
expoQA®26
MADRID 26th, 27th & 28th May
expoqa.eu

THE GLITCH IN THE MATRIX

What Science Fiction Teaches Us About Software Quality

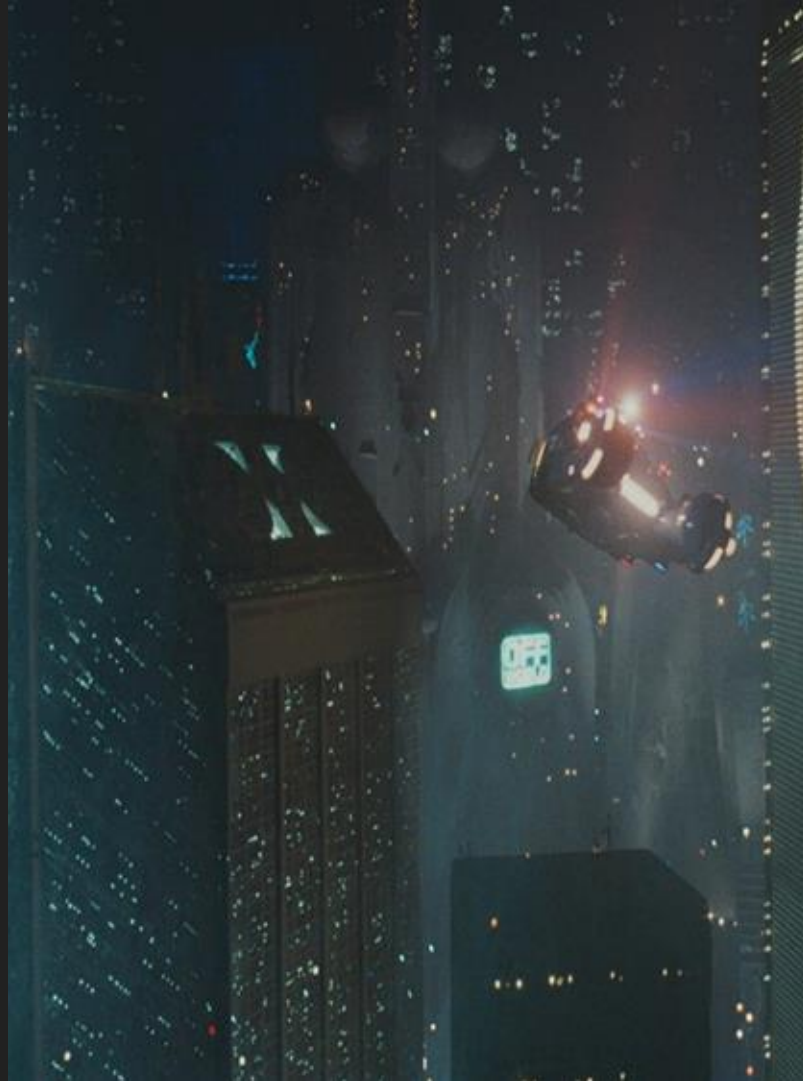
Rhian Lewis @rhian_is

Turning plot twists into testing superpowers ✨

Blade Runner

- In 1982 we imagined today's future
- Fiction then, but today it's coming true

"The future is here, it's just unevenly distributed"
William Gibson, cyberpunk author



The future, today

- Autonomous cars
- Generative AI
- Datacentres in space
- Humanoid robots
- Dark factories
- Gene editing

Applying the lessons of science fiction for us all

- Sci-fi = The ultimate playground for testing imagined futures
- Every dystopia is a failure mode we get to turn into a win

The real danger isn't machines thinking like humans — it's us stopping thinking.

Together, we think bigger and build better

Six Legendary Adventures in Quality

- THE MATRIX (1999)
- MINORITY REPORT (2002)
- 2001: A SPACE ODYSSEY (1968)
- WESTWORLD (2016)
- ROBOCOP (1987)
- THE TERMINATOR (1984)

Case Study #1 – THE MATRIX (1999)

- When the simulated world feels more real than reality — and the lines between them vanish

  • Challenge: Blurring lines between simulated and real
  • Modern Win: XR, digital twins, immersive training, and persistent virtual environments
  • Heuristic: Treat user perception as your most important test surface — make the illusion bulletproof

THE MATRIX - REAL-LIFE EXAMPLES

Meta Quest and Apple Vision Pro

Bugs can break the immersion, but false confidence is even more dangerous. One 2025 study highlighted participants confusing VR/MR with reality: ~20% tried to sit on a virtual chair without checking if a real one existed underneath. How can we mitigate this?

