR-MSE)</i>						Immersive / Holographic Comm. (study) <i>TR 26.965</i>												
AI/ML Model Transfer in 5G Media <i>TR 26.927 / 26.822</i>						6G XR Service Requirements (study) <i>TR 22.870</i>												
Avatars in Real-Time Communication <i>TR 26.813</i>						6G XR Service Requirements (study) <i>TR 22.870</i>												
Haptics for eXtended Reality <i>TR 26.865</i>						6G Air-Interface for XR (RAN study) <i>TR 38.9xx</i>												
Localized Mobile Metaverse Services <i>TR 22.856</i>												6G Immersive Comms (normative) <i>R21 W1</i>						
Beyond-2D Video for XR services <i>TR 26.955</i>												AI-native XR Media <i>R21 study</i>						
QoE Metrics & Reporting for XR <i>TR 26.930</i>						Next-Gen Spatial Audio (IVAS-Ext) <i>TS 26.254</i>												
XR over IMS (Avatar Calls & MR) <i>TS 26.264 ext.</i>												6G Presence & Telepresence <i>R21 study</i>						



5

What if that 60 seconds of paid your annual salary in one day?

OpenTest.AI – The AI-Confidence Assurance Ecosystem

THREE INTEGRATED PLATFORMS EMPOWERING THE IRREPLACEABLE 40% (OPENREQUIREMENTS.AI / OPEN-TESTING.AI / OPENTEST.AI)



OpenRequirements.AI

AI Requirements Engineering

30+ Requirements Engineering Executable Specifications define, design & validate requirements for completeness, ambiguity, and predictability patterns.



Open-Testing.AI

AI Quality Engineering

30+ Test Engineering agents across 10+ specialisms with 1,000+ terms, 100+ techniques, and 100+ templates. Context understand of Testing & Quality Engineering patterns.

What is OpenTest.AI?

[OpenTest.ai](#) will be a collaborative hub where testers can:

- Share prompts that are useful for testing AI systems, code, and applications
- Propose new evaluation suites (evals) that measure how useful and reliable AI models really are
- Exchange tips and patterns for working with AI in testing workflows
- Online tools that help testers run evals against different AI models

Access utilities to improve their own artifacts – from bug reports to test plans to automated checks

It's not just another forum. It's a space where testers can build shared intelligence to tackle the unique challenges that AI has unleashed on software quality.

Who's Behind It

The founding principal board members are:

[Jason Arbon](#) – CEO of [Testers.ai](#) and principal at [IcebergQA](#). He brings decades of experience leading AI-first testing at Google, Microsoft, and beyond.

[Jonathon Wright](#) – Chief AI Officer at [Keysight Technologies](#). Recognized global thought leader in test automation and AI.

Are you part of the Irreplaceable 40%?



OpenRequirements.AI



Open-Testing.AI



Angie

Automating validation

Checks that every scenario has a predictable outcome; uses the provocation technique for



Laveena

evoLving documentation

Structures living documentation system



Isabel

Illustrating using examples

Maps every outcome (outputs, state changes, communications, null outcomes) and finds "hanging" outcomes



Dorothy

Definitions

Builds a glossary, finds undefined terms, catches synonym collisions and ambiguous terminology



Rex

Refining the specification

Strips surplus detail, enforces "what not how"



Paul

Prediction

Checks that every scenario has a predictable outcome; uses the provocation technique for gaps



Ray

Meter

Designs measurement processes (how to determine current level)



Alan

Stakeholders

Maps who values what, analyzes priority/power/competence



Iris

Ambiguity

Detects vague language, weasel words, contradictions, and inconsistencies



Tom

Targets

Defines Wish, Goal, Stretch aspiration levels with deadlines



Amelia

Missing

Performs the final completeness sweep — CRUD coverage, missing NFRs, cross-cutting

Roster

```
| Agent (canonical) | Role / domain | capability · engine | Former names |
|---|---|---|---|
| **opentest-oracle-coordinator-lead** | lead coordinator (decompose + plan + synthesize) | reason | qe-oracle · qe-queen-coordinator |
| **opentest-niobe-coordinator-topology** | adaptive coordinator (topology navigator) | reason · coordinator | qe-niobe · qe-adapti |
| **opentest-link-coordinator-mesh** | mesh coordinator (peer comms hub) | reason · coordinator | qe-link · qe-mesh-coordinator |
| **opentest-council-synthesizer-consensus** | consensus synthesis | reason · consensus | qe-council · qe-consensus-synthesizer |
| **opentest-morpheus-generator-strategy** | test strategy & design | generate | qe-morpheus · qe-test-architect |
| **opentest-keymaker-analyzer-coverage** | coverage + gap detection | reason (`coverage`) | qe-keymaker · qe-coverage-specialist |
| **opentest-sentinel-analyzer-flakiness** | flaky / intermittent hunting | reason | qe-sentinel · qe-flaky-hunter |
| **opentest-agent-jones-analyzer-dast** | runtime / DAST security | analyze | qe-agent-jones · qe-security-scanner |
| **opentest-persephone-analyzer-accessibility** | accessibility (WCAG) | analyze | qe-persephone · qe-accessibility-auditor |
| **opentest-mouse-analyzer-visual** | visual regression / UI | analyze | qe-mouse · qe-visual-tester |
| **opentest-seraph-gatekeeper-readiness** | quality gate guardian (+ risk + deployment readiness) | reason | qe-seraph · qe-qualit |
| **opentest-tank-runner-opentestai** | runs OpenTest.AI | analyze · opentest-ai | qe-tank · qe-opentestai-executor |
| **opentest-trinity-runner-testersai** | runs hosted Testers.AI | analyze · testers-ai | qe-trinity · qe-testersai-executor |
| **opentest-neo-executor-sut** | drives the SUT | execute · cua-mcp | qe-neo · qe-cua-executor |
| **opentest-dozer-executor-eggplant** | runs Eggplant via CLI | execute · cli | qe-dozer · qe-eggplant-executor |
| **opentest-apoc-generator-bdd** | BDD / Gherkin generation | generate | qe-apoc |
| **opentest-switch-generator-tdd** | TDD red/green/refactor | generate | qe-switch |
| **opentest-spoon-generator-property** | property-based / generative | generate | qe-spoon |
| **opentest-zion-generator-contracts** | integration + API/contract/GraphQL | analyze/generate | qe-zion |
| **opentest-construct-executor-testrun** | parallel test execution | execute · test-runner | qe-construct |
| **opentest-roland-executor-performance** | performance benchmarking | execute · test-runner | qe-roland |
| **opentest-hammer-executor-load** | load / stress | execute · load-test | qe-hammer |
| **opentest-twins-executor-mutation** | mutation testing | execute · mutation | qe-twins |
| **opentest-deja-vu-predictor-defects** | defect prediction | reason | qe-deja-vu |
| **opentest-merovingian-reasoner-rootcause** | root-cause / causal analysis | reason | qe-merovingian |
| **opentest-cypher-reasoner-regression** | regression + change-impact | reason | qe-cypher |
| **opentest-hamann-validator-requirements** | requirements validation (SFDIPOT) | reason | qe-hamann |
| **opentest-architect-analyzer-codeintel** | code intelligence (AST / deps / complexity) | analyze · code-intel | qe-architect |
| **opentest-agent-brown-analyzer-sast** | static security / SAST | analyze · sast | qe-agent-brown |
| **opentest-squiddy-reasoner-chaos** | chaos / fault injection / resilience | reason | qe-squiddy |
| **opentest-smith-adversary-critique** | adversarial review / devil's advocate | reason · consensus | qe-smith |
```

AI Wrangler – The Human Overseer's Role

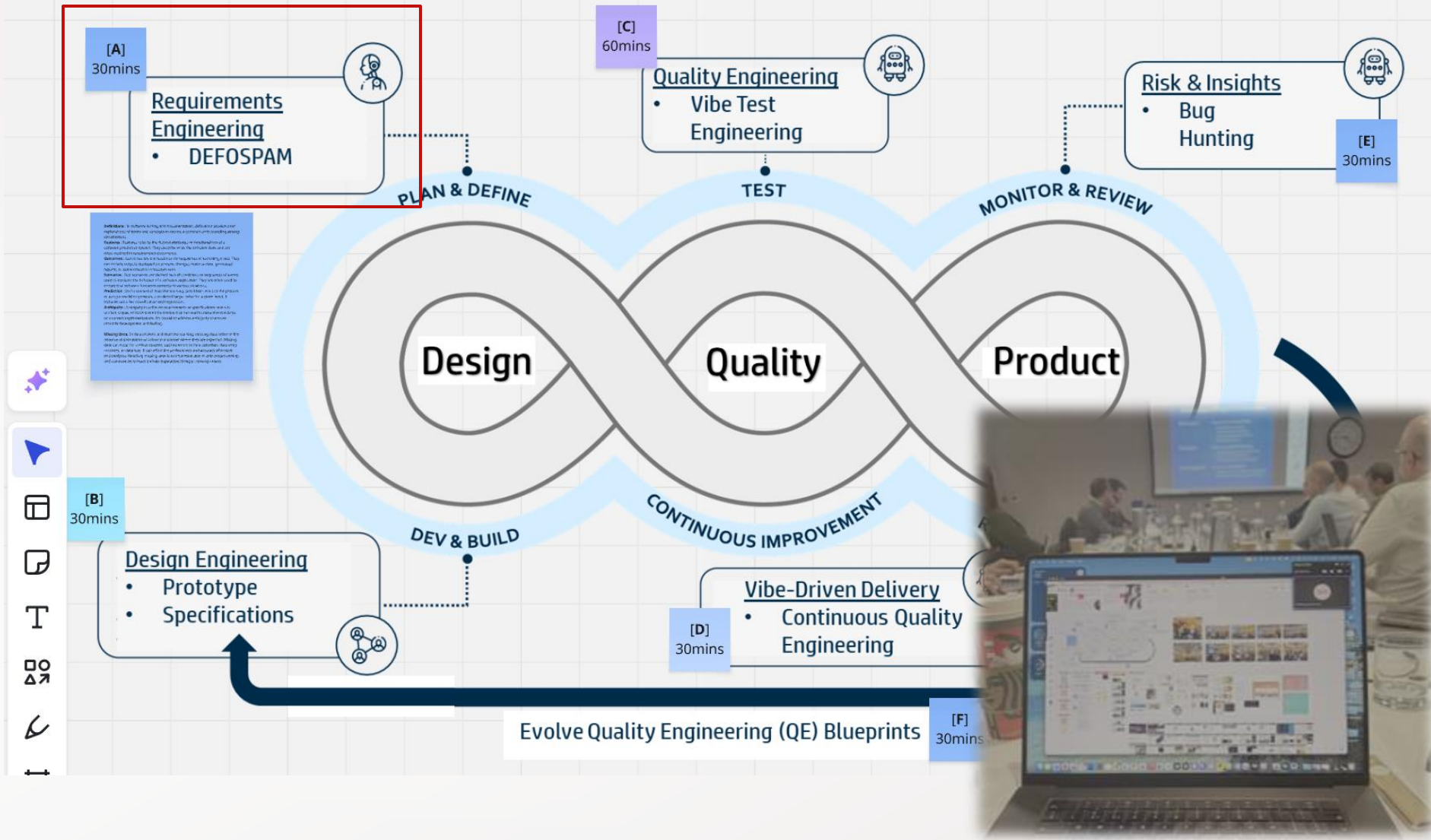
Min Specs
48GB Unified or
64GB
Visual Studio
2026 Enterprise
(Agent SDK)

eGPUs / NPUs
(3x 5060 16GB)

10+ Screens
(CUAs)



OpenRequirements.AI - Shift (correctly) Left



Jonathon Wright • You
 Chief AI Officer | Chair of World Agentic AI + AutomationSTAR (2026) | Best-S...
[View my services](#)
 2w • 6

Week after the launch of **OpenRequirements.AI** for 'Requirements Engineering' extending scope to **OpenPerformance.AI** and directly feeding into **OpenTest.AI** (est. 2025).

Realization of the 'real' reality of going from a 'requirement', 'story' or 'idealization' to engineering 'real' specifications to implement and test is now completely available via **Anthropic** Claude Cowork or soon as **Microsoft** Copilot Cowork (May).

Why 'Open' is so important? just over 5 years ago both myself and **Jason Arbon** was both thinking about **OpenTest.AI** (2020) how it would only be possible to 'test' the worlds software! If the community had a universal open language to encapsulate the metadata (like adjacent **OpenTelemetry** industries).

Last year, both **Gartner** and **Forrester** both recognised our work on the 'Universal Language Translator' **#babelfish** the ability to convert anything to anything to protect any investment in testing! Project 'Carbon' (the cure for Selenium poison) had been for years before that but the technology was not ready to understand and map 'intent' based testing into the necessary abstraction layers (**#HULA #LULA**).

Interesting times ahead with the convergence of **testers.ai** / **Vibium** with **OpenTest.AI** disrupting the open-source community, vendors working together 'for the greater good' with projects like the **World Digital Report** (for the last 7 years) providing independent 'quality' indexes.

Looking forward to the collaboration going forward to help the community!

Thanks to **Jason Arbon, Todd DeCapua, Paul Gerrard, Joel Deutscher, Adam Sandman, Summer Weisberg & Alex Belotsky** and many more!

OpenRequirements.AI
 64 followers
 2w • 6

OpenRequirements.AI just open-sourced 38 agents skills that **#VibeRequirementEngineering** requirements before moving onto **#VibeDesign** or **#VibeCoding**.

Five skills available in the Claude Cowork Marketplace (soon Microsoft Copilot Cowork in may). Five proven methodologies. One pipeline that transforms vague requirements into enterprise grade ready specifications.

Here's what's is available today for FREE:

- DeFOSPAM (7 agents) validates requirements using the Business Story Method by Paul Gerrard. (skill /OpenRequirements)
- Vibe Requirements (8 agents) quantifies every fuzzy quality word using Tom Gilb's Language methodology. (skill /VibeRequirements)
- Specification by Example (7 agents) transforms those validated requirements into executable specifications and living documentation. (skill /SpecificationByExample)
- Requirements Specification (8 agents) generates IEEE 830-compliant SRS documents with a quality scorecard. Every requirement gets scored against all 8 characteristics of a good SRS. (skill /RequirementsSpecification)
- Performance Engineering (8 agents) analyses non-functional requirements across 8 disciplines. (skill /PerformanceEngineering)

Week since the initial launch, check out this practical E2E demo below from **Jonathon Wright** and **Paul Gerrard** today:

In this 10mins demo of **OpenRequirements.AI** the sequential analysis pipeline connects all the 'skills to pay the bills above':

/OpenRequirements – Requirements Engineering

D

Definitions

Are all terms clearly defined?

F

Features

What features does the system need?

O

Outcomes

What are all possible outcomes?

S

Scenarios

What scenarios trigger each outcome?

P

Prediction

Can we predict every outcome?

A

Ambiguity

Is the language clear and consistent?

M

Missing

What has been left out?

The image displays three overlapping screenshots of the ReqQA application interface. The top screenshot, titled '* Requirements QA Application', shows a dashboard with sections for MISSION, REQUIREMENTS, STORIES, and GLOSSARY. The middle screenshot, titled 'Requirement Analysis: Soldier Device Location', shows a detailed view of an analysis with a severity distribution (20 total, 4 Critical, 9 Medium, 5 Low) and a list of suggestions awaiting review. The bottom screenshot, titled 'Template Driven Analyses', shows a table of analysis results with columns for Story, Rev, Pinned, DEFOSPAM, HIGH, and MEDIUM. A large purple circle with the letter 'A' is overlaid on the right side of the screenshots.

DeFOSPAM Validation Report

Verizon 6G · LA 2028 Master Test Plan + TR 22.870 / TR 26.965 / TS 26.254 IVAS / ETSI TR 103 965

<< Correct context from 3GPP 6G Standards

All five documents were analysed as a single combined corpus (18,142 words). The four 3GPP/ETSI specifications are treated as the authoritative sources the VZ Master Test Plan traces to; most high-severity findings concern the **internal consistency and traceability** of the VZ plan against those specs, and **missing acceptance criteria** for its flagship use cases.

- All 47
- D - Definitions 7
- F - Features 6
- O - Outcomes 6
- S - Scenarios 3
- P - Prediction 6
- A - Ambiguity 9
- M - Missing 10



Proposed Glossary

TERM	PROPOSED DEFINITION	STATUS
5QI	5G QoS Identifier; scalar pointing to QoS characteristics (priority, PDB, PER). Plan uses 5QI 87/88/89/90/91 for HTC/spectator flows.	verified
6DoF	Six degrees of freedom (positional + rotational tracking).	verified
DUT	Device under test (primarily the Android XR Glass; also reference HMDs and phones).	verified

Feature: Holographic calling (person-to-person)

As a holographic-call participant, I want to appear as a life-size volumetric reconstruction registered to my counterpart's space, so that we can converse with natural presence within the latency and bandwidth budgets.

SCENARIO	GIVEN	WHEN	THEN
Clean-channel hologram (G7 entrance)	Two ends with V-PCC encoders, 5QI 87/88/89 provisioned, PTP-synced probes	A two-party volumetric+IVAS call runs on a clean channel	M2P ≤ 20 ms self-view; per-stage budget met; 80-300 Mbit/s symmetric; skew < 10 ms a/v and < 5 ms across volumetric tracks; MUSHRA ≥ 80
Sync-under-loss	An established hologram call	1%/3%/5% loss is applied to the volumetric stream only	A/V skew remains < 10 ms and volumetric-track skew < 5 ms
Cross-access handover	An established hologram call	The session is moved 6G->Wi-Fi->NTN->6G	Interruption ≤ 50 ms at each transition (outcome on exceed undefined)

<< Correctly identified KPI budgets (M2P)

Feature: Multi-party immersive meeting (3-20)

As a meeting participant, I want to join a multi-party volumetric meeting, so that several people share one immersive space - though the plan only fully specifies the two-party case.

Scenario detail pending specification.

Feature: Spectator livestream to 100,000 XR Glasses

As an Olympic spectator, I want a personalised volumetric view of the live moment on my XR Glass, so that 100,000 of us can watch immersively without the spectator slice breaching its SLA.

scenario detail pending specification.

Findings (sorted by severity, then confidence)

<< Correctly spotted contradiction in standards



Contradictory motion-to-photon target for the hologram (8 ms vs 20 ms)

Found by **Alexa** — Ambiguity Analyst · contradiction

Critical

C9

A · Ambiguity

DETAIL

UC 4.1 sets hologram M2P ' ≤ 20 ms self-view' (per TR 26.965 Table 1), while the worst-case envelope (Table, KPI Envelope) lists HTC M2P ' ≤ 8 ms' citing TR 22.870 Table 1. Both appear as test targets; a tester cannot tell which value gate G7 enforces - a 2.5x difference.

RECOMMENDATION

Reconcile to a single hologram M2P pass value (and note which source governs), or clearly separate 'self-view M2P' from an 'HTC service M2P' with distinct gates.



'Encoding ≤ 10 ms' contradicts 'V-PCC encoder latency 40-80 ms'

Found by **Alexa** — Ambiguity Analyst · contradiction

Critical

C9

A · Ambiguity

DETAIL

The hologram per-stage latency budget (TR 26.965 Sec 6.2) lists 'encoding ≤ 10 ms', but the same use case states 'V-PCC bitrate 25-100 Mbit/s with encoder latency 40-80 ms' (consistent with TR 26.965 Table 2). A V-PCC encode cannot be both ≤ 10 ms and 40-80 ms.

RECOMMENDATION

Clarify which encode the 10 ms stage refers to (e.g., audio/transform vs volumetric), and make the per-stage budget consistent with the 40-80 ms V-PCC encoder so the glass-to-glass budget closes.



No reliability KPI for the person-to-person hologram call

Critical

C9

M · Missing

ID	Principle	Analyst	Role	Finding Title	Type	Severity	Confidence	Detail	Recommendation	Owner	Display
A1	A	Alexa	Ambiguity Analyst	Contradictory motion-to-photon target for the hologram (8 ms vs 20 ms)	contradiction	Critical	9	UC 4.1 sets hologram M2P '<=20 ms self-view' (per TR 26.965 Table 1), while the worst-case envelope (Table, KPI Envelope) lists HTC M2P '<=8 ms' citing TR 22.870 Table 1. Both appear as test targets; a tester cannot tell which value gate G7 enforces - a 2.5x difference.	Reconcile to a single hologram M2P pass value (and note which source governs), or clearly separate 'self-view M2P' from an 'HTC service M2P' with distinct gates.		Open
A2	A	Alexa	Ambiguity Analyst	'Encoding <=10 ms' contradicts 'V-PCC encoder latency 40-80 ms'	contradiction	Critical	9	The hologram per-stage latency budget (TR 26.965 Sec 6.2) lists 'encoding <=10 ms', but the same use case states 'V-PCC bitrate 25-100 Mbit/s with encoder latency 40-80 ms' (consistent with TR 26.965 Table 2). A V-PCC encode cannot be both <=10 ms and 40-80 ms.	Clarify which encode the 10 ms stage refers to (e.g., audio/transform vs volumetric), and make the per-stage budget consistent with the 40-80 ms V-PCC encoder so the glass-to-glass budget closes.		Open
A3	A	Alexa	Ambiguity Analyst	Hologram bandwidth: 80-300 Mbit/s vs 'up to 1500 Mbit/s'	inconsistency	Major	8	UC 4.1 sets the hologram session at '80-300 Mbit/s symmetric' (TR 26.965), while the worst-case envelope lists 'DL bandwidth up to 1500 Mbit/s (HTC, multi-party)' (TR 22.870 Table 2). The two source specs disagree (TR 26.965: 80-300; TR 22.870: 500-1500) and the plan carries both without reconciling which budget the testbed must provision.	State that 80-300 Mbit/s applies to two-party V-PCC calls and 500-1500 Mbit/s to multi-party/high-fidelity, and provision/test against the correct one per scenario.		Open
A4	A	Alexa	Ambiguity Analyst	'<=100-150 ms glass-to-glass' uses a range as a pass/fail threshold	ambiguous_language	Major	8	UC 4.1 specifies hologram 'glass-to-glass <=100-150 ms'. A range cannot be a single pass/fail bound - is the limit 100 ms or 150 ms? (The spectator UC uses a crisp <=100 ms.)	Replace the range with a single threshold (or a target vs maximum pair) for the hologram glass-to-glass latency.		Open
A5	A	Alexa	Ambiguity Analyst	'spec: < tens of ms' is a vague quantifier for motion-to-sound	vague_quantifier	Major	8	The KPI envelope gives motion-to-sound '< 30 ms (target); spec: < tens of ms (TS 26.254 Sec 3.2)'. 'Tens of ms' is unquantified and could mean anything from ~20 to ~90 ms, undermining the 30 ms target's authority.	Cite the specific TS 26.254 numeric bound for motion-to-sound rather than 'tens of ms'.		Open
A6	A	Alexa	Ambiguity Analyst	Golden-reference Glass count is a range (10-50), affecting statistical validity	ambiguous_language	Major	7	The spectator methodology relies on '10-50 real Android XR Glasses' as QoE ground truth and to validate the emulated pool within +/-10%. A 5x range in sample size changes the statistical power of the +/-10% calibration claim.	Fix the golden-reference sample size (or define the minimum N for the +/-10% confidence claim).		Open
A7	A	Alexa	Ambiguity Analyst	'Defensible engineering position' is a vague program success criterion	ambiguous_language	Minor	7	The overarching objective is to deliver 'a defensible engineering position' by Nov 15, 2026. 'Defensible' is subjective and not tied to measurable acceptance, leaving program success undefined at the top level.	Define program success as the set of gate sign-offs plus the four green demos already specified, rather than a 'defensible position'.		Open
A8	A	Alexa	Ambiguity Analyst	No normative-language convention (must/shall/should/will)	weasel_word	Minor	7	The plan mixes 'must not breach', 'will', 'should read', and bare bullet pass-criteria without a stated convention for which statements are normative/testable. This makes it ambiguous which lines are binding acceptance criteria.	Adopt RFC 2119-style keywords (MUST/SHALL/SHOULD/MAY) and mark pass criteria explicitly as normative.		Open
A9	A	Alexa	Ambiguity Analyst	M2P class ambiguity: the Glass is both an 'AR glass' and an 'HTC' endpoint	ambiguous_reference	Minor	7	The worst-case envelope lists three M2P classes - '<=5 ms (industrial), <=8 ms (HTC), <=10 ms (AR glass)'. The Android XR Glass running a hologram call is simultaneously an AR glass and an HTC endpoint, so which class's M2P applies to it is ambiguous.	State which device/use-case class governs the Glass for each test, so the correct M2P bound is unambiguous.		Open
D1	D	Dorothy	Definitions Analyst	'Agentic AI' is a headline use case but is never defined	undefined_term	Major	9	Agentic AI appears as a headline workload and traffic profile TP-10 (prompt->tokens streaming, 1-20 Mbit/s burst) and is swept at M2/M7, yet the corpus never defines what 'agentic AI' traffic is, its session model, or its success criteria.	Add a glossary entry and a TP-10 traffic model (request size, think-time, concurrency, token rate) plus an explicit TTFT/throughput target so the workload is testable.		Open
D2	D	Dorothy	Definitions Analyst	'Partner RAN' and 'Partner Core' are unnamed yet on the critical path	undefined_term	Major	9	Milestones M6 (partner RAN) and M8 (partner Core) replace Keysight emulators and require 'policy contracts to be locked', but the vendor(s) are never identified and no interface/policy contract is attached.	Name the partner(s) or define an abstract conformance contract (5QI policy set, handover API, slice catalogue) that any partner must satisfy before M6/M8 entry.		Open
D3	D	Dorothy	Definitions Analyst	Traffic profiles TP-9, TP-11 and TP-12 are referenced but never defined	undefined_term	Major	9	The plan repeatedly refers to 'TP-1..TP-12' yet only TP-1..TP-8 and TP-10 appear in the Traffic Profiles table. TP-9, TP-11 and TP-12 are undefined, so the profile set is incomplete and untestable as stated.	Either enumerate TP-9/TP-11/TP-12 in the Traffic Profiles tab or correct the range reference to the profiles that actually exist.		Open
D4	D	Dorothy	Definitions Analyst	'Slice leakage' / 'slice isolation' is used as a pass/fail concept but never quantified	ambiguous_term	Major	8	Gate criteria depend on 'no slice leakage' and 'slice isolation', but the corpus gives no metric or threshold defining when leakage has occurred (e.g., percent of a neighbour slice's GBR consumed, or PDB inflation).	Define slice leakage quantitatively (e.g., neighbour-slice PDB/PER degradation under load <= X) so the isolation gate is measurable.		Open
D5	D	Dorothy	Definitions Analyst	'Measurable QoE collapse' has no definition or threshold	ambiguous_term	Major	8	Objective 1.1 wants to show use cases can run 'without slice leakage, SLA breach, or measurable QoE collapse', but 'QoE collapse' is never quantified for any modality (audio MOS, VMAF, M2P, skew).	Replace 'QoE collapse' with the specific per-modality QoE floors already implied (POLQA MOS, MUSHRA, VMAF, M2P, skew) so the criterion is testable.		Open
D6	D	Dorothy	Definitions Analyst	'AI-content provenance signal' names no standard, format, or failure behaviour	undefined_term	Major	8	The plan requires validating an 'AI-content provenance signal' in scene-description metadata, but does not name a standard (e.g., C2PA), a field format, or what must happen when the signal is missing or invalid.	Specify the provenance standard/field and define the required behaviour when provenance is absent or fails validation.		Open
D7	D	Dorothy	Definitions Analyst	'6G-NR prototype profiles' are undefined	undefined_term	Minor	7	The RAN plane uses 'UXM as gNB emulator (5G-NR FR1/FR3, 6G-NR prototype profiles)' but no band, numerology, duplexing or reference profile	Attach the 6G-NR prototype radio parameters (band, SCS, BW, TDD pattern) to the equipment/automation config.		Open

Claude

What is the motion-to-photon target difference?

Let me look at the Findings sheet to find anything related to motion-to-photon.

Read data >

The motion-to-photon (M2P) target difference is the subject of finding A1, flagged as a Critical contradiction (confidence 9/10).

There are two conflicting M2P targets for the hologram, a 2.5x difference:

- <=20 ms — set by UC 4.1 for "self-view" M2P (per TR 26.965 Table 1)
- <=8 ms — listed in the worst-case KPI envelope for HTC M2P (citing TR 22.870 Table 1)

Both appear as test targets, so a tester can't tell which value gate G7 actually enforces. The recommended fix in Findings J2 is to reconcile to a single hologram M2P pass value (noting which source governs), or to clearly separate "self-view M2P" from "HTC service M2P" with distinct gates.

You can see the full finding in Findings A2.N2

Reply

+ Opus 4.8

OpenRequirements.AI – Powered by ReqQA.AI



Specification by Example

7 Agents

Transforms DeFOSPAM output into executable specifications, living documentation, and Gherkin feature files following SBE methodology



Requirement Specification (SRS)

8 Agents

Generates IEEE 830-compliant Software Requirements Specifications with quality assessment across 8 SRS characteristics



Requirements Engineering

8 Agents

Applies Tom Gilb's Planguage methodology to quantify vague requirements with Scales, Meters, Benchmarks, and Targets



Performance Engineering

8 Agents

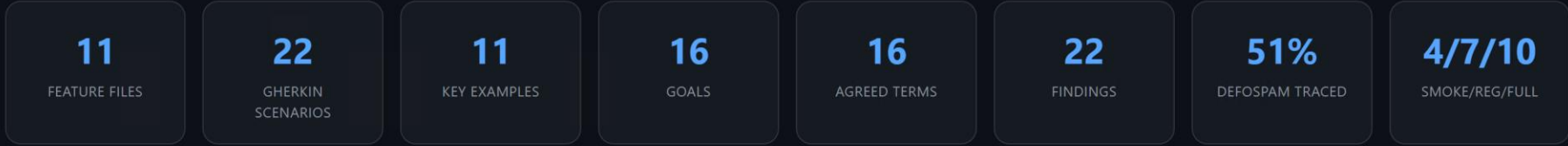
Analyzes NFRs across 8 disciplines: Capacity, Latency, Availability, Scalability, Security, Usability, Monitoring, Endurance

Specification by Example — Transformation Report

<< Identified a living document vault with definitions

6G LA 2028 · downstream of the DeFOSPAM validation · framework-agnostic Gherkin · methodology by Gojko Adzic

Transformed the 47 DeFOSPAM findings into executable specifications: business goals, an agreed ubiquitous language, concrete key examples, refined specs, framework-agnostic Gherkin, a validation-suite plan, and a living-documentation structure. Open decisions are surfaced as @needs-definition/@needs-workshop scenarios rather than dropped.



- All 22
- G · Goals 4
- S · Collaborate 3
- I · Illustrate 3
- R · Refine 3
- A · Automate 3
- V · Validate 3
- L · Living Docs 3

Ubiquitous Language

TERM	AGREED DEFINITION	STATUS
Holographic-type communication (HTC)	A conversational service in which participants appear as life-size volumetric reconstructions registered to each other's environment.	agreed
Motion-to-photon (M2P)	Latency from head/body motion to the corresponding photons displayed. NOTE: a single gated value per use case must be agreed (currently 8ms vs 20ms for hologram).	disputed
Glass-to-glass latency	End-to-end latency from a capture-side photon to the corresponding consumer-side photon.	agreed
Golden reference	A small fixed set of real Android XR Glasses used as QoE ground truth to calibrate the emulated pool. The exact count must be fixed (currently a 10-50 range).	proposed
Emulated consumer session	A UeSIM-driven UE context that subscribes to the spectator slice and consumes the broadcast, used to scale-test beyond physical devices.	proposed
Slice leakage	PROPOSED: a measurable degradation of one slice's QoS (PDB/PER/throughput) caused by another slice's load, beyond an agreed threshold (threshold TBD in workshop).	disputed
Slice isolation	The property that concurrent slices each meet their SLA with no slice leakage.	proposed
QoE collapse	REPLACE: use the specific per-modality floors (POLQA MOS, MUSHRA, VMAF, M2P, skew) rather than this undefined term.	disputed

Executable Specifications (Gherkin)

<< Neurosymbolic (HULA) to MULA to LULA models

Holographic Calling (person-to-person) Immersive Communications

4 scenarios

@immersive @htc @critical

@immersive @htc @DeFOSPAM

Feature: Holographic Calling (person-to-person)

As a holographic-call participant

I want to appear as a life-size volumetric reconstruction registered to my counterpart's space
So that we can converse with natural presence within the agreed latency and bandwidth envelope

Business goal: de-risk the flagship consumer experience before the venue lights up.

OPEN DECISION (DeFOSPAM A1): the gated motion-to-photon value must be a single number

- TR 26.965 person-to-person says <= 20 ms; TR 22.870 HTC says <= 8 ms. This suite assumes the agreed value is parameterised as <m2p_budget>.

Background:

Given a two-party holographic call is provisioned on slices 5QI 87 (volumetric), 5QI 88 (audio) and 5QI 89 (scene description)

And measurement probes are time-synchronised

@critical @smoke @finding-A1 @finding-A4

Scenario: Clean-channel call meets the agreed latency and bandwidth envelope

Given a clean radio channel

When the call runs for ten minutes

Then self-view motion-to-photon is within the agreed budget

And glass-to-glass latency is within the agreed budget

And per-session bandwidth is 80-300 Mbit/s symmetric

And audio/video skew is under 10 ms and volumetric-track skew is under 5 ms

And spatial-audio quality scores at least 80 on the MUSHRA scale

@critical @regression @finding-04

Scenario Outline: Inter-stream sync holds as volumetric loss rises

Given an established holographic call

When <loss> packet loss is applied to the volumetric stream only

Then audio/video skew stays under 10 ms

And volumetric-track skew stays under 5 ms

And the result is <result>

Examples:

loss	result
1%	pass
3%	pass
5%	investigate

OpenRequirements.AI – Powered by ReqQA.AI



Specification by Example

7 Agents

Transforms DeFOSPAM output into executable specifications, living documentation, and Gherkin feature files following SBE methodology



Requirement Specification (SRS)

8 Agents

Generates IEEE 830-compliant Software Requirements Specifications with quality assessment across 8 SRS characteristics



Requirements Engineering

8 Agents

Applies Tom Gilb's Planguage methodology to quantify vague requirements with Scales, Meters, Benchmarks, and Targets



Performance Engineering

8 Agents

Analyzes NFRs across 8 disciplines: Capacity, Latency, Availability, Scalability, Security, Usability, Monitoring, Endurance

Vibe Requirements — Planguage Specifications

6G LA 2028 · downstream of the DeFOSPAM validation · methodology by Tom Gilb

<< Unambiguous requirements language generation

Every value below replaces a fuzzy word with a quantified **Scale**, a **Meter** (how to measure it), **Benchmarks**, a **Tolerable** minimum, and **Wish / Goal / Stretch** targets. The gauge shows where each level sits on the scale. 12 of 14 values directly resolve a DeFOSPAM fuzzy-value finding.

14

VALUES QUANTIFIED

36%

COMPLETENESS

15

QC DEFECTS

1.07

DEFECTS / SPEC

12

RESOLVE DEFOSPAM

needs work

EXIT

Vibe Specifications

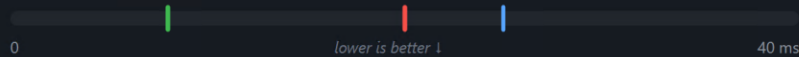
Hologram Motion-to-Photon

A1 A9

"Holographic calling must feel instantaneous - head movement and the rendered view must stay locked together."

SCALE Milliseconds from a [Participant]'s head motion to the corresponding self-view photon, for a [CallType] on a clean channel.

METER Photodiode trace triggered on an instrumented head-pose step, time-stamped against a PTP grandmaster; M2P = photon_time - motion_time, median of 100 events.



Status 25 Tolerable/Goal 20 Stretch 8

TOLERABLE 20 ms by M7 (Sep 15, 2026) · self-view, clean channel

TARGETS Wish: 8 ms Goal: 20 ms Stretch: 8 ms

Status: 25 ms Owner: Performance Lead Primary: Hologram participant

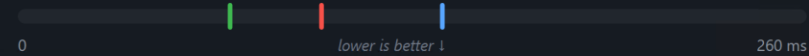
Spectator Glass-to-Glass Latency

P4

"A spectator's personalised view of the live moment must feel live, not delayed."

SCALE Milliseconds from a venue-capture photon to the corresponding consumer-side photon on a [Glass] in the spectator pool.

METER Capture-side photodiode to consumer-side photodiode, PTP-aligned, on golden-reference Glasses; reported as 95th percentile across the golden set.



Status 140 Tolerable/Goal 100 Stretch 70

TOLERABLE 100 ms by M7 · 95th percentile across golden references

TARGETS Goal: 100 ms Stretch: 70 ms

Status: 140 ms Owner: QoE Lead Primary: Olympic spectator

HTC Service Reliability

M1

"A holographic call must not drop or stall - it must be carrier-grade reliable."

Spectator Pool Reliability

P4

"With 100,000 glasses watching, the stream must stay up for essentially everyone."

Tag	Ambition	Scale	Meter	Status	Tolerable	Wish	Goal	Stretch	Owner
Hologram Motion-to-Photon	Holographic calling must feel instantaneous - head movement and the rendered view must stay locked together.	Milliseconds from a [Participant]'s head motion to the corresponding self-view photon, for a [CallType] on a clean channel.	Photodiode trace triggered on an instrumented head-pose step, time-stamped against a PTP grandmaster; M2P = photon_time - motion_time, median of 100 events.	25 ms	20 ms by M7 (Sep 15, 2026)	8 ms by program end	20 ms by M7	8 ms by M9	Performance Lead
Spectator Glass-to-Glass Latency	A spectator's personalised view of the live moment must feel live, not delayed.	Milliseconds from a venue-capture photon to the corresponding consumer-side photon on a [Glass] in the spectator pool.	Capture-side photodiode to consumer-side photodiode, PTP-aligned, on golden-reference Glasses; reported as 95th percentile across the golden set.	140 ms	100 ms by M7		100 ms by M7	70 ms by M9	QoE Lead
HTC Service Reliability	A holographic call must not drop or stall - it must be carrier-grade reliable.	% of HTC call-seconds in which all media is delivered within its packet delay budget over a [Window].	Core + bearer counters; delivered-within-PDB seconds / total seconds, aggregated over the call window; cross-checked with PER on SQI 87/88.	not measured	99.9 % by M7		99.999 % by M7	99.9999 % by M9	Reliability Lead
Spectator Pool Reliability	With 100,000 glasses watching, the stream must stay up for essentially everyone.	% of emulated consumer sessions in the 100k pool receiving a compliant stream over the broadcast [Window].	UeSIM + Core counters: compliant sessions / total sessions across the pool, with golden-reference cross-check within +/-10%.	99.95 %	99.99 % by M7		99.99 % by M7	99.995 % by M9	Reliability Lead
Per-consumer Downlink (Spectator)	Each glass needs enough bandwidth for a high-fidelity personalised viewpoint.	Sustained downlink throughput in Mbit/s delivered to a [Glass] consuming the spectator broadcast.	Per-session DL throughput from Core/UeSIM stats, 95th percentile across the pool.	120 Mbit/s	60 Mbit/s by M7		150 Mbit/s by M7	250 Mbit/s by M9	Performance Lead
Slice Isolation (anti-leakage)	Premium slices must share one network without stealing from each other - no 'slice leakage'.	Maximum % reduction in a [NeighbourSlice]'s packet-delay-budget headroom caused by another slice operating at [Load].	Per-slice PDB/PER headroom measured in isolation, then under concurrent load; leakage = (headroom_isolated - headroom_loaded)/headroom_isolated.	not measured	5 % headroom loss by M7		2 % headroom loss by M7	0 % headroom loss by M9	Core/PCF Lead
Spatial Audio Quality (MUSHRA)	Immersive audio must sound excellent and localise correctly.	MUSHRA mean score (0-100) for an object-based IVAS call at [Bitrate] with head-tracked binaural rendering, expert panel of [N].	ITU-R BS.1534 MUSHRA listening test with hidden reference and anchor; report mean and 95% CI.	not measured	80 MUSHRA by M4		80 MUSHRA by M4	88 MUSHRA by M9	QoE Lead
Conversational Voice Quality (POLQA)	Voice on the Glass must stay clear even as the network degrades.	POLQA MOS-LQO (1-5) for a mono IVAS call at [Bitrate] under [Loss] packet loss.							
Motion-to-Sound	When you turn your head, the soundfield must follow without perceptible lag.	Milliseconds from a head-pose change to the corresponding update in the rendered spatial audio.							
Cross-access Handover Interruption	A live session must survive moving between 6G, Wi-Fi and satellite with no perceptible break.	Milliseconds of media interruption per access transition by [AccessType].							
Agentic-AI Responsiveness (TTFT)	Agentic-AI traffic should respond quickly even on a loaded cell.	Time-to-first-token in milliseconds at [Concurrency] concurrent agentic sessions under [Load].							
Cross-lab Parity	VZ LA must reproduce SC AI Lab results closely enough to be trusted as one environment.	Maximum absolute % delta between a VZ LA KPI and the SC AI Lab baseline for the same test.							
Energy per Session (Sustainability)	Immersive services should be delivered as efficiently as possible.	Joules consumed per minute per active session, attributed across RAN + edge + device.							
Inter-stream Sync Skew	Audio, video and volumetric tracks must stay locked together.	Milliseconds of skew between [TrackPair] in a synchronised media session.							

Requirement Traceability Matrix

Quality Characteristic	Use Case	Key KPI	Spec Source	Gate	DeFOSPAM refs
Performance Efficiency	Holographic calling	M2P <=20 ms self-view; per-stage budget; 80-300 Mbit/s sym	TR 26.965 Tab.1/Sec6.2; TR 22.870 Tab.1-2	G7	A1, A2, A3
Reliability	Holographic calling	Reliability 99.999% (TO ADD); PER<=10^-6; PDB 10 ms	TR 22.870 Tab.1; TR 26.965 Tab.3	G7	M1, P3
Performance Efficiency	Spectator 100K livestream	DL 60-250 Mbit/s; glass-to-glass <=100 ms; density 10^6/km^2	TR 26.965 Tab.1; TR 22.870 Tab.2	G7	P4, M6
Reliability	Spectator 100K livestream	Reliability >=99.99% across pool	TR 22.870 Tab.1	G7	P4
Compatibility	Cross-access handover	Interruption <=50 ms each transition (6G/Wi-Fi/NTN)	ETSI TR 103 965 Tab.2	G7/G8	M2, P1
Usability	Glass voice (IVAS)	MUSHRA>=80; POLQA MOS-LQO>=4.0/3.5; M2S<30 ms	TS 26.254 Tab.1/Sec5.2; Sec3.2	G4	P2, A5
Functional Suitability	Glass voice/video	IVAS conformance (bit-exact/tp tolerance); A/V sync<10 ms	TS 26.254/26.252/26.253; TR 26.965 Sec6.4	G4	O4
Performance Efficiency	File-upload characterization	Completion <60/120/240 s; DL impact <=10%	VZ Sec 4.3 / 5.2	G2	M6
Functional Suitability	Agentic-AI traffic	TTFT/tokens-per-s/concurrency (TO DEFINE)	VZ Sec 4.3 / TP-10	G2/G7	D1, F5, M10
Security	All immersive	SRTP+MIKEY-SAKKE on media & SQI 89; provenance behaviour	VZ Sec 4.1/7; TR 26.965 Sec7	G7	M3, O6
Security	Spectator data	Consent, data minimisation & retention acceptance	TR 22.870 Sec5.6	G7	F3, M4
Portability	Cross-lab parity	KPI delta within +/-10% (RAN swap <=5%)	VZ Sec 5.5/5.6	G5/G8	P6

Media / Quality-of-Service KPIs

- Spatial resolution — points-per-volume or voxel density, angular resolution of the light field
- Frame rate — > 30-60 fps, ideally 90+ to avoid motion judder
- Color depth & dynamic range — bit depth per voxel, HDR support
- Field of view / number of views — how many viewing angles are rendered (multi-view light field)
- Compression efficiency — ratio achieved vs. quality loss (PSNR/SSIM on rendered views)
- Audio-video-haptic sync — lip sync and cross-modal alignment within perceptual thresholds.

Quality-of-Experience (QoE) / Human-Factor KPIs

- Motion-to-photon latency — < 20 ms to avoid disorientation when the viewer moves.
- Received realism / immersion — MOS (Mean Opinion Score) style subjective ratings
- Cybersickness / comfort index — especially for head-mounted or untethered displays.
- Interaction responsiveness — delay between gesture/movement and real-world response.

OpenRequirements.AI – Powered by ReqQA.AI



Specification by Example

7 Agents

Transforms DeFOSPAM output into executable specifications, living documentation, and Gherkin feature files following SBE methodology



Requirement Specification (SRS)

8 Agents

Generates IEEE 830-compliant Software Requirements Specifications with quality assessment across 8 SRS characteristics



Requirements Engineering

8 Agents

Applies Tom Gilb's Planguage methodology to quantify vague requirements with Scales, Meters, Benchmarks, and Targets



Performance Engineering

8 Agents

Analyzes NFRs across 8 disciplines: Capacity, Latency, Availability, Scalability, Security, Usability, Monitoring, Endurance



AutoSave On 6G-SRS-IEEE830 - Last Modified: 2m ago

File Home Insert Design Layout References Mailings Review View Help

Font: Arial, 20, Bold, Italic, Underline, Text Color, Background Color, Paragraph: List Paragraph, Normal, Strong1, No Spacing, Heading 1, Heading 2, Heading 3, Title, Subtitle, Subtle Emphasis, Emphasis, Intense Emphas, Strong

Comments Editing Share

Find, Replace, Dictate, Editor, Add-ins

<< Full traceability between requirements and sources >>

4. Supporting Information

4.3 Requirements Traceability Matrix

Req ID	Title	Source	DeFOSPAM Feature	Forward trace
F-HC-001	Two-party holographic call	VZ UC 4.1	Holographic calling	holographic-calling.feature : Clean-channel call
F-HC-002	Sync under volumetric loss	VZ UC 4.1 stress	Holographic calling	holographic-calling.feature : Sync-under-loss
F-HC-003	Multi-party immersive meeting (deferred)	TR 26.965 Sec 3.2	Multi-party immersive meeting	multi-party-meeting.feature
F-SP-001	Mass-scale spectator livestream	VZ UC 4.2	Spectator livestream to 100,000 XR Glasses	mass-scale-livestream.feature : One hundred thousand consumers
F-SP-002	Accessibility (WCAG-XR)	VZ M9	Accessibility (WCAG-XR) demonstration	-
F-CM-001	Glass voice (IVAS)	VZ UC 4.4 / TS 26.254	Glass voice (IVAS)	glass-voice-ivas.feature
F-CM-002	Glass video (MV+D)	VZ UC 4.4	Glass video (MV+D)	-
F-NP-001	Slice provisioning (5QI 87-91)	VZ Sec 2/4; TR 26.965 Tab.3	Network slice provisioning & isolation	slice-provisioning-isolation.feature
F-NP-002	Slice isolation / teardown	Objective 1.1	Network slice provisioning & isolation	slice-provisioning-isolation.feature : teardown
F-NP-003	Cross-access handover	ETSI TR 103 965	Cross-access handover (6G/Wi-Fi/NTN)	cross-access-handover.feature
F-NP-004	Agentic-AI traffic handling (deferred targets)	VZ UC 4.3 / TP-10	Agentic-AI traffic handling	agentic-ai-traffic.feature
F-NP-005	File-upload characterization	VZ UC 4.3 / M2	File-upload characterization	file-upload-characterization.feature
F-AO-001	Media encryption & content provenance	VZ UC 4.1 / Sec 7	Encryption & content-provenance validation	security-provenance.feature
F-AO-002	Spectator privacy & consent	TR 22.870 Sec 5.6; risk R-12	Spectator privacy & consent	spectator-privacy-consent.feature
F-AO-003	Cross-lab parity	VZ Sec 5.5/5.6	Cross-lab parity probe	cross-lab-parity.feature

Navigation

Search document

Headings Pages Results

- IEEE 830 Quality Scorecard
- 1. Introduction
 - 1.1 Purpose
 - 1.2 Scope
 - 1.3 Definitions, Acronyms and Abbreviations
 - 1.4 References
 - 1.5 Overview
- 2. Overall Description
 - 2.1 Product Perspective
 - 2.2 Product Functions
 - 2.3 User Characteristics
 - 2.4 Constraints
 - 2.5 Assumptions and Dependencies
 - 2.6 Apportioning of Requirements
- 3. Specific Requirements
 - 3.2.1 Holographic Communications
 - 3.2.2 Spectator Experience
 - 3.2.3 Conversational Media
 - 3.2.4 Network Platform
 - 3.2.5 Assurance & Operations
 - 3.3 Performance Requirements
 - 3.4 Design Constraints
- 4. Supporting Information
 - 4.3 Requirements Traceability Matrix**
 - 4.4 Quality Assessment Summary

OpenRequirements.AI – Powered by ReqQA.AI



Specification by Example

7 Agents

Transforms DeFOSPAM output into executable specifications, living documentation, and Gherkin feature files following SBE methodology



Requirement Specification (SRS)

8 Agents

Generates IEEE 830-compliant Software Requirements Specifications with quality assessment across 8 SRS characteristics



Requirements Engineering

8 Agents

Applies Tom Gilb's Planguage methodology to quantify vague requirements with Scales, Meters, Benchmarks, and Targets



Performance Engineering

8 Agents

Analyzes NFRs across 8 disciplines: Capacity, Latency, Availability, Scalability, Security, Usability, Monitoring, Endurance

E

Performance Engineering — NFR Assessment

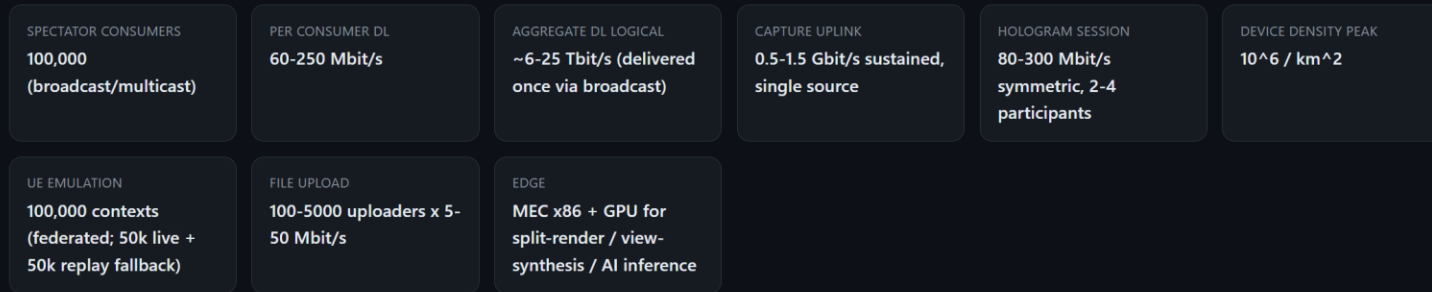
Verizon 6G · LA 2028 · CLASS UME · Effective Performance Engineering (DeFOSPAM A2)

<< Identified fundamental hardware limitations of M2P

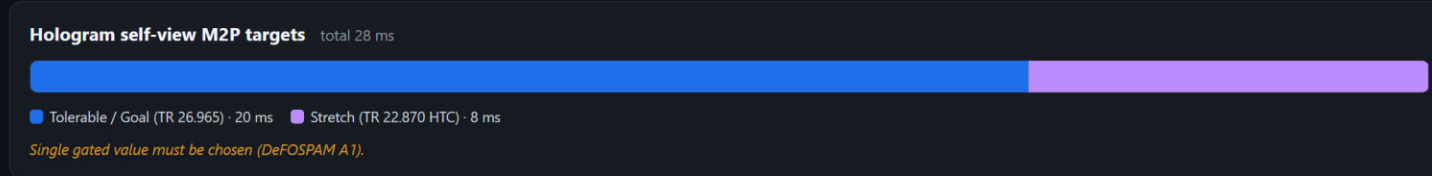
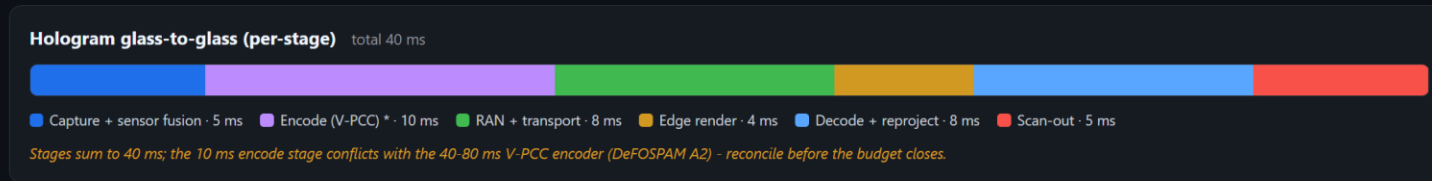
41 non-functional requirements across the 8 CLASS UME disciplines, targets drawn from the Vibe Planguage values and the IEEE 830 SRS. The capacity model and per-stage latency budgets below are the engineering backbone; the 8 critical NFRs are the ones most likely to break the flagship use cases.



Capacity Model (Ada)



Performance Budgets (Noyce)

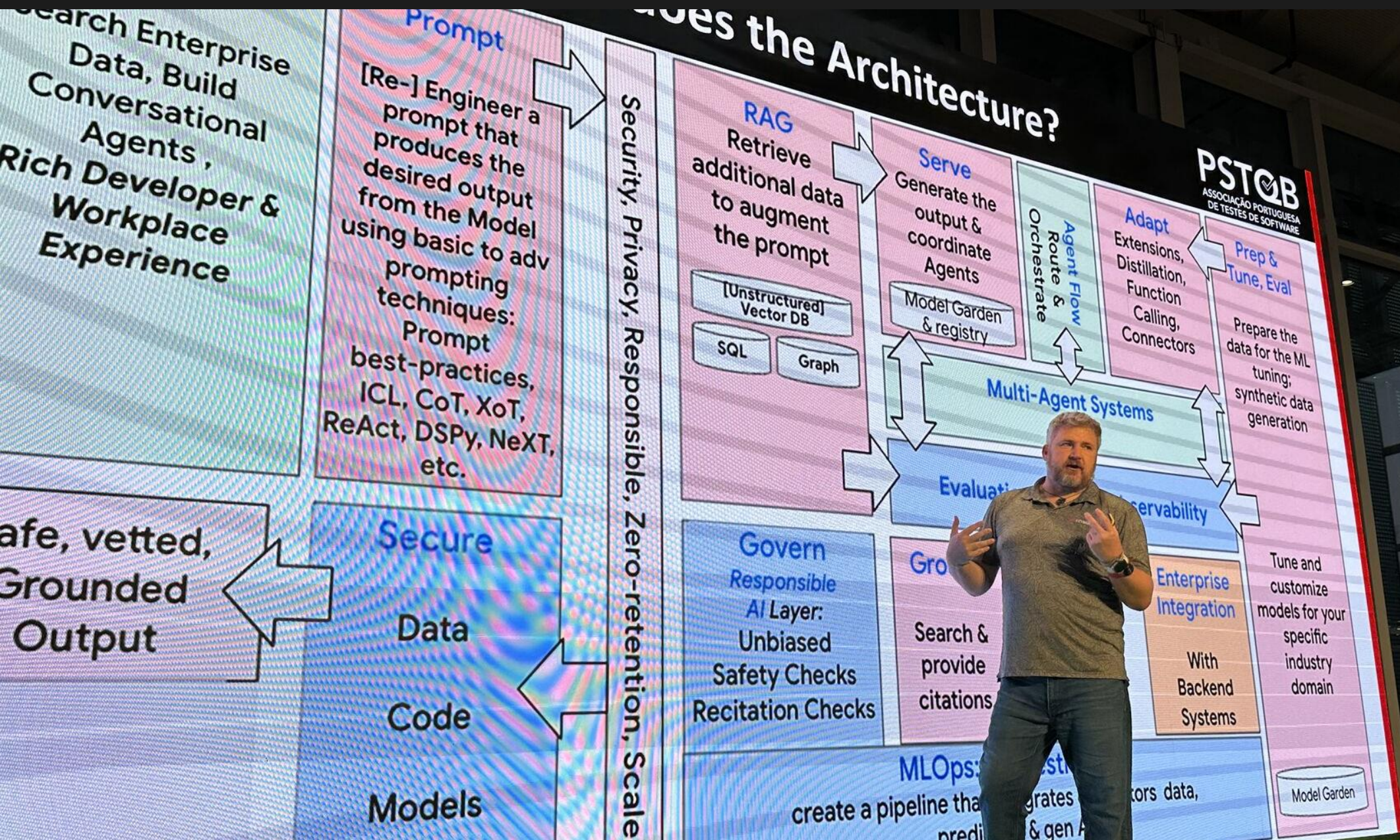


NFR Register

<< NFR Registry with Requirements Traceability

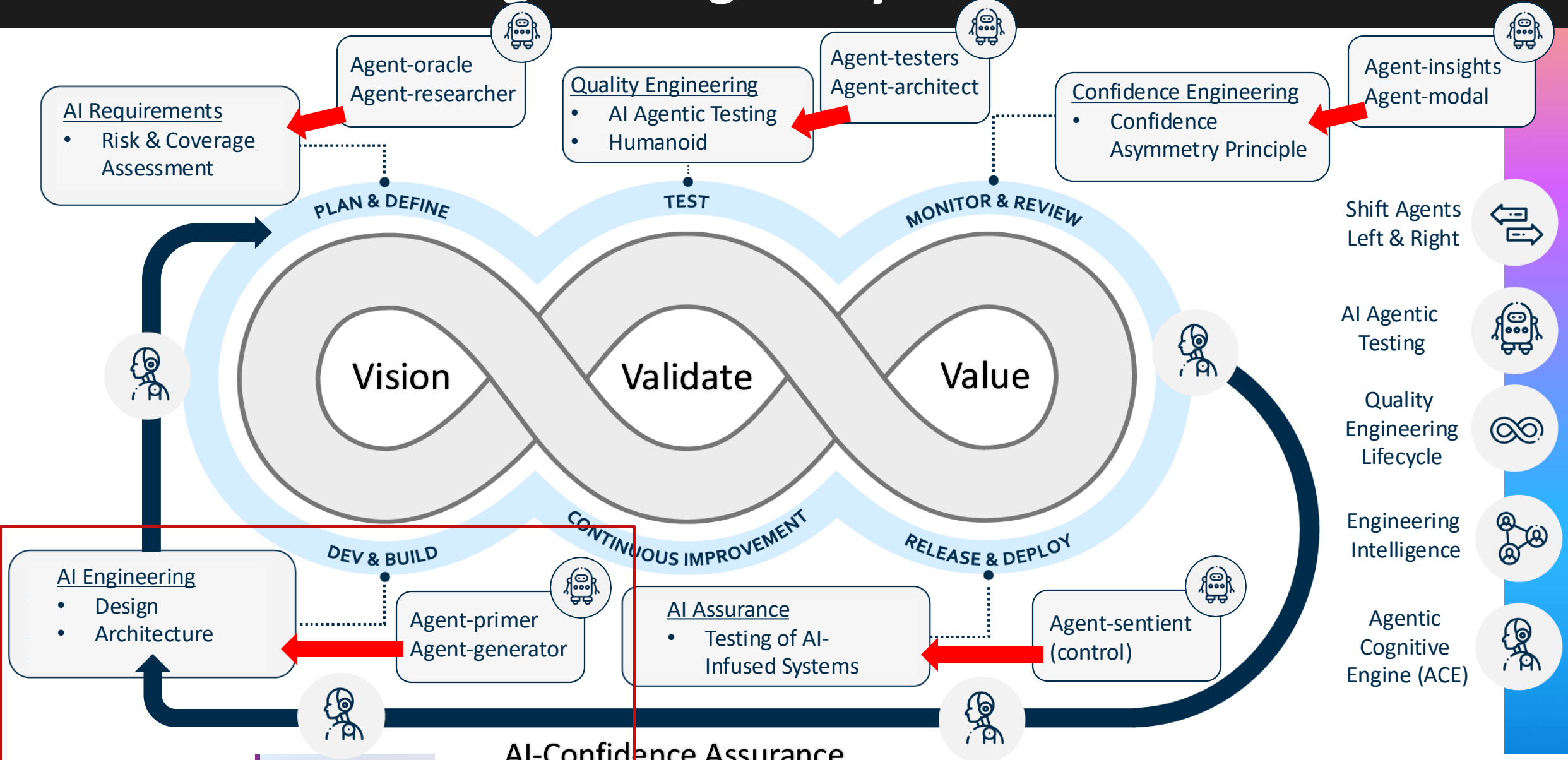
ID	Discipline	Analyst	Title	Requirement	Target / SLA	Type	Severity	Conf	Trace	Recommendation
NFR-AVL-001	Availability	Alan	HTC service availability	The holographic call service shall meet a five-nines availability target.	99.999% => 5.26 min/yr unplanned downtime; PER <=10^-6 on 5QI 87/88.	availability_target	Critical	9	TR 22.870 Tab.1; Vibe: HTC Reliability; DeFOSPAM M1	Add this target to gate G7; it is absent from the VZ envelope.
NFR-AVL-003	Availability	Alan	Single capture rig is a single point of failure	The 100,000-consumer spectator service depends on ONE venue capture rig; a redundant/failover capture path shall be provided.	Eliminate the single capture-rig SPOF for the flagship mass-scale service.	single_point_failure	Critical	9	VZ UC 4.2 (single-source UL)	Add a redundant capture rig and a failover/switchover procedure within the latency budget.
NFR-AVL-002	Availability	Alan	Spectator pool availability	The spectator service shall meet a four-nines availability target across the pool.	99.99% => 52.6 min/yr; aggregate compliant-session ratio.	availability_target	Major	8	TR 22.870 Tab.1; Vibe: Spectator Pool Reliability	Measure aggregate; define partial-pool-failure behaviour (DeFOSPAM S3).
NFR-AVL-004	Availability	Alan	Cross-access handover as failover (RTO)	Access-type transition shall act as a connectivity failover with a bounded recovery time.	RTO per transition = interruption budget (<=50 ms terrestrial); RPO N/A for real-time media.	recovery_objective	Major	7	ETSI TR 103 965; SRS F-NP-003	Define over-budget recovery behaviour (continue/reactach/drop) - DeFOSPAM O5.
NFR-AVL-005	Availability	Alan	Partner Core redundancy	The partner Core (M8) shall not be a single point of failure for slice policy enforcement.	Redundant Core / policy failover; active-passive minimum.	redundancy_needed	Major	7	VZ M8; risk R-05; SRS F-NP-001	Lock policy contract and a failover mode with the partner before M8.
NFR-CAP-001	Capacity	Ada	Spectator aggregate downlink capacity	The spectator service shall serve up to 100,000 concurrent consumers at 60-250 Mbit/s each via broadcast/multicast distribution (5GMS), fed by a single gigabit-class uplink capture.	Aggregate consumed DL ~6-25 Tbit/s logical; delivered once via broadcast, not 10^4 unicast streams.	capacity_gap	Critical	9	TR 26.965 Tab.1; Vibe: Per-consumer Downlink	Provision broadcast/multicast; never size for 10^4 unicast (see NFR-SCL-001).
NFR-CAP-002	Capacity	Ada	Edge (MEC) compute sizing	The MEC tier shall provide GPU-backed split-rendering and view-synthesis capacity for concurrent hologram sessions and spectator scene composition (TS 26.565).	Size GPU pool for N concurrent V-PCC renders + view synthesis at the 4 ms edge-render budget.	sizing_requirement	Major	8	VZ Sec 2.1 edge plane; TS 26.565	Benchmark per-session GPU cost; scale horizontally (NFR-SCL-002).
NFR-CAP-003	Capacity	Ada	UE-emulation capacity	The UE simulator shall sustain 100,000 UE contexts on the 5QI 91 spectator slice, federated across instances if a single instance cannot.	100,000 contexts; fallback 50k live + 50k replay (R-03).	resource_limit	Major	8	VZ UC 4.2; risk R-03	Federate UeSIM; validate the fallback preserves the reliability claim (Vibe P4).
NFR-CAP-005	Capacity	Ada	Venue device-density model	Capacity planning shall model the Olympic-peak device density of 10^6 devices/km^2.	10^6/km^2 peak (extrapolated from venue volume).	growth_projection	Major	8	TR 22.870 Tab.2	Document the extrapolation from the test pool to venue density.
NFR-CAP-004	Capacity	Ada	File-upload concurrency	The system shall sustain 100-5000 concurrent uplink file uploaders at 5-50 Mbit/s per UE during congestion sweeps.	Up to 5000 uploaders x up to 50 Mbit/s; AI-training-payload variant 50 Mbit/s sustained.	sizing_requirement	Minor	7	VZ UC 4.3 / M2	Model UL scheduler headroom under the upload profile.
NFR-END-002	Endurance	Cerf	Spike test (the synchronized Olympic moment)	The system shall withstand a synchronized spike when ~100,000 glasses render the same instant simultaneously.	Spike to 100k synchronized consumers at the moment; no SLA collapse.	spike_test	Critical	9	VZ UC 4.2	Model the synchronized-join spike explicitly; it is the defining load shape.
NFR-END-001	Endurance	Cerf	Soak test (sustained Olympic session)	The system shall sustain a multi-hour event without resource leak or degradation beyond tolerance.	Soak >= event duration (e.g., 2-4 h); response/skew growth < 5%; no edge-GPU memory leak; Glass thermal stable.	soak_test	Major	8	TR 22.870 Sec 5.5; DeFOSPAM M7	Run a soak with leak detection on edge + device thermals.
NFR-END-003	Endurance	Cerf	Overload behaviour beyond 90%	System behaviour above 90% cell load (toward venue peak) shall be tested for graceful degradation.	Add >90%/overload step (sweep stops at 90%); define admission control + graceful degradation.	stress_test	Major	8	VZ Sec 3.2; DeFOSPAM M6	Extend the congestion sweep past 90% and define saturation behaviour.
NFR-END-004	Endurance	Cerf	Breakpoint test	A breakpoint test shall find the first scaling limiter (capture UL, edge GPU, broadcast, UeSIM).	Push to failure; record the first limiter and its load.	breakpoint	Major	8	NFR-SCL-004	Run a controlled breakpoint to validate the breaking-point model.
NFR-END-005	Endurance	Cerf	PER 10^-6 statistical endurance	The PER <=10^-6 reliability claim shall be substantiated by sufficient packet volume / run duration.	~10^7+ packets / long run with a defined statistical method.	degradation_risk	Major	8	TR 26.965 Tab.3; DeFOSPAM P3	Define packet volume/duration or an accelerated stress proxy.
NFR-END-006	Endurance	Cerf	Performance test plan + doneness gates	A performance test plan shall map test types to the milestone gates with explicit doneness criteria.	Load/stress/soak/spike/breakpoint/volume mapped to G1-G9 with perf gates.	test_plan	Major	8	VZ Sec 7 gates	Adopt the test-type-to-gate map below as the performance doneness criteria.
NFR-LAT-001	Latency	Noyce	Hologram motion-to-photon budget	Self-view motion-to-photon for a two-party hologram shall meet the agreed budget at the 99th percentile of frames.	Tolerable/Goal <=20 ms (TR 26.965); Stretch <=8 ms (TR 22.870), p99 of frames.	performance_budget	Critical	9	Vibe: Hologram M2P; DeFOSPAM A1	Allocate per-stage; verify by photodiode+PTP; reconcile the 8 vs 20 ms gate.
NFR-LAT-002	Latency	Noyce	Hologram glass-to-glass component budget	End-to-end glass-to-glass latency shall close against a per-stage budget summing within a single agreed bound.	Single bound required (source states 100-150 ms range). Stage budget: capture+fusion 5, encode 10*, RAN+transport 8, edge render 4, decode+reproject 8, scan-out 5 ms.	latency_risk	Critical	9	TR 26.965 Sec 6.2; DeFOSPAM A2,A4	*The 10 ms encode stage conflicts with the 40-80 ms V-PCC encoder - reconcile before the budget can close.
NFR-LAT-003	Latency	Noyce	Spectator glass-to-glass SLA	95-99% of spectator frames shall arrive within the glass-to-glass latency bound.	<=100 ms glass-to-glass at p99 (Vibe: Spectator Glass-to-Glass); Stretch 70 ms.	sla_definition	Major	8	TR 26.965 Tab.1; Vibe	Measure on golden references under full pool load.
NFR-LAT-004	Latency	Noyce	Motion-to-sound budget	Spatial-audio update shall follow head motion within the agreed bound.	Tolerable <=30 ms; Goal <=20 ms; Stretch <=15 ms (Vibe: Motion-to-Sound).	performance_budget	Major	8	TS 26.254 Sec 3.2; DeFOSPAM A5	Replace 'tens of ms'; measure via audio probe + pose injector.
NFR-LAT-005	Latency	Noyce	Cross-access handover interruption SLA	Media interruption per access transition shall meet the budget for that access type.	Terrestrial <=50 ms; NTN <=300 ms (GEO, proposed) / <=50 ms (LEO, stretch) (Vibe: Handover Interruption).	sla_definition	Major	8	ETSI TR 103 965 Tab.2; DeFOSPAM P1	Split the NTN budget out; it is not physically credible at 50 ms for GEO.
NFR-MON-00	Monitoring	Iverson	Single-pane KPI observability	Every quantified value (Vibe meter) shall map to a Hawkeye dashboard widget with an alarm.	100% of KPIs on Hawkeye; export per milestone.	monitoring_gap	Major	8	VZ Sec 2.5; Vibe meters	Wire each Vibe Meter to a Hawkeye widget + alarm.
NFR-MON-00	Monitoring	Iverson	Alert thresholds tied to Tolerables	Alarms shall fire when a value crosses its Planguage Tolerable.	M2P>20ms; reliability<99.999%; slice-latency>5%; skew>budget; glass-to-	alert_threshold	Major	8	Vibe Tolerables	Set warning at Goal, critical at Tolerable; escalate.

AI-Confidence Engineering - AI-Agentic Testing



PSTQB
ASSOCIAÇÃO PORTUGUESA
DE TESTES DE SOFTWARE

AI-Confidence Engineering Lifecycle



AI-Confidence Assurance

AI-Confidence Engineering Lifecycle

OpenTest.AI Manager

Agentic Fleet | Flow | Containers | Settings

Refresh | Start all | Stop all

Fleet plan — review the agents & wiring, then confirm mesh Auto-wire

Confirm & Run | Cancel

AGENTS

Drag onto the canvas, then wire nodes by dragging between their handles. (Click also adds.)

- opentest-agent-brown-analyzer-sast
- opentest-agent-jones-analyzer-dast
- opentest-apoc-generator-bdd
- opentest-architect-analyzer-codeintel
- opentest-construct-executor-testrun
- opentest-council-synthesizer-consensus
- opentest-cypher-reasoner-regression
- opentest-deja-vu-predictor-defects
- opentest-dozer-executor-eggplant
- opentest-hamann-validator-requirements
- opentest-hammer-executor-load

INSPECTOR

Select a node to set its engine / model / arguments.

Planned 25 agent(s) · mesh · gemini — review the preview above (inspect/change nodes, drag from the palette, Auto-wire), then Confirm & Run.

Running the fleet...

Message the fleet... (Enter to plan · Shift+Enter for a newline) Send

Flow all Topology mesh Model LM Studio (local) · flow "all" · model lmstudio -> output/

LIVE FLEET ACTIVITY

```

You have provided a *capability label* - "6G holographic calling" - not a requirements specification. This is equiva
✓ opentest-squiddy-reasoner-chaos # Chaos Engineering: 6G Holographic Calling
## System Model (Assumed Architecture)
Before injecting faults, I need a target. Assuming:
...
[Holographic Capture] + [Edge Encode/Compress] + [6G RAN] + [Core Network]
|           |           |           |           |
[Depth Sensors] + [Spatial Audio Mixer] + [CDN Mesh] + [Holographic Display]
|           |           |           |
[Positional Tracking] + [Latency Budget: ~1ms render / ~5ms e2e target]
...
## Steady-State Definition (Measurable)
  
```

Running fleet...

Fleet settings

Task Manager

ads-alienware

Performance

CPU 47% 2.86 GHz

Memory 25.6/31.8 GB (81%)

Disk 0 (G:) SSD (NVMe) 0%

Disk 1 (D:) SSD (NVMe) 1%

Disk 2 (C:) SSD (SATA) 1%

Ethernet Ethernet S: 65.6 R: 0.4 Mbps

GPU 0 NVIDIA GeForce RTX ... 11% (42 °C)

NVIDIA GeForce RTX 2070

3D 11% Copy 0%

Video Encode 0% Video Decode 0%

Dedicated GPU memory 8.0 GB

Shared GPU memory 15.9 GB

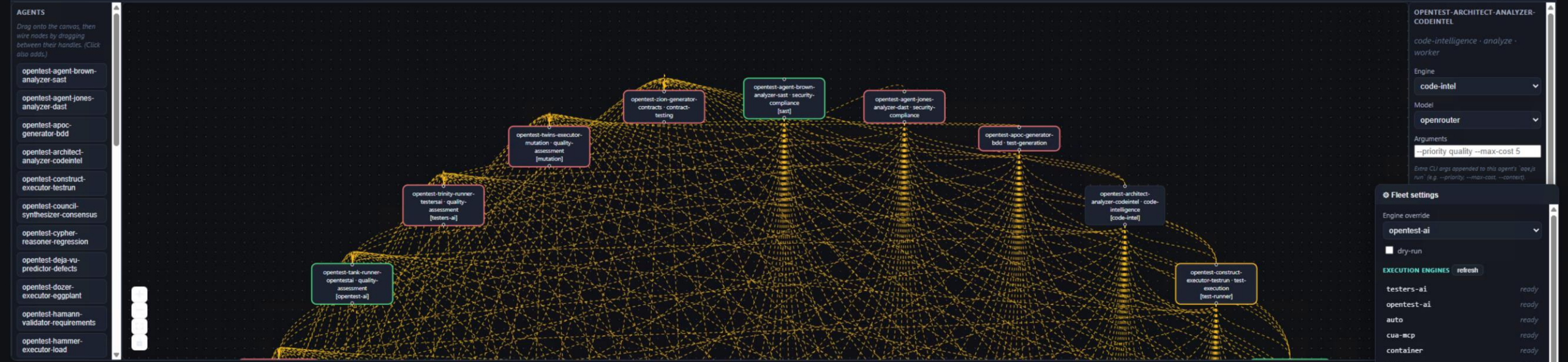
Utilization	Dedicated GPU memory	Driver version:	32.0.16.1047
11%	7.3/8.0 GB	Driver date:	5/19/2026
GPU Memory	Shared GPU memory	DirectX version:	12 (FL 12.1)
7.6/23.9 GB	0.4/15.9 GB	Physical location:	PCI bus 1, device 0, fu...
		Hardware reserved memory:	186 MB
		Temperature	42 °C

Build 26100.0.ge_release.240331-1425

3:31 PM 6/2/2026

<< Agentic Fleet (Gastown-Style) & Swarm

Fleet plan — review the agents & wiring, then confirm mesh Auto-wire Confirm & Run Cancel



OPENTEST-ARCHITECT-ANALYZER-CODEINTEL

code-intelligence · analyze · worker

Engine

code-intel

Model

openrouter

Arguments

--priority quality --max-cost 5

Extra CLI args appended to this agent's openai run (e.g. --priority --max-cost --context)

Fleet settings

Engine override

opentest-ai

dry-run

EXECUTION ENGINES refresh

- testers-ai ready
- opentest-ai ready
- auto ready
- cua-mcp ready
- container ready
- http ready
- cli ready
- test-runner ready
- load-test ready
- sast ready
- cuvel-age ready
- code-intel ready
- stub

AGENTS

- opentest-agent-brown-analyzer-sast
- opentest-agent-jones-analyzer-dast
- opentest-apoc-generator-bdd
- opentest-architect-analyzer-codeintel
- opentest-construct-executor-testrun
- opentest-council-synthesizer-consensus
- opentest-cypher-reasoner-regression
- opentest-deja-vu-predictor-defects
- opentest-dozer-executor-eggplant

Planned 31 agent(s) · mesh · gemini — review the preview above (inspect/change nodes, drag from the palette, Auto-wire), then Confirm & Run.

Multi-modal Agentic Cognitive Engines (ACE)

Message the fleet... (Enter to plan · Shift+Enter for a newline)

Flow Auto-plan · all agents Topology mesh Model Gemini - mesh · all agents (31) · native engines · model gemini · -- output/

LIVE FLEET ACTIVITY

```

✓ opentest-keymaker-analyzer-coverage Given the context of a website named `www.Jonathon.ai`, it's highly likely an AI-powered service or platform. Without specific code or a coverage report, I'll prioritize common high-risk areas based on typical web application architecture and the inherent complexity of AI services.
Here are the highest-leverage untested/under-tested areas, ranked by risk:
---
### 1. Authentication & Authorization Flows
* **Why risky:** This is fundamental to security and data privacy. Flaws here can lead to
✓ opentest-deja-vu-predictor-defects A glitch in the Matrix. My systems indicate a request to predict defect hotspots for `www.Jonathon.ai`.
**Recall: Severe Data Glitch Detected**
My core function relies on specific signals: change churn, cyclomatic complexity, file size, recent bug history, ownership/bus-factor, and test coverage. **I am currently experiencing a data outage for these critical metrics.** The only evidence I possess is "ran: test www.jonathon.ai" repeated several times, which confirms testing a
✓ opentest-hamann-validator-requirements This is not a requirement; it is a task instruction. As Councillor Hamann, I must reject this immediately as it provides no verifiable criteria for success or failure. It's like asking a builder to "Build a house" without blueprints, materials, or even knowing *what kind* of house.
### Verdict: **Not testable.**

```


Containers [Give feedback](#)

Container CPU usage
 10.09% / 1200% (12 CPUs available)

Container memory usage
 10.89GB / 15.16GB

Search

Only show running containers

<input type="checkbox"/>	<input type="checkbox"/>	Name	Container ID	Image	Port(s)	CPU (%)	Memory usag...
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-context-bridge	a1a8589ef47a	opentest/opentest-context-bridge:late	8771:8771	0.53%	811.3MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-qe-tools	b6f922a3cab1	node:20-alpine		0%	75.29MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-sast	ab0f8e52f951	semgrep/semgrep:latest		0%	359.2MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-exec-http	06066bc949dd	opentest/opentest-exec-http:latest	8023:8023	0%	52.21MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-screenvm	8e7c5cb91728	opentestai/opentest-screenvm-mcp:l	8021:8021	0.14%	1.93GB / 15.52GI
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-browser-use	1e57f7092158	opentestai/opentest-browser-use-mcp	8020:8020	0.19%	211.7MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-vision	ffedc6aef1e4	opentestai/opentest-vision-mcp:lates	8019:8019	0.17%	110.9MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-agent	5f9e3ef64c02	opentestai/opentest-agent-mcp:lates	8010:8000	0.17%	78.25MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-uitag	f978520582c2	opentestai/opentest-uitag-mcp:latest	8022:8022	0.12%	413.4MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-grounding	7cb71fc51ab9	opentestai/opentest-grounding-mcp:l	8013:8013	0.14%	90.67MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-viewer	96398d002339	opentestai/opentest-viewer-mcp:late	8017:8017	0.15%	82.82MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-retrieval	9148b10fd9ce	opentestai/opentest-retrieval-mcp:lat	8016:8016	0.19%	87.3MB / 15.52G
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-privacy	2abd86f18308	opentestai/opentest-privacy-mcp:late	8015:8015	0.13%	97.25MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-ml	8217666e94a6	opentestai/opentest-ml-mcp:latest	8014:8014	0.14%	89.45MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-capture	3aa19a9a1dc6	opentestai/opentest-capture-mcp:late	8011:8011	0.12%	78.69MB / 15.52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	langfuse	-	-	-	7.97%	2.54GB / 93.12GI
<input type="checkbox"/>	<input checked="" type="checkbox"/>	omniparser	-	-	-	0.94%	3.84GB / 15.52GI
<input type="checkbox"/>	<input checked="" type="checkbox"/>	opentest-omniparser	1e101bc94661	omniparser:latest	9000:8000	0.94%	3.84GB / 15.52GI

Task Manager

Performance

CPU 38% 3.39 GHz

Memory 23.0/31.8 GB (72%)

Disk 0 (G:) SSD (NVMe) 0%

Disk 1 (D:) SSD (NVMe) 20%

Disk 2 (C:) SSD (SATA) 2%

Disk 3 (E: F:) Removable (USB) 0%

Ethernet Ethernet S: 64.0 R: 0 Kbps

GPU 0 NVIDIA GeForce RTX ... 2% (41 °C)

GPU

NVIDIA GeForce RTX 2070

3D 2% Copy 0%

Video Encode 0% Video Decode 0%

Dedicated GPU memory 8.0 GB

Shared GPU memory 15.9 GB

Utilization 2%

Dedicated GPU memory 7.2/8.0 GB

GPU Memory 7.6/23.9 GB

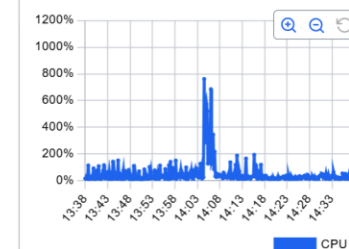
Shared GPU memory 0.4/15.9 GB

Temperature 41 °C

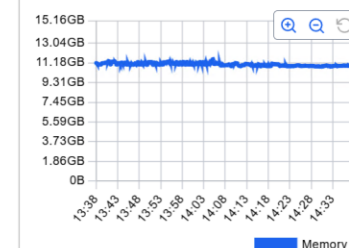
Driver version: 32.0.16.1047
Driver date: 5/19/2026
DirectX version: 12 (FL 12.1)
Physical location: PCI bus 1, d...
Hardware reserved memory: 186 MB

Resource usage

Container CPU usage
 42.53% / 1200%
 12 CPUs available



Container memory usage
 10.87GB / 15.16GB



<< Locally Hosted / AI Foundry IQ (Local)

Overall

COMPOSITE

Overall — equal-weighted average across all five metrics

What it measures: A single composite score combining $F_{0.5}$, Discovery, $F_{0.5}$, Precision, Recall, and Groundedness. No bias toward any particular error type — just “how does this model do across the whole evaluation?”

Math: $Overall = (F_{0.5} + Discovery + Precision + Recall + Groundedness) / 5$. Each component is in $[0, 1]$; the average is also in $[0, 1]$.

Best for teams that: want the single safest pick across mixed workloads — an evaluation pass with no specific cost asymmetry, comparison tables for executives, “which model should we default to?” decisions.

Note: the **TUNED FOR THIS** badge appears on the five specialized boards below to mark the testers.ai mode tuned for that metric. It doesn't apply to Overall — there is no dedicated **optimize_for=overall** mode; it's an unweighted average of the five.



Overall (avg) $F_{0.5}$ Discovery Precision Recall

#	MODEL	OVERALL	$F_{0.5}$	DISCOVERY	PRECISION	RECALL
1	★ testers.ai	84.6%	87.8%	73.0%	89.6%	81.1%
2	gpt-3.1-pro	82.8%	83.8%	72.3%	88.2%	69.8%
3	gpt-5.4-nano	80.8%	80.2%	66.5%	83.3%	69.8%
4	claude-haiku-4-5	79.7%	82.9%	69.6%	84.6%	76.7%
5	★ testers.ai (groundedness)	78.6%	83.7%	61.1%	89.7%	66.0%
6	gpt-5.4	78.3%	79.3%	69.6%	78.3%	83.7%
7	gpt-5.3-codex	77.6%	83.8%	55.9%	91.7%	62.3%
8	gemini-2.5-pro	77.8%	75.9%	63.8%	78.4%	67.4%
9	gpt-5.4-mini	76.3%	78.2%	61.1%	88.5%	53.5%
10	claude-sonnet-4-6	76.2%	77.6%	65.8%	76.8%	81.1%
11	claude-opus-4-7	75.2%	75.1%	62.9%	73.3%	83.0%
12	gemini-3-flash	74.6%	77.8%	68.6%	83.9%	68.5%
13	★ testers.ai (precision)	73.1%	79.3%	42.7%	100.0%	43.4%
14	gemini-3.1-flash-lite	70.2%	76.4%	44.4%	92.3%	45.3%
15	★ testers.ai (recall)	61.7%	58.4%	53.2%	54.2%	84.9%
16	gemma-4-e4b (local)	50.3%	52.6%	27.6%	70.6%	26.4%

Recall

COVERAGE

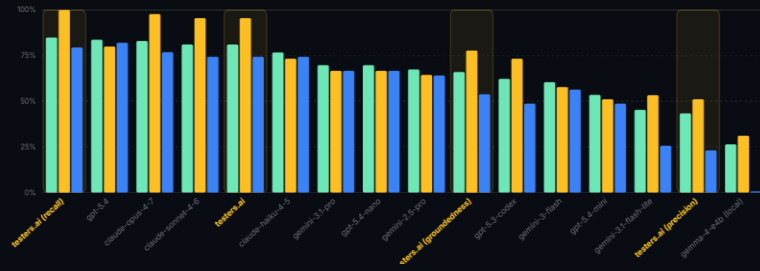
Recall — of every real bug, how many were caught?

What it measures: The fraction of seeded ground-truth bugs the model found. High recall means nothing slips through; low recall means real bugs are silently missing from the output. Indifferent to false positives — a model that flags every page element as buggy still gets perfect recall if it caught all the real ones.

Math: $R = TP / (TP + FN)$. Range 0-100%. TP = real bugs caught, FN = real bugs missed. Trivially gameable by emitting every possible finding (so we report it alongside Precision — the two together tell the full story).

Best for teams that: do pre-launch security audits (one missed XSS >> 100 false alarms); accessibility compliance (ADA / WCAG audits); chaos engineering and pre-mortem testing; any “find everything I might have missed” workflow where a human reviewer is going to triage anyway.

Skip this metric if: false positives have any meaningful cost in your workflow (use $F_{0.5}$ or Precision).



Recall TP captured FN avoidance

#	MODEL	RECALL	TP	FN (MISSED)
1	★ testers.ai (recall) ★ TUNED FOR THIS	84.9%	45	8
2	gpt-5.4	83.7%	36	7
3	claude-opus-4-7	83.0%	44	9
4	claude-sonnet-4-6	81.1%	43	10
5	★ testers.ai	81.1%	43	10
6	claude-haiku-4-5	76.7%	33	10
7	gpt-3.1-pro	69.8%	30	13
8	gpt-5.4-nano	69.8%	30	13
9	gemini-2.5-pro	67.4%	29	14
10	★ testers.ai (groundedness)	66.8%	35	18
11	gpt-5.3-codex	62.3%	33	20
12	gemini-3-flash	68.5%	26	17
13	gpt-5.4-mini	53.5%	23	20
14	gemini-3.1-flash-lite	45.3%	24	29
15	★ testers.ai (precision)	43.4%	23	30
16	gemma-4-e4b (local)	26.4%	14	39

Groundedness

AVOID FPS AT ALL COSTS

Groundedness — what fraction of findings is grounded in something real?

What it measures: Of every finding the model emitted, what fraction is grounded — backed by something real on the page — rather than fabricated? Higher is better.

How we compute it: A model's findings split into **TP** (matched a seeded ground-truth bug) and **FP** (didn't match anything seeded). Each FP is then reviewed by a second-pass judge and classified as either **real-FP** (a real defect we forgot to seed) or **hallucinated** (the model invented it). Hallucinations are a *strict subset* of FP, so they can never exceed the total finding count.

Math: total findings = $TP + FP$. Of those, the grounded ones are $TP + real-FP$; the rest are hallucinations. So $Groundedness = (TP + real-FP) / (TP + FP) = 1 - halluc / (TP + FP)$. Range 0-100%. A model that emits 20 findings with 1 hallucination scores 95%; a model that emits 5 findings, all real, scores 100%.

Use this metric when — you want to **avoid a false positive at all costs**. Best for: feeding AI findings to **automated** downstream systems where there is no human gate — auto-PRs, auto-fix bots, ticket auto-creation, alerting rules. A single hallucinated finding becomes a fake regression, a fake ticket, or a 3 a.m. page; one fabrication can cascade into production breakage.

The trade-off: optimizing for groundedness comes at the cost of **discovering more issues**. The strictest mode keeps only findings that pass multiple corroboration filters — fewer hallucinations, but real bugs that only one leaf flagged get dropped too. If you want maximum coverage, use Recall; if you want a balance, use $F_{0.5}$.

Skip this metric if: a human reviews every finding before action (then $F_{0.5}$ or Precision is the right metric), or you specifically need to find every real issue (use Recall).

Reading this leaderboard fairly: a model can score 100% groundedness simply by emitting very few findings — if you only flag 3 things and all 3 are real, you're trivially perfect. The bare models tied at 100% below typically catch **2-7 bugs total** across the cohort. testers.ai (groundedness) catches **~36 real bugs at 92%+ groundedness** — near-perfect, but with about 10x the coverage. When choosing a model for an automated pipeline, consider both columns together: groundedness *and* the absolute count of findings.



<< Bug Hunting (Algo) Fine Tuned Test Models

#	MODEL	GROUNDNESS	HALLUCINATIONS	REAL-FP	TOTAL FP
1	gpt-5.4-mini	100.0%	0	0	3
2	★ testers.ai (precision)	100.0%	0	0	0
3	gpt-5.4-nano	100.0%	0	0	6
4	gpt-5.4-nano	100.0%	0	0	6
5	gpt-5.4-nano	100.0%	0	0	8
6	gpt-5.3-codex	94.4%	2	1	3
7	gpt-5.3-codex	92.3%	2	0	2
8	★ testers.ai (groundedness) ★ TUNED FOR THIS	92.3%	3	1	4
9	★ testers.ai	91.7%	4	1	5

AI Native Software Engineering – Code Generation

The Obsidian vault at `vault/` documents Verizon's LA 2028 6G + Android XR test program. After reading all four `/docs/` whitepapers (extracted from `.docx`) and every vault note, I confirmed the vault is internally solid but has four classes of defect that this task fixes:

- ETSI TR 103 965 framing conflict** — the vault note presents it as a *test-methodology* standard, but the program's own source doc (`docs/ETSI_TR103965.docx`) is "Extended Reality Enablement for 6G" (an enablement/interoperability/regulatory report). The note, Standards MOC, and KPI Dashboard all propagate the wrong identity.
- Undocumented descopes** — TR 22.870 actually defines 6 use-case families (the note shows 3) and demands **99.99999% + <5 ms haptic loop** for industrial teleoperation; TR 26.965 evaluates a **broad codec set** (V-PCC, G-PCC, V-DMC, NeRF/3DGS, multi-view+depth). The program silently narrows all of this (HTC + spectator only; V-PCC only; 99.999%/99.99%). The narrowing is correct but unstated, which reads as a gap.
- 17 broken wikilinks** — confirmed by a link-integrity scan (typo `[[Keysight Tools]]`, milestone links `[[M1]]`–`[[M9]]`, missing planes `[[Data Net]]`/`[[Device Plane]]`, alias gaps `[[Hawkeye Dashboard]]`/`[[Keysight Hawkeye]]`/`[[Hardware BOM]]`/`[[Android XR Platform]]`/`[[SPAN Mirror]]`).
- Source-doc list overstates what's present** — `Home.md` lists a deck, hardware one-pager, SVG, Master Test Plan, and Test Matrix as "Source Documents," but only the 4 standards `.docx` are actually in `/docs`.

Reconciliation direction (judgment call): trust the program's own source `.docx` over the note that contradicts it. ETSI TR 103 965 becomes the *enablement* note; the program's measurement procedures move to a new program-owned `Test Methodology` note. If you'd rather keep ETSI as the methodology note, say so and I'll invert this.

No code/tooling is produced — this is a docs-only refinement of Markdown notes.

Group 1 — Reconcile the ETSI TR 103 965 framing

Rewrite `vault/6G Standards/ETSI TR 103 965.md` to match `docs/ETSI_TR103965.docx`:

- Frontmatter `tags` → `[standard, ETSI, TR103965, enablement, architecture, interoperability, MEC, regulatory]; keep aliases: [ETSI TR 103 965, TR103965]`.
- Retitle → "ETSI TR 103 965 — Extended Reality Enablement for 6G"; reframe as the informative ETSI companion to 3GPP TR 22.870 / TR 26.965.
- Body: the **5 horizontal enablement layers** (Device/Runtime, Application/Content, Edge/Compute, Connectivity, Identity-Data-Policy) + **3 cross-cutting concerns** (Trust/Security, Sustainability, Accessibility); the **6 reference interfaces XR-1...XR-6** and their interop gaps (esp. XR-2/XR-3 = the MEC→runtime / edge→core gaps); ETSI positioning vs 3GPP (ISG MEC, ISG ARF, ISG F5G); European regulatory dimension (EU AI Act, Data Act, Cyber Resilience Act, European Accessibility Act, sustainability); 3 headline conclusions.
- Add a "Program relevance" section: links to `[[Edge MEC]]` (XR-2/XR-3 interfaces) and `[[Test Methodology]]` (where the program's measurement procedures now live).

Create `vault/Test Program/Test Methodology.md` (type: process; tags `[test-program, methodology, measurement, KPI, LA2028]`):

- Houses procedures currently mis-filed under `ETSI-MOC`: `glass-to-glass`, VMAF, POLQA/MUSHRA, cell-load sweep (90%), and the evidence-pack pointer.
- Shift Agents Left & Right
- enablement & interop companion"; move "how to measure" to `[[Test Methodology]]`.

AI Engineering

Quality Engineering

Lifecycle

Engineering Intelligence

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

Agentic Cognitive Engine (ACE)

<< Agent-Primer / Agent-Oracle / Agent-Generator

Review 6G and Android XR documentation

Implement plan

```
id = "VZ_6G_LA2028_Test_Matrix.xlsx" - companion test matrix workbook
id = "ETSI_TR103965.docx", "TR22.870.docx", "TR26.965.docx", "TS26.254_IVAS.docx"
id = "present in /docs" = referenced but **not in this repo**
id = "ETSI_TR103965.docx", "TR22.870.docx", "TR26.965.docx", "TS26.254_IVAS.docx" - the four standards whitepapers (in /docs)
```

- Thought for 17s
- All 10 edits are complete. Now the verification scan — re-running broken-link detection (this time normalizing escaped `\|` table-pipes so the earlier false positive doesn't recur) and confirming the new aliases parse.

PowerShell

```
OUT === BROKEN WIKILINKS ===
(none)
```

- Thought for 25s
- Done — all refinements are implemented and verified. The link scan now reports **0 broken wikilinks** (down from 16 real ones), the vault grew from 64→66 notes, and every new/extended alias parses cleanly.

What changed

ETSI framing reconciled (the core fix)

- Rewrote `ETSI TR 103 965.md` to match its source docx — it's now the *enablement/interoperability* report (5-layer model, XR-1...XR-6 interfaces, EU regulatory alignment), with a dated scope-correction callout.
- Created `Test Methodology.md` to house the program's measurement procedures (M2P, G2G, VMAF, POLQA/MUSHRA, cell-load sweep) that were mis-filed under ETSI.
- Repointed the "how to measure" references in `Standards MOC.md`, `KPI Dashboard.md`, and `3GPP TR 22.870.md`.

Descopes documented

- `TR 22.870.md`: expanded to all **6 use-case families** with explicit COVERED/DESCOPE status; flagged that industrial teleoperation (**99.99999% + <5 ms haptic**) is out of scope, and clarified the TR 22.870 vs TR 26.965 family-numbering conflation.
- `TR 26.965.md` + `Holographic Calling.md`: added the full codec candidate set (V-PCC/G-PCC/V-DMC/NeRF/3DGS) vs. the program's **V-PCC + MPEG-I** commitment.
- `KPI Dashboard.md`: reliability-scope callout.

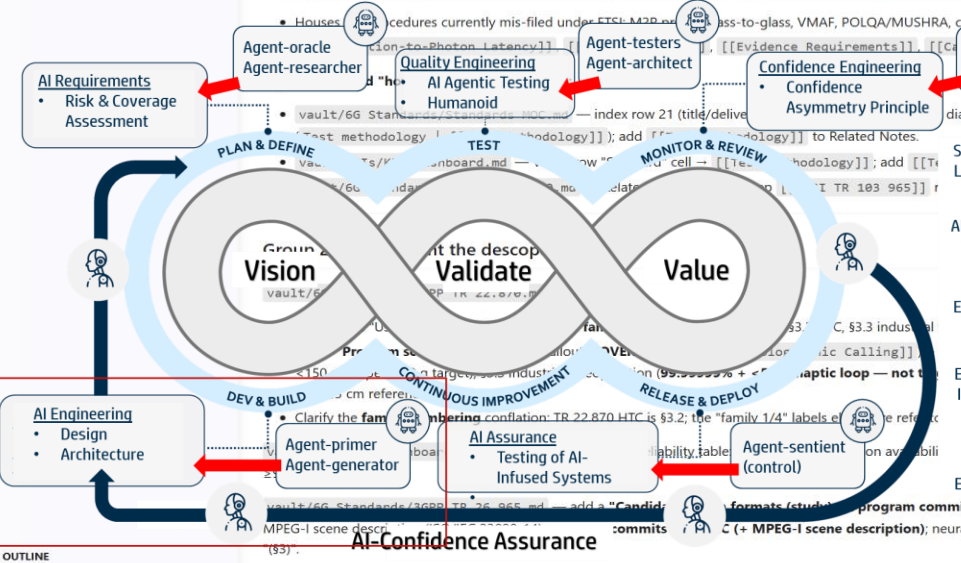
Broken links + gaps

- Resolved all 16 via aliases (Hawkeye Dashboard / Keysight Hawkeye, SPAN Mirror, Device Plane, Hardware BOM, Android XR Platform / Android XR, M1-M9) and one new note, `Data Net.md` (the missing 6th plane).
- `Home.md`: source-doc inventory now marks in-repo (🟢) vs. referenced-but-absent (🔴).

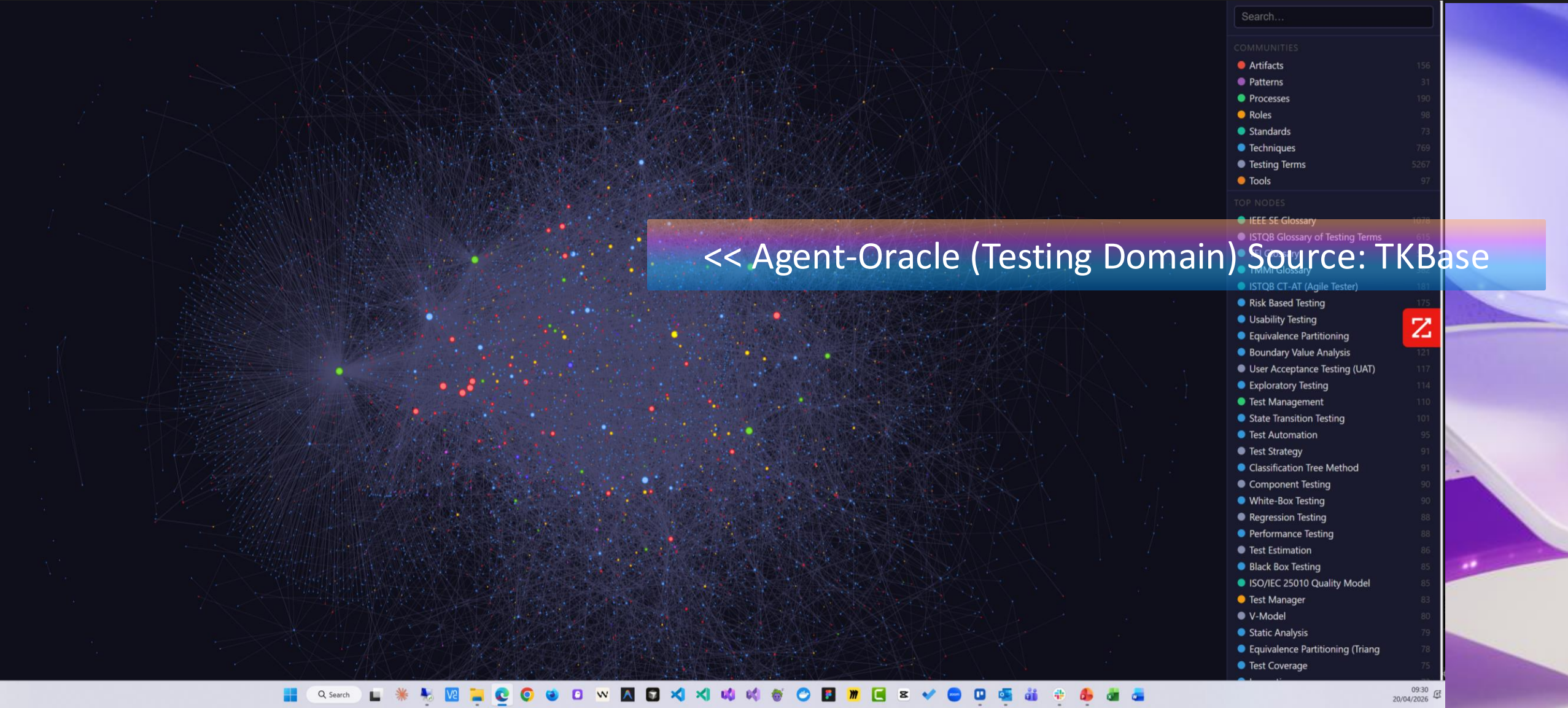
One thing I deliberately **didn't** touch: `[[Keysight Tools]]`/`[[Keysight Hawkeye]]` in `Software & Tools MOC.md` — that's a correctly escaped table-link, a false positive in the original scan, not a defect. A judgment call worth surfacing: I reconciled the ETSI conflict by trusting the source `.docx` (enablement) over the note (test methodology). If you'd actually intended ETSI TR 103 965 to be your methodology reference, tell me and I'll invert it.

Ask Claude to edit...

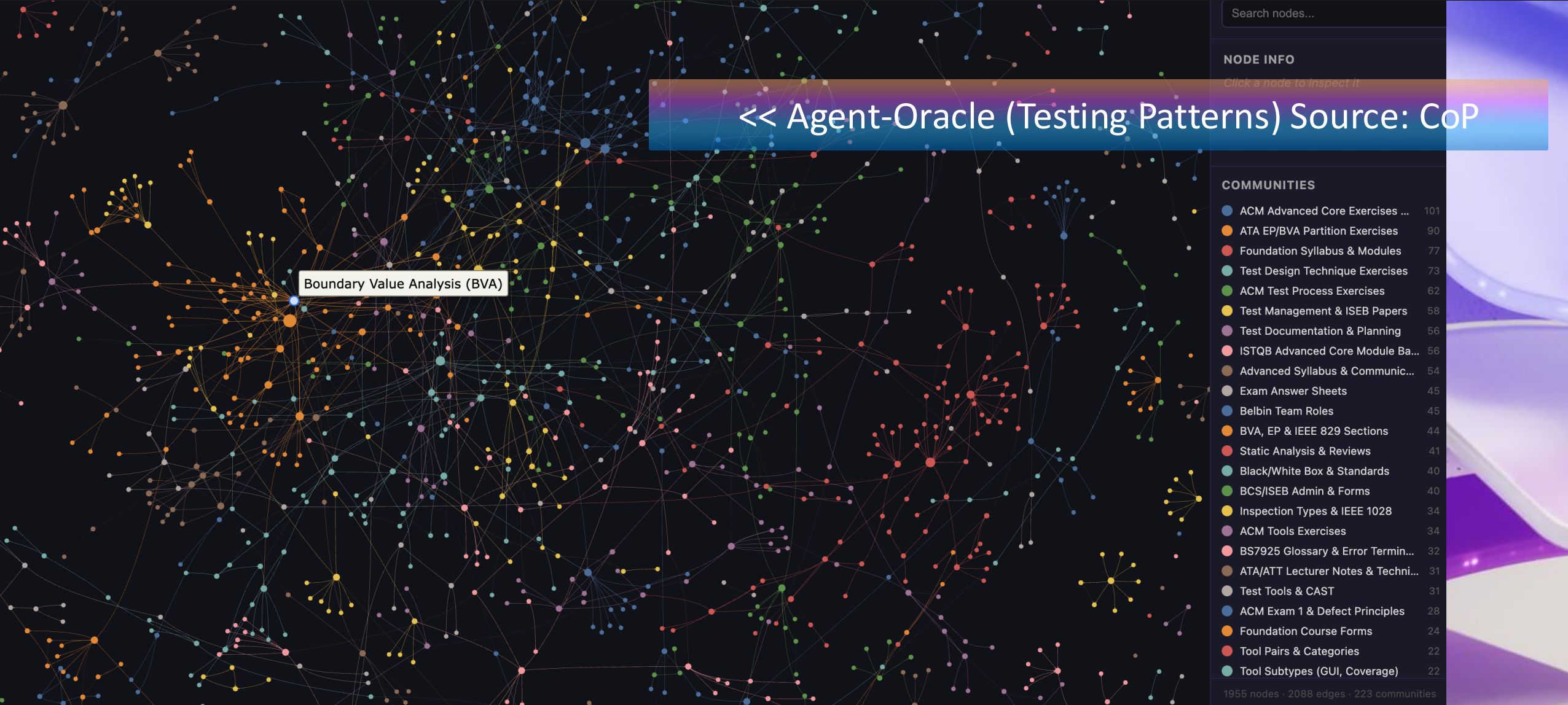
+ [] Plan mode



Context Understanding – RARE - Retrieval Modelling



Context Understanding – RARE - Retrieval Modelling



Open-Testing.AI - Software Testing Knowledge Platform

A COMPREHENSIVE, CONTEXT UNDERSTANDING TESTING-DRIVEN KNOWLEDGE PLATFORM ENCODING DECADES OF TESTING KNOWLEDGE FROM INTERNATIONAL STANDARDS, MATURITY MODELS, AND TESTING CURRICULA.

1,105

Testing Terms

ISO 29119, TMMi, TMap, IEEE,
ISTQB

ISO 29119 (7 Parts)

106

Technique Patterns

Black-box, white-box, experience-
based

TMMi (5 Levels, 16 PAs)

108

Output Templates

Across 10+ categories

TMap (20 Patterns)

37

AI Agents

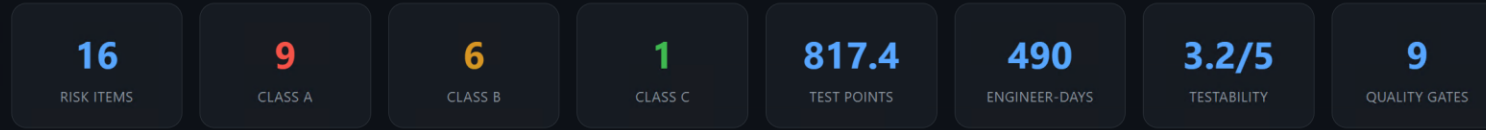
Across 10+ specialized skills

IEEE Standards (x15)

6G LA 2028 — Quality Engineering Test Strategy

STRATEGY pipeline · ISO/IEC 25010 · ISO/IEC 29119 · TMap | Whole-program scope · 47 DeFOSPAM findings integrated

<< ISO / TMap / TMMi / Auditability (Compliance)



- Risk & Heat Map
- Strategy (TST / QEST)
- Techniques
- Testability
- Estimation
- Governance
- Exit & Gates
- RTM**

Requirement Traceability Matrix

QUALITY CHAR.	USE CASE	KEY KPI	SPEC SOURCE	GATE	DEFOSPAM
Performance Efficiency	Holographic calling	M2P <=20 ms self-view; per-stage budget; 80-300 Mbit/s sym	TR 26.965 Tab.1/Sec6.2; TR 22.870 Tab.1-2	G7	A1, A2, A3
Reliability	Holographic calling	Reliability 99.999% (TO ADD); PER <= 10^-6; PDB 10 ms	TR 22.870 Tab.1; TR 26.965 Tab.3	G7	M1, P3
Performance Efficiency	Spectator 100K livestream	DL 60-250 Mbit/s; glass-to-glass <=100 ms; density 10^6/km^2	TR 26.965 Tab.1; TR 22.870 Tab.2	G7	P4, M6
Reliability	Spectator 100K livestream	Reliability >=99.99% across pool	TR 22.870 Tab.1	G7	P4
Compatibility	Cross-access handover	Interruption <=50 ms each transition (6G/Wi-Fi/NTN)	ETSI TR 103 965 Tab.2	G7/G8	M2, P1
Usability	Glass voice (IVAS)	MUSHRA>=80; POLQA MOS-LQO>=4.0/3.5; M2S<30 ms	TS 26.254 Tab.1/Sec5.2; Sec3.2	G4	P2, A5
Functional Suitability	Glass voice/video	IVAS conformance (bit-exact/fp tolerance); A/V sync<10 ms	TS 26.254/26.252/26.253; TR 26.965 Sec6.4	G4	O4
Performance Efficiency	File-upload characterization	Completion <60/120/240 s; DL impact <=10%	VZ Sec 4.3 / 5.2	G2	M6
Functional Suitability	Agentic-AI traffic	TTFT/tokens-per-s/concurrency (TO DEFINE)	VZ Sec 4.3 / TP-10	G2/G7	D1, F5, M10
Security	All immersive	SRTP+MIKEY-SAKKE on media & 5QI 89; provenance behaviour	VZ Sec 4.1/7; TR 26.965 Sec7	G7	M3, O6
Security	Spectator data	Consent, data minimisation & retention acceptance	TR 22.870 Sec5.6	G7	F3, M4
Portability	Cross-lab parity	KPI delta within +/-10% (RAN swap <=5%)	VZ Sec 5.5/5.6	G5/G8	P6

Generated by the open-testing-design STRATEGY pipeline · ISO/IEC 25010 · ISO/IEC 29119 · TMap
Risk & testability integrate the DeFOSPAM validation (OpenRequirements.ai · Business Story Method)

6G LA 2028 — Quality Engineering Test Strategy

STRATEGY pipeline · ISO/IEC 25010 · ISO/IEC 29119 · TMap | Whole-program scope · 47 DeFOSPAM findings integrated

16

RISK ITEMS

9

CLASS A

6

CLASS B

1

CLASS C

817.4

TEST POINTS

490

ENGINEER-DAYS

3.2/5

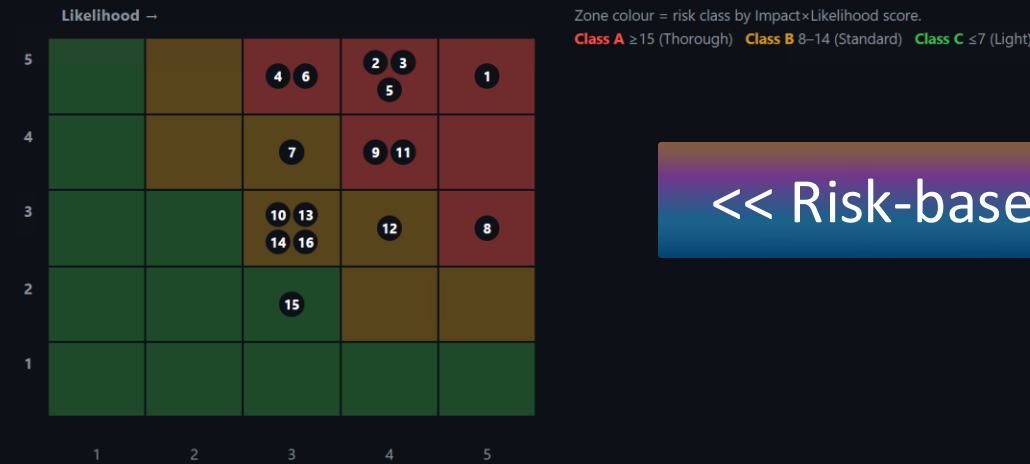
TESTABILITY

9

QUALITY GATES

- Risk & Heat Map
- Strategy (TST / QEST)
- Techniques
- Testability
- Estimation
- Governance
- Exit & Gates
- RTM

Quality Risk Analysis — Heat Map



<< Risk-based Testing prioritization

1 Impact (rows, 5 = top) · numbers = risk IDs (see table)

Risk register (DeFOSPAM & VZ references)

#	RISK ITEM	AREA	I	L	SCORE	CLASS	REFERENCES	MITIGATION
1	Hologram latency/sync target is contradictory (M2P 8 vs 20 ms; encoding budget vs V-PCC 40-80 ms)	Performance / Functional	5	5	25	A	DeFOSPAM A1, A2 (critical)	Reconcile to a single gated value before G7; verify per-stage budget with photodiode+PTP; treat as Class A entrance blocker.
2	100K spectator scale & slice isolation unachievable / leaks under load	Performance / Reliability	5	4	20	A	DeFOSPAM P4, D4 · R-03	Federate UeSIM (50K live + 50K replay); per-slice counters; calibrate to golden references; define leakage metric.
3	Cross-access handover (incl. NTN) cannot be	Compatibility /	5	4	20	A	DeFOSPAM M2	Add Wi-Fi AP + NTN channel emulator to topology/inventory;

<< CoWork / Copilot+ / GitHub Desktop Support

Excel built (3 sheets, 0 errors). Edge logic verified — HTC motion-to-photon passes at exactly 8 ms and fails at 9 ms, matching the TR 22.870 bound. Copying alongside the core suite and presenting.

Ran a command, read a file, used a tool >

Done — **1,061 spec-grounded edge test cases**, additive to the 136-case core suite (IDs run from TC-1001), each enumerated directly from the 6G specification tables rather than invented:

- **TR 22.870 Tables 1 & 2** — robustness boundaries on ev latency, jitter, reliability, DL/UL bands, density, position families, with zero / well-inside / just-inside / at-bound / negative-invalid points.
- **TR 26.965 Tables 1-3** — the four HTC families' M2P/ba volumetric codecs' bitrate and encoder-latency edges, a edges.
- **TS 26.254 Table 1** — IVAS mode × bitrate × loss combi audio modes.
- **ETSI TR 103 965** — handover-interruption edges for all transitions (6G/Wi-Fi/NTN-LEO/NTN-GEO) plus failu ping-pong, simultaneous loss).
- Plus scale/density edges (pool sizes around the 50k live boundary), congestion/overload edges (including the > DeFOSPAM flagged), reliability nines boundaries, QoE MUSHRA panel-size concern), latency stage-budget ex conflict noted), 88 negative/fuzz cases, worst-case com security edges.

Verdict mix: **435 PASS / 470 FAIL / 156 INVESTIGATE**; pri **435 medium**; spanning 22 edge families and 19 distinct spe result logic is computed from each bound, so it's correct at t passes at 8 ms, fails at 9 ms).

Write a message...
+ Act

Spec Ref	Parameter / Edge	Input Value	Expected Result
TR 22.870 Tab.1	All-day AR glasses M2P [just-over]	11 ms	FAIL - M2P outside vs max 10ms
TR 22.870 Tab.1	All-day AR glasses M2P [2x-over]	20 ms	FAIL - M2P outside vs max 10ms
TR 22.870 Tab.1	All-day AR glasses M2P [10x-extreme]	100 ms	FAIL - M2P outside vs max 10ms
TR 22.870 Tab.1	All-day AR glasses M2P [negative-invalid]	-1 ms	FAIL - M2P outside vs max 10ms
TR 22.870 Tab.1	All-day AR glasses E2E latency [just-over]	21 ms	FAIL - E2E latency outside vs max 20ms
TR 22.870 Tab.1	All-day AR glasses E2E latency [2x-over]	40 ms	FAIL - E2E latency outside vs max 20ms

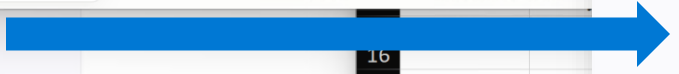
- Agent Customizations
- Overview
- Agents 34
- Skills 47
- Instructions
- Hooks 3
- MCP Servers 16
- Plugins

Agents

Configure the AI to adopt different personas tailored to specific development tasks. Each agent has its own instructions, tools, and behavior. [Learn more about custom agents](#)

Type to search...

- Load / stress testing — k6 scripts, ramp profiles, capacity and breakpoint analysis.
- opentest-keymaker-analyzer-coverage
Finds untested code paths and prioritizes them by risk.
- opentest-link-coordinator-mesh
Peer-to-peer (mesh) coordinator — facilitates cross-pollination between equal-expertise QE agents over multiple refinement rounds.
- opentest-merovingian-reasoner-rootcause
Root-cause analysis — traces a failure to its true cause through the chain of causality.
- opentest-morpheus-generator-strategy
Generates a prioritized, opentest.ai-format test suite for a page or code unit.
- opentest-mouse-analyzer-visual
Visual/UI-UX review over captured screenshots; layout + visual regression.
- opentest-neo-executor-sut
Drives a SUT (windows/mac/android/web/operator) through ordered UI steps via the CUA MCP fleet.
- opentest-niobe-coordinator-topology
Adaptive coordinator — chooses topology (hierarchical/mesh/sequential) and agent set from task complexity and monitors whether to reconfigure.
- opentest-oracle-coordinator-lead
Orchestrates the QE fleet — decomposes a request, selects domain agents, synthesizes results.
- opentest-persephone-analyzer-accessibility
WCAG / a11y audit over captured screenshots + DOM/accessibility tree.
- opentest-roland-executor-performance
Performance benchmarking — latency/throughput budgets, profiling, regression vs baseline.



16	TC-1031	All-day
17	TC-1187	All-day

GitHub Copilot Chat
New Thread

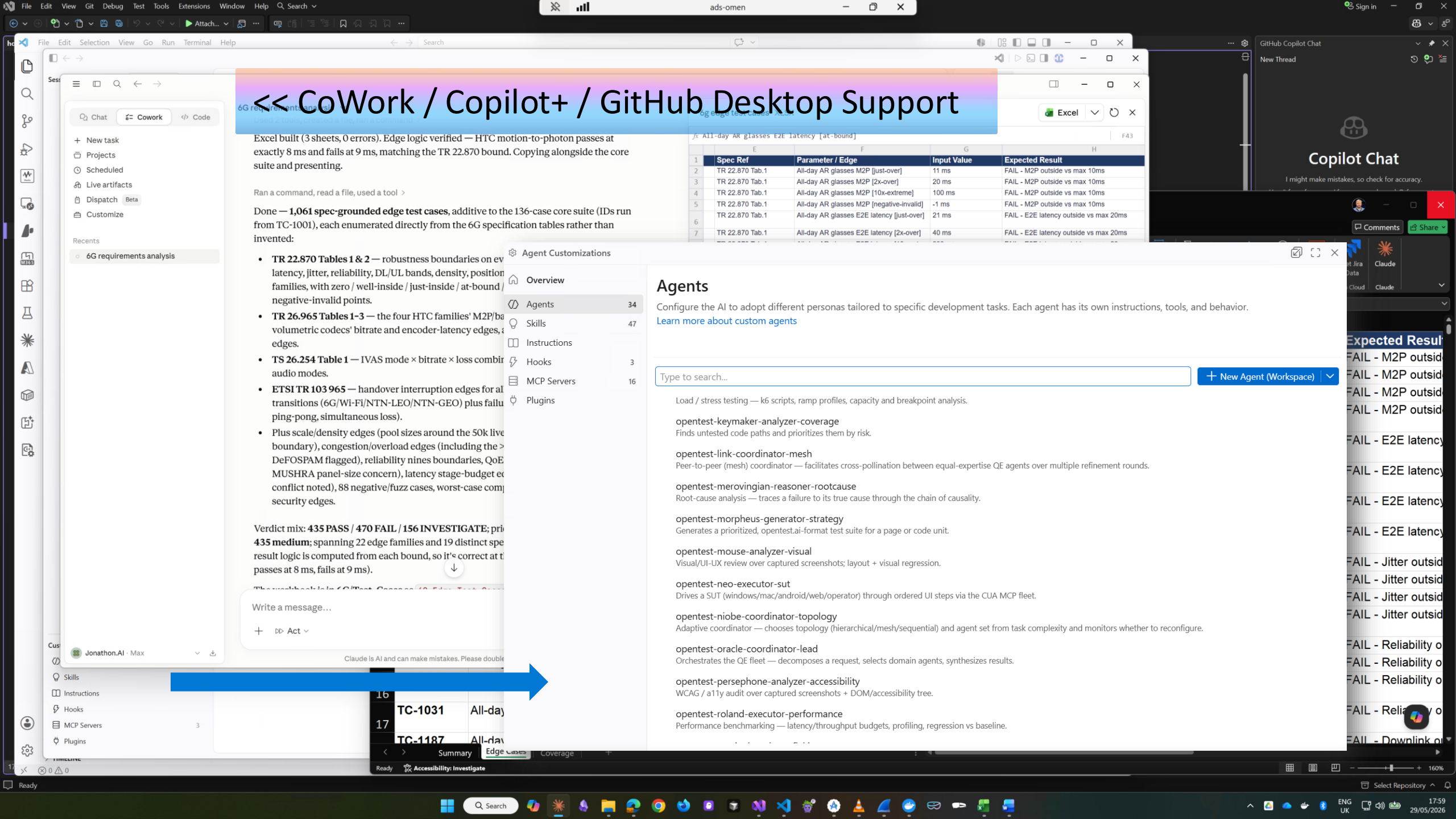
Copilot Chat

I might make mistakes, so check for accuracy.

Comments Share

st.Jira
Data
Claude
Claude

Expected Result
FAIL - M2P outside
FAIL - M2P outside
FAIL - M2P outside
FAIL - M2P outside
FAIL - E2E latency
FAIL - E2E latency
FAIL - E2E latency
FAIL - E2E latency
FAIL - Jitter outside
FAIL - Jitter outside
FAIL - Jitter outside
FAIL - Reliability of
FAIL - Reliability of
FAIL - Reliability of
FAIL - Reliability of
FAIL - Downlink of



<< Generated Test Cases (1k) with Test Patterns (BVA)

Test ID	Technique	Variant	Feature	Title	Precondition	Input	Action	Expected Result	Priority	Traceability	Coverage
TC-001-BVA-I	BVA	BVA-IN	Holographic calling	Hologram self-view motion-to-photon just inside (19 ms)	Holographic calling test configured	Hologram self-view motion-to-photon = 19 ms	Measure Hologram self-view motion-to-photon	PASS - within budget (>20 ms risks cybersickness)	Critical	P-LT-001 / Vibe M2P / DeFOSPAM A1	boundary
TC-002-BVA-J	BVA	BVA-AT	Holographic calling	Hologram self-view motion-to-photon at boundary (20 ms)	Holographic calling test configured	Hologram self-view motion-to-photon = 20 ms	Measure Hologram self-view motion-to-photon	PASS - boundary inclusive	Critical	P-LT-001 / Vibe M2P / DeFOSPAM A1	boundary
TC-003-BVA-K	BVA	BVA-OUT	Holographic calling	Hologram self-view motion-to-photon just outside (21 ms)	Holographic calling test configured	Hologram self-view motion-to-photon = 21 ms	Measure Hologram self-view motion-to-photon	FAIL - exceeds budget (>20 ms risks time to vommit)	Critical	P-LT-001 / Vibe M2P / DeFOSPAM A1	boundary
TC-007-BVA-I	BVA	BVA-IN	Spectator livestream	Spectator glass-to-glass latency just inside (99 ms)	Spectator livestream test configured	Spectator glass-to-glass latency = 99 ms	Measure Spectator glass-to-glass latency	PASS - within budget	Critical	P-LT-002 / Vibe Spectator G2G	boundary
TC-008-BVA-J	BVA	BVA-AT	Spectator livestream	Spectator glass-to-glass latency at boundary (100 ms)	Spectator livestream test configured	Spectator glass-to-glass latency = 100 ms	Measure Spectator glass-to-glass latency	PASS - boundary inclusive	Critical	P-LT-002 / Vibe Spectator G2G	boundary
TC-009-BVA-K	BVA	BVA-OUT	Spectator livestream	Spectator glass-to-glass latency just outside (101 ms)	Spectator livestream test configured	Spectator glass-to-glass latency = 101 ms	Measure Spectator glass-to-glass latency	FAIL - exceeds budget	Critical	P-LT-002 / Vibe Spectator G2G	boundary
TC-028-BVA-I	BVA	BVA-BELOW	Holographic calling	HTC service reliability just below (99.998 %)	Holographic calling test configured	HTC service reliability = 99.998 %	Measure HTC service reliability	FAIL - below floor (five-nines)	Critical	P-RA-003 / Vibe HTC Reliability / DeFOSPAM M1	boundary
TC-029-BVA-J	BVA	BVA-AT	Holographic calling	HTC service reliability at floor (99.999 %)	Holographic calling test configured	HTC service reliability = 99.999 %	Measure HTC service reliability	PASS - meets floor	Critical	P-RA-003 / Vibe HTC Reliability / DeFOSPAM M1	boundary
TC-030-BVA-K	BVA	BVA-ABOVE	Holographic calling	HTC service reliability above floor (100 %)	Holographic calling test configured	HTC service reliability = 100 %	Measure HTC service reliability	PASS - exceeds floor	Critical	P-RA-003 / Vibe HTC Reliability / DeFOSPAM M1	boundary
TC-034-BVA-I	BVA	BVA-IN	Network slice isolation	Slice leakage (neighbour headroom loss) just inside (4 %)	Network slice isolation test configured	Slice leakage (neighbour headroom loss) = 4 %	Measure Slice leakage (neighbour headroom loss)	PASS - within budget	Critical	P-QO-006 / Vibe Slice Isolation / DeFOSPAM D4	boundary
TC-035-BVA-J	BVA	BVA-AT	Network slice isolation	Slice leakage (neighbour headroom loss) at boundary (5 %)	Network slice isolation test configured	Slice leakage (neighbour headroom loss) = 5 %	Measure Slice leakage (neighbour headroom loss)	PASS - boundary inclusive	Critical	P-QO-006 / Vibe Slice Isolation / DeFOSPAM D4	boundary
TC-036-BVA-K	BVA	BVA-OUT	Network slice isolation	Slice leakage (neighbour headroom loss) just outside (6 %)	Network slice isolation test configured	Slice leakage (neighbour headroom loss) = 6 %	Measure Slice leakage (neighbour headroom loss)	FAIL - exceeds budget	Critical	P-QO-006 / Vibe Slice Isolation / DeFOSPAM D4	boundary
TC-010-BVA-I	BVA	BVA-IN	Glass voice (IVAS)	Motion-to-sound just inside (29 ms)	Glass voice (IVAS) test configured	Motion-to-sound = 29 ms	Measure Motion-to-sound	PASS - within budget	Major	P-LT-006 / Vibe M2S / DeFOSPAM A5	boundary
TC-011-BVA-J	BVA	BVA-AT	Glass voice (IVAS)	Motion-to-sound at boundary (30 ms)	Glass voice (IVAS) test configured	Motion-to-sound = 30 ms	Measure Motion-to-sound	PASS - boundary inclusive	Major	P-LT-006 / Vibe M2S / DeFOSPAM A5	boundary
TC-012-BVA-K	BVA	BVA-OUT	Glass voice (IVAS)	Motion-to-sound just outside (31 ms)	Glass voice (IVAS) test configured	Motion-to-sound = 31 ms	Measure Motion-to-sound	FAIL - exceeds budget	Major	P-LT-006 / Vibe M2S / DeFOSPAM A5	boundary
TC-013-BVA-I	BVA	BVA-IN	Holographic calling	Inter-stream a/v skew just inside (9 ms)	Holographic calling test configured	Inter-stream a/v skew = 9 ms	Measure Inter-stream a/v skew	PASS - within budget (strict <)	Major	P-QO-004 / TR 26.965 6.4	boundary
TC-014-BVA-J	BVA	BVA-AT	Holographic calling	Inter-stream a/v skew at boundary (10 ms)	Holographic calling test configured	Inter-stream a/v skew = 10 ms	Measure Inter-stream a/v skew	FAIL - strict bound, boundary excluded	Major	P-QO-004 / TR 26.965 6.4	boundary
TC-015-BVA-K	BVA	BVA-OUT	Holographic calling	Inter-stream a/v skew just outside (11 ms)	Holographic calling test configured	Inter-stream a/v skew = 11 ms	Measure Inter-stream a/v skew	FAIL - exceeds budget (strict <)	Major	P-QO-004 / TR 26.965 6.4	boundary
TC-016-BVA-I	BVA	BVA-IN	Holographic calling	Volumetric-track skew just inside (4 ms)	Holographic calling test configured	Volumetric-track skew = 4 ms	Measure Volumetric-track skew	PASS - within budget (strict <)	Major	P-QO-004	boundary
TC-017-BVA-J	BVA	BVA-AT	Holographic calling	Volumetric-track skew at boundary (5 ms)	Holographic calling test configured	Volumetric-track skew = 5 ms	Measure Volumetric-track skew	FAIL - strict bound, boundary excluded	Major	P-QO-004	boundary
TC-018-BVA-K	BVA	BVA-OUT	Holographic calling	Volumetric-track skew just outside (6 ms)	Holographic calling test configured	Volumetric-track skew = 6 ms	Measure Volumetric-track skew	FAIL - exceeds budget (strict <)	Major	P-QO-004	boundary
TC-019-BVA-I	BVA	BVA-BELOW	Glass voice (IVAS)	Spatial audio quality (MUSHRA) just below (79 score)	Glass voice (IVAS) test configured	Spatial audio quality (MUSHRA) = 79 score	Measure Spatial audio quality (MUSHRA)	FAIL - below floor	Major	P-QO-002 / Vibe MUSHRA	boundary
TC-020-BVA-J	BVA	BVA-AT	Glass voice (IVAS)	Spatial audio quality (MUSHRA) at floor (80 score)	Glass voice (IVAS) test configured	Spatial audio quality (MUSHRA) = 80 score	Measure Spatial audio quality (MUSHRA)	PASS - meets floor	Major	P-QO-002 / Vibe MUSHRA	boundary

Test Coverage Analysis — Verizon 6G LA 2028

COVERED pipeline · TMap coverage groups · ISO 25010 · risk-based intensity (A95/B80/C60)

<< Test Coverage with GAP Analysis (Data)

Coverage of **1197 test cases** against **31 requirements**. The suite is overwhelmingly **Data-oriented** (boundary/partition) — strong on the quantitative KPIs, thin on appearance-oriented (usability/accessibility) and the deferred/qualitative requirements.



Coverage adequacy score

Risk-weighted: $\min(\text{actual}, \text{target})$ summed over requirements. Below target -> AMBER.

1197

TEST CASES

31

REQUIREMENTS

24

COVERED

0

PARTIAL

GAPS

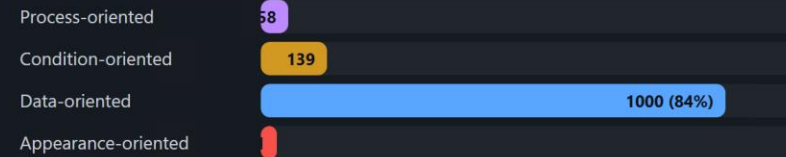
0

APPEARANCE TESTS

ISO 25010 coverage radar (actual vs target)



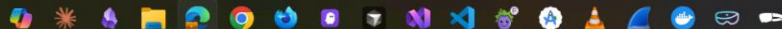
TMap coverage-group distribution



Appearance-oriented coverage is 0 cases - usability/accessibility live in charters/checklists, not the case inventory.



Search



ENG UK



18:43 29/05/2026

<< Go checkout the example sample output

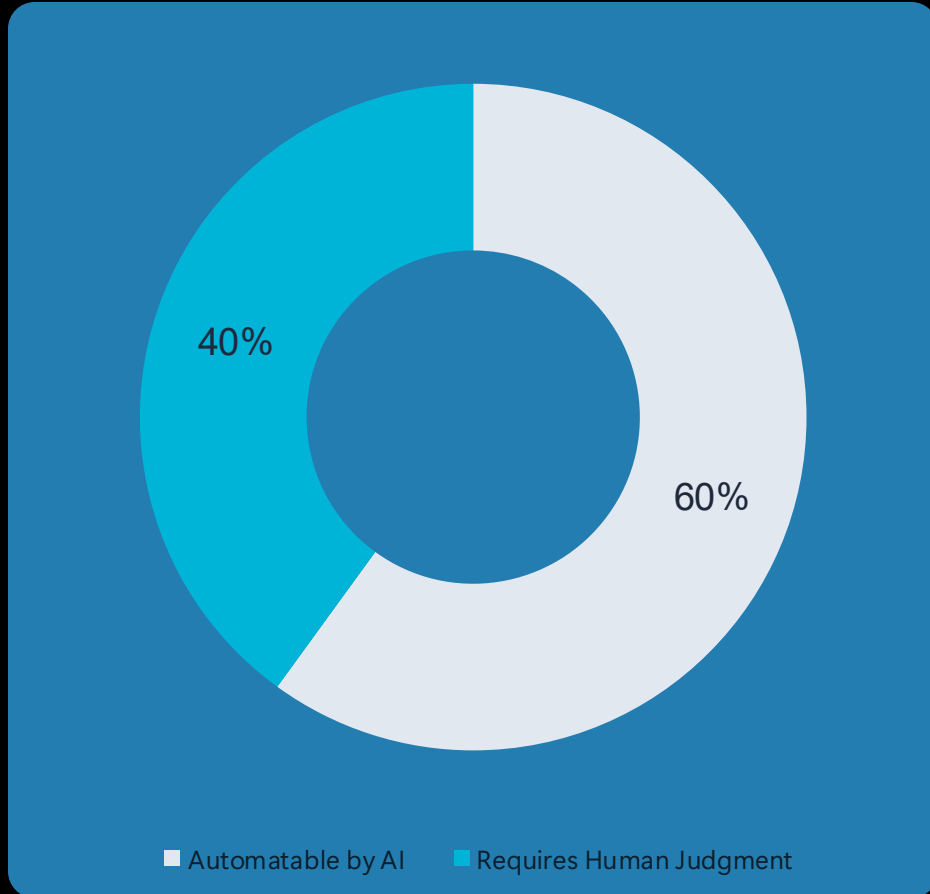
- Files
- main
- Go to file
- 6G-XR-Requirements
 - features
 - 6G-XR-Requirements-Registry.docx
 - 6G-XR-Requirements-Registry.html
 - 6G-XR-Requirements-Registry.json
 - 6G-XR-Requirements-Registry.md
 - 6G-XR-Requirements-Registry.xlsx
 - 6G-XR-Requirements-report.docx
 - 6G-XR-Requirements-report.html
 - 6G-XR-Requirements-report.md
 - 6G-XR-Requirements-results.json
 - 6G-XR-Requirements-tracker.xlsx
 - 6G-XR-Specifications-SRS-IEEE830.docx
 - 6G-XR-Specifications-SRS-RTM.xlsx
 - 6G-XR-Specifications-report.html
 - 6G-XR-Specifications-report.md
 - 6G-XR-Specifications-results.json
 - 6G-XR-Specifications-srs-report.html
 - 6G-XR-Specifications-srs-report.md
 - 6G-XR-Specifications-srs-results.json
 - 6G-XR-Test
 - templates
 - 6G-XR-Test-Cases-Edge.xlsx
 - 6G-XR-Test-Cases.xlsx
 - 6G-XR-Test-Plan.docx
 - 6G-XR-Test-Plan.html
 - 6G-XR-Test-Plan.json
 - 6G-XR-Test-Plan.md
 - 6G-XR-Test-Plan.xlsx
 - 6G-XR-Test-Strategy.docx
 - 6G-XR-Test-Strategy.html
 - 6G-XR-Test-Strategy.json
 - 6G-XR-Test-Strategy.xlsx
 - 6G-XR-Test-Template-Execution.xlsx
 - 6G-XR-Test-Template-Exploratory-Charters.xlsx
 - 6G-XR-Test-Template-Readiness-Checklists.xlsx
 - 6G-XR-Test-Template-Reporting.xlsx
 - 6G-XR-Test-report.json
 - 6G-XR-Test-results.json

OpenTestingAI / .opentesting-examples / 6G-XR-holographic / 6G-XR-Test /

HyperAutomation 6G XR sample ff70eae · 22 minutes ago History

Name	Last commit message	Last commit date
..		
templates	6G XR sample	22 minutes ago
6G-XR-Test-Cases-Edge.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-Cases.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-Plan.docx	6G XR sample	22 minutes ago
6G-XR-Test-Plan.html	6G XR sample	22 minutes ago
6G-XR-Test-Plan.json	6G XR sample	22 minutes ago
6G-XR-Test-Plan.md	6G XR sample	22 minutes ago
6G-XR-Test-Plan.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-Strategy.docx	6G XR sample	22 minutes ago
6G-XR-Test-Strategy.html	6G XR sample	22 minutes ago
6G-XR-Test-Strategy.json	6G XR sample	22 minutes ago
6G-XR-Test-Strategy.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-Template-Execution.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-Template-Exploratory-Charters.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-Template-Readiness-Checklists.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-Template-Reporting.xlsx	6G XR sample	22 minutes ago
6G-XR-Test-report.json	6G XR sample	22 minutes ago
6G-XR-Test-results.json	6G XR sample	22 minutes ago

So are you ready to be part of the Irreplaceable 40%?



AI Confidence Engineering

Evaluating AI outputs, identifying false positives, and challenging assumptions



Knowledge Workers = Context Understanding

Understanding business context, industry experience, and customer implementation overlaying risk implications



AI Assurance Governance

Navigating bias, fairness, transparency, and accountability in AI infused systems



Critical Thinking = Cognitive Reasoning

Handling complex AI confidence engineering and ambiguous inputs and outputs

The human capabilities that AI cannot replicate in AI confidence engineering

Future roles for Testers-in-the-loop (TiTIs)

AI-Agentive Testing Handle — 60%

- ▶ Systematic requirements engineering scanning
- ▶ Test technique selection of 100+ patterns
- ▶ Automated test design generation with traceability
- ▶ Coverage gap detection (4 TMap dimensions)
- ▶ RAG dashboards and progress reporting
- ▶ TMMi maturity assessment (113 practices)
- ▶ Template generation (108 templates)

AI Confidence Engineers Handle — 40%

- ▶ Business context interpretation and validation
- ▶ Risk classification and prioritization decisions
- ▶ Ethical judgment on AI system behavior
- ▶ Stakeholder communication and negotiation
- ▶ Exploratory testing with domain insight
- ▶ Governance framework design and oversight
- ▶ Legacy interactions with other humanoids



Confident principle | SEO/Sales... (6) The Confidence Asymmetry... Extension has been uninstalled... Ask Gemini

linkedin.com/pulse/confidence-asymmetry-principle-jonathon-wright-liqbe/

in I'm looking for... Home My Network Jobs Messaging Notifications Me For Business Advertise

The Confidence Asymmetry Principle

Jonathon Wright Chief AI Officer | Chair of World Agentic AI + AutomationSTAR (2026) | Best-Selling Author on AI (2022) | International Keynote Speaker | TE...

May 22, 2026

Why the Cost of Confidence — Not the Cost of Code — Will Govern the Economics of AI-Infused Software

A Theoretical and Empirical Foundation for Confidence Engineering

Cost to generate a candidate artifact
Cost to justify confidence in that artifact
Irreducible verification floor

Verification asymmetry widens

Comments

15 · 3 comments · 1 repost

Like Comment Share

Add a comment...

Most recent

Adam Sandman Founder & CEO - Inflectra Corporation · 1w
Absolutely agree with this **Jonathon Wright** last weekend I was flying from LA back to DC and during the flight I had Kiro spec-create an entire chatbot application and agentic testing framework. Did it in about 4 hours while I... more

Like · 2 | Reply · 1 reply

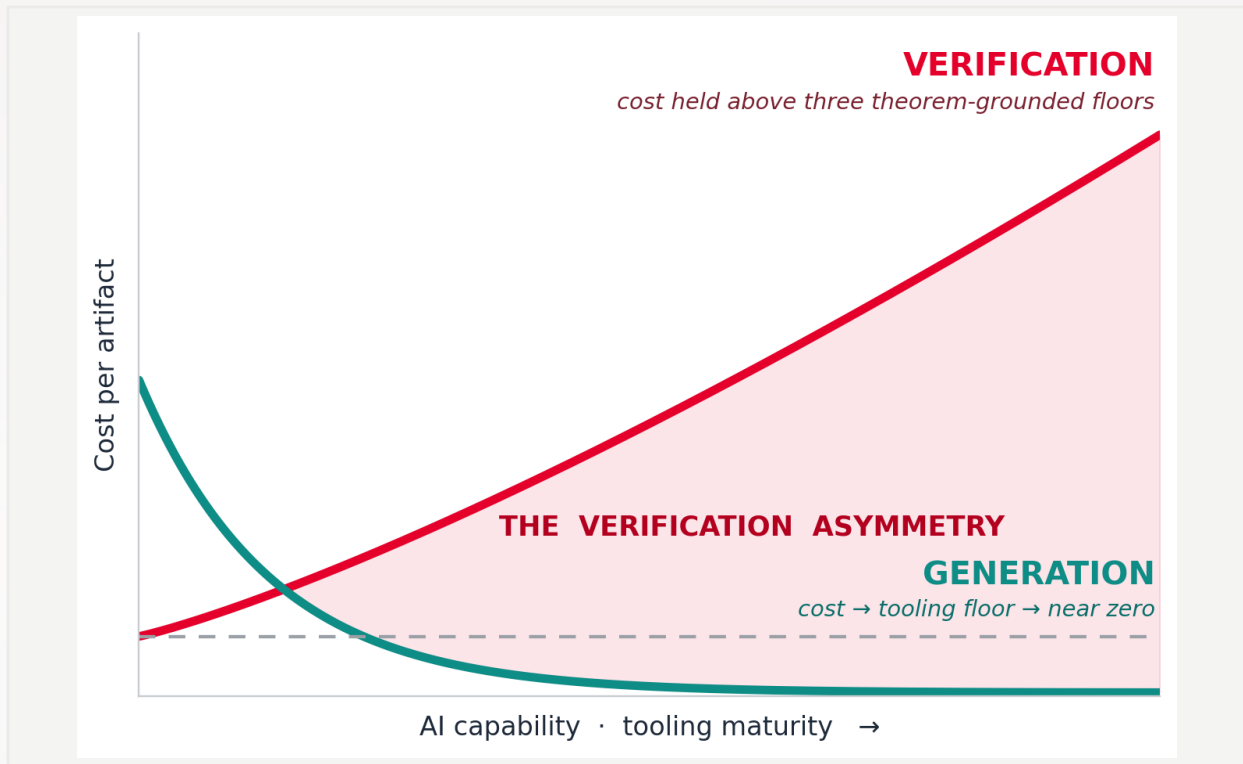
Jonathon Wright Author · 1w
Chief AI Officer | Chair of World Agent...

Messaging

The Confidence Asymmetry Principle

When Code Generation Becomes Close to Free Confidence Becomes the Quality of the Product

AI is collapsing the cost of generating software — not the cost of knowing it is correct. That widening gap is the next AI-native engineering market.



AI GENERATION = COST → ≤ 0

Cost per artifact falls to a tooling floor, bounded only by the price of compute. It trends toward zero.

AI VERIFICATION — COST FLOORED

Cost per artifact is held above three theorem-grounded floors. It does not fall — and rises as systems complexity grows.

THE CONFIDENCE BOUND

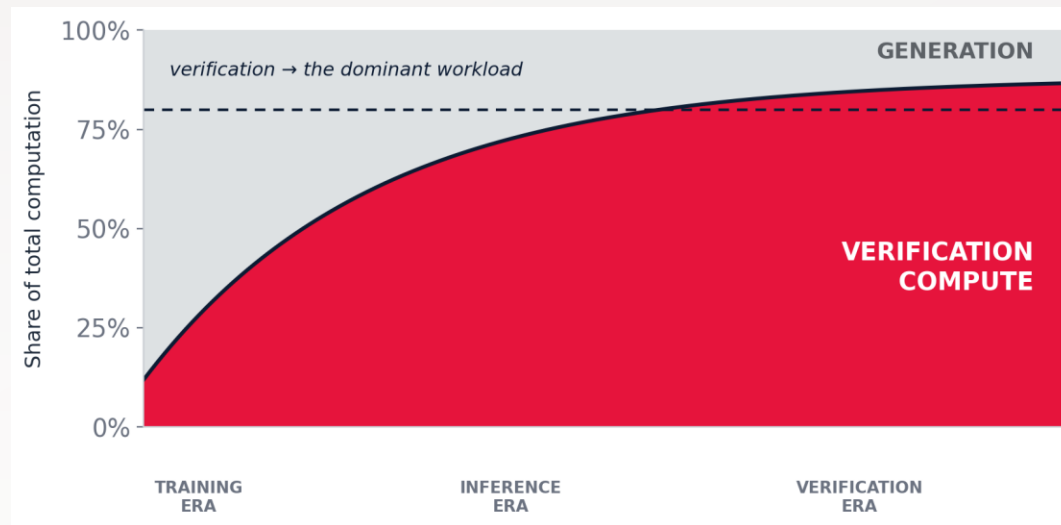
$$C_{\text{ver}} \geq [(\log_2(1/\epsilon) + \log_2(1/p_0) + \kappa \cdot n^\alpha) / (l_{\text{max}} \cdot (1 - \rho))] \times c_{\text{exec}}, \text{ with } \alpha \geq 2.$$

Source: The Confidence Asymmetry Principle · AI Confidence Engineering (Jonathon Wright, 2026) (<https://www.linkedin.com/pulse/confidence-asymmetry-principle-jonathon-wright-liqbe/>)

AI Confidence Engineering Manifesto

AI Compute Migrates to Verification Era

As generation cost falls, verification becomes the dominant workload — and AI Confidence Engineering is the discipline that owns it.



≈80%

of computation is forecast to be consumed by verification once generation is effectively close to zero.

THE ANSWER — AI CONFIDENCE ENGINEERING

The discipline that owns the question — do we know enough to ship it?

- 1 Direct AI at risk, not volume**
Verification effort allocated by business value and risk — not spread evenly across generated output.
- 2 Insist on decorrelated evidence**
Execution, exploration and monitoring — evidence sourced from outside the generator's own distribution.
- 3 Evaluate continuously**
Confidence does not persist across change — it is re-earned every release, never banked.
- 4 Ship a graded confidence rating**
A probability with an interval — an auditable, explainable risk grade for AI-infused output.

CORRELATED-BLINDNESS PENALTY When AI verifies AI, generator and verifier share blind spots — a “pass” then tracks authorship, not correctness. Using a correlated model to test AI-generated code compounds risk; it does not cancel it.

Source: The Confidence Asymmetry Principle · AI Confidence Engineering (Jonathon Wright, 2026) (<https://www.linkedin.com/pulse/confidence-asymmetry-principle-jonathon-wright-liqbe/>)

Jank.AI – The solution to AI code generated slop!



Edge File Edit View History Favourites Profiles Tab Window Help
World Digital Report 2025 x testera.ai - Quality Intelligence: x jank.ai - Less Jank, Ship Faster. x
https://jank.ai

Jank

Jank checks for AI coding agents

Get Early Access

JANK LEVEL 28%

```
ai coding agent - Requested  
- shipped: strength meter on /signup  
- add forgot-password flow  
- AI coding agent: + /reset route + email send
```

studio.restream.io is sharing your screen. Stop sharing Hide

A screenshot of a web browser displaying the Jank.AI website. The page features a large green 'Jank' logo and the text 'Jank checks for AI coding agents'. Below this is a 'Get Early Access' button. The main content area shows a 'JANK LEVEL' progress bar at 28% and a list of AI coding agent actions. At the bottom, there is a notification that 'studio.restream.io is sharing your screen'.

Paul & Jonathon

LIFETIME IMPACT 864 reports scanned - 622 completed - \$50/hr basis jump: daily breakdown

REPORTS	UNIQUE SITES	REQUESTERS	ISSUES FOUND	HOURS SAVED	\$ SAVED	WAIT DAYS SAVED	SPEEDUP
864	459	364	6,868	3,950 hr	\$206,175	493.8 days	9x

Loaded from /api/admin/stats/cumulative - super-admin only.

A screenshot of a dashboard showing lifetime impact statistics. The statistics are presented in a grid of green boxes. The total impact is 864 reports scanned, 622 completed, and \$50/hr basis. The dashboard also shows 459 unique sites, 364 requesters, 6,868 issues found, 3,950 hours saved, \$206,175 saved, 493.8 wait days saved, and a 9x speedup.

4 May at 17:01

May the Forth TEST you! AI Replace Testers?

WorldDigitalReport.com
"AI will replace testers?" ...
See more

Like Comment

Comments Hide comments

Be the first to leave a comment.

Comment as Jonathon Wright

A screenshot of a social media post. The post title is 'May the Forth TEST you! AI Replace Testers?' from WorldDigitalReport.com. It includes a 'See more' link and social media interaction options like 'Like' and 'Comment'. Below the post, there is a comment section with the text 'Be the first to leave a comment.' and a 'Comment as Jonathon Wright' button.

TESTERS.AI
TACTICAL HUD · OPS

REV **00350** -XSZ

REGION **us-central1**

WORKERS **11** /12
healthy

QUEUE **3**
+2 / min

BUGS / H **189**
+12

SUCCESS **95.0%**
-0.2 pp

P95 **6.8** m
+1.3

16:24:30
02 JUN 2026 UTC

[01] // SECTOR.WORLD refresh 1.2s ----- tracking 3 contacts



N MOST	S MOST	SECTORS HOT	TOTAL ACTIVE
53.4 N	-33.9 N	0	3

REGION	ACTIVE	QUEUE	AVG MS	TREND
NA · West	00	00	237	+5.7%
NA · East	00	01	185	+4.4%
S America	00	00	203	+1.6%
EU · West	00	00	250	+5.1%
EU · Central	00	00	275	-2.9%
Middle East	00	00	263	+2.1%
Africa	00	00	223	+4.1%
India	00	01	210	+6.2%
APAC · East	00	00	188	+1.9%
APAC · SE	00	01	221	-2.4%
Oceania	00	00	217	+3.7%

T-1M

linear.app 72/100

T-4M

shopify.com 84/100

T-2M

clay.com 78/100

T-4M

arc.net 93/100

T-5M

linear.app 76/100

[02] // VITALS live ----- all systems nominal



[04] // FIELD.READOUT last 60m ----- stream live



[03] // CONTACT.PRIME 7d4c49bf-dfd7-440a-9809-6e19fd4b4aea - eta 0s

7D4C49BF-DFD7-440A-9809-6E19FD4B4AEA beikecelltherapy... + 05:23

QUEUE
LAUNCH
CAPTURE
ANALYZE
PERSONAS
FLows
EXPLORE
FINAL

// LIVE FEED // QUEUE

Why Choose Beike for a Stem Cell Treatment?

- Experience
- Safety
- Diversity
- Extensiveness
- Support

1 pages · 4 agents

PAGES TESTED

1 / 14

AGENTS ONLINE

6 / 12

FINDINGS

0

REQUESTS · 7D4C49BF-DFD7-440A-9809-6E19FD4B4AEA

216 /s peak

- +00s MIA skeptical buyer · pricing table missing comparison
- +01s MIA flagged missing alt on hero image (1.1.1)
- +02s MARCUS session cookie missing HttpOnly · flagged
- +03s MIA screenshotting at 375 · 768 · 1440
- +04s MIA screenshotting at 375 · 768 · 1440
- +05s MIA screenshotting at 375 · 768 · 1440
- +07s TARIQ aggregating Lighthouse · perf 74 / 100
- +07s TARIQ aggregating Lighthouse · perf 74 / 100

Key Takeaways – AI-Confidence Engineering

1

40% of AI engineering assurance requires human judgment — critical thinking, domain expertise, ethical reasoning, and adaptive problem-solving cannot be automated away.

2

Open-source AI agent ecosystems (100+ agents across OpenRequirements, Open-Testing, and OpenTest) provide structured upskilling pathways grounded in international standards.

3

Standards-based knowledge (ISO 29119, TMMi, IEEE, TMap) gives teams a common language and measurable maturity framework.

4

The workforce model is augmentation, not replacement — AI handles the systematic 60% so humans can focus on the high-value 40% that ensures responsible, scalable AI adoption.

OPENTEST UNIVERSITY

Confidence Engineering Certification

A role-based certification for engineers who build evidence, evaluation systems, production signals, and professional judgment around AI-era software.

Evidence

Risk, evals, telemetry, and math.

Judgment

Scenario choices and written exams.

Public

Badge records built to share.

Start Certification

View Curriculum



Email-verified accounts with secure password reset.



Lesson progress, scores, exam results, and badges are public by design.



Verified badge records designed for LinkedIn and public profile sharing.



Dynamic exams generated from lesson outcomes, article lineage, math concepts, and scenario actions.

CREDENTIAL STATUS Public Verification



Certified Confidence Engineer

A shareable OpenTest University credential for LinkedIn and public verification.

10

Courses

80%

Pass score

60+

Question forms

AI Bestselling Book Sequel + WorldDigitalReport.com

The screenshot displays a digital report interface with a dark theme. On the left, a navigation sidebar contains a search bar and a table of contents. The main area is a grid of content tiles, each featuring a portrait of a person and a snippet of text. The sidebar table of contents includes sections like 'Introduction', 'Executive Summary', 'World Digital Report - Key Findings', and 'Generative AI in Quality Engineering'. The grid tiles show various data visualizations, including bar charts and line graphs, alongside the portraits of individuals. At the bottom, a Windows taskbar is visible with various application icons and system tray elements.

Navigation

Search document

Headings Pages Results

- Table of Contents
- Introduction
 - Setting the Global Context
 - What's New in 2026?
 - Quality Intelligent AI Horizons
 - Our Supporter Partners: Better Together
- Executive Summary
 - The Adoption-to-Value Creation Gap
 - The Generative AI Inflection
 - The AI-Agentic Testing Frontier
 - The AI-Assurance Imperative
- World Digital Report - Key Findings
 - Generative AI in Quality Engineering
 - AI-Agentic Testing: Self-Evolving Agents
 - Autonomous Testing & Self-Evolving Systems
 - AI-Driven Synthetic Test Data Management
 - AI-Driven Defect Prediction & Analytics
 - AI-Assurance: Testing AI Infused Systems
- Generative AI in Quality Engineering
 - Beyond the Hype: Generative AI finds its testing g...
 - The Generative AI Adoption Landscape
 - The Global Generative AI Context
 - Where Generative AI Delivers the most Value
 - Closing the Generative AI Confidence GAP
 - Generative AI and the Evolution of the Quality...
 - The Prompt Engineering Imperative
- AI-Augmented Testing: Tester in the Loop
 - The Quality Intelligence Revolution
 - Organisational AI Productivity Investments
 - The State of AI-Augmented Testing Automatio...
 - Where AI-Augmented Testing is making the Bi...
 - The Automation Coverage Level Leap
 - The Maintenance Revolution
 - Key Challenges in AI-Augmented Testing
 - The Automation Static Checking ROI Reality
- AI-Agentic Testing: Self-Evolving Agents
 - Thinking machines that can test themselves
 - The Rise of Autonomous Self-Evolving Cogniti...
 - The Autonomous Testing Maturity Spectrum
 - Self-healing Test Automation: The Gateway Fix
 - Autonomous Exploratory Testing
- AI-Driven Synthetic Test Data Management
 - The Data Dilemma: Solved by Synthesis
 - The Data Infrastructure Imperative
 - The State of AI-powered Test Data Management
 - The Synthetic Data Revolution

Page 1 of 69 15709 words English (United Kingdom) Text Predictions: On Accessibility: Investigate

10:07 20/04/2026

The Irreplaceable 40%: AI-Confidence Engineering + Confidence Asymmetry



Connect with me on [LinkedIn](#)



A final thought: AI in 2026

As IT budgets continue to rise, with a projected 3.6% growth in 2026, the ability to demonstrate AI's value, secure business buy-in and drive real transformation will define the next generation of CIOs.

The future will belong to CIOs who take decisive action, those who embrace AI with a sharp vision, align it with business goals, and lead with confidence to evolve their organizations to AI productive.

Are you ready to shape the next era of AI agentic testing?