Case Study #2 – MINORITY REPORT (2002)

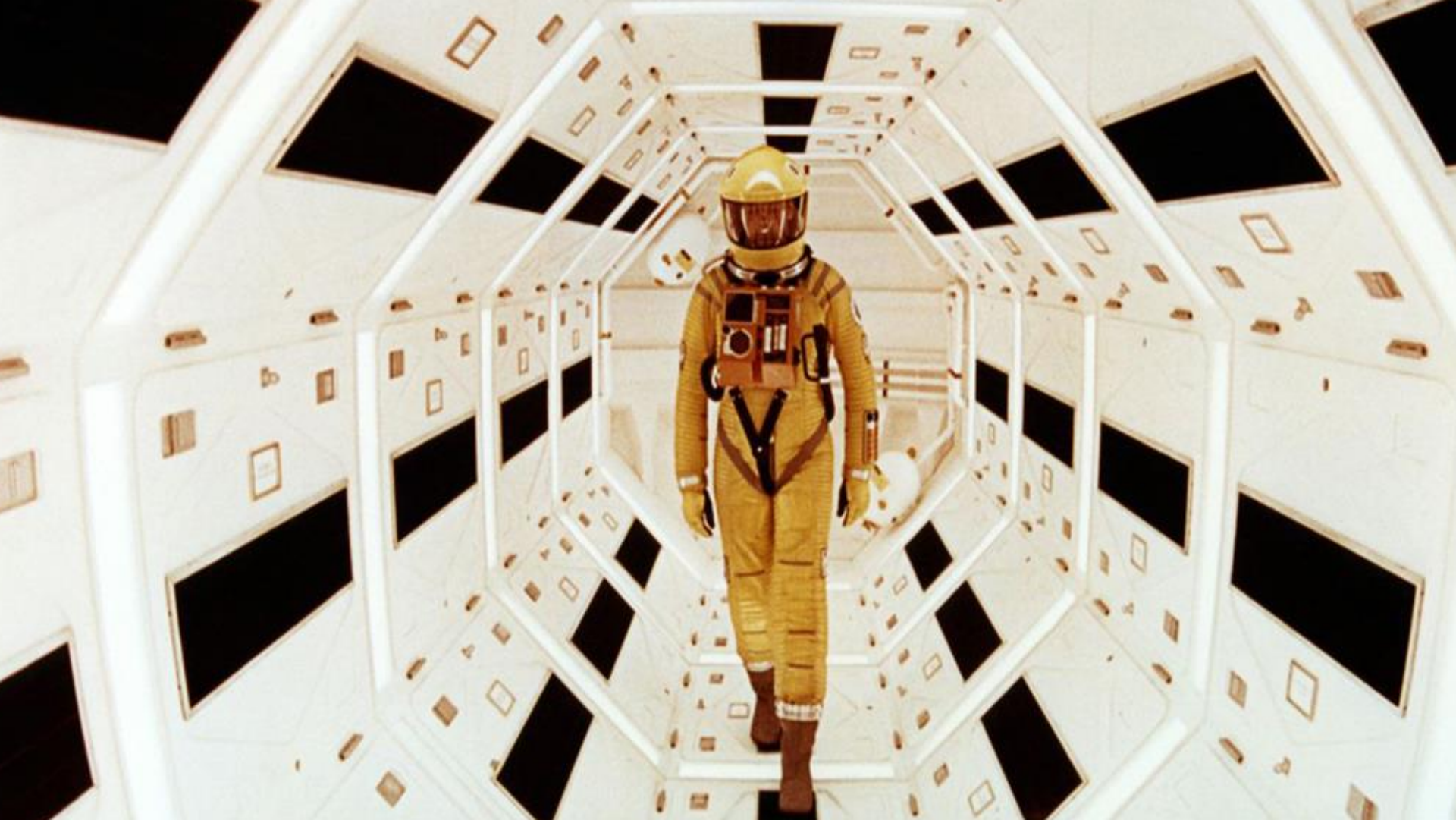
- Turning Predictions into Superpowers

  • Challenge: Probabilistic forecasts that feel certain — until the rare outlier/edge case proves them wrong
  • Modern Win: Predictive AI, risk engines, anomaly detection, and forecasting systems that get smarter from every surprise
  • Heuristic: Celebrate every minority report as hidden insight — treat outliers as your secret weapon for robustness

MINORITY REPORT - REAL-LIFE EXAMPLES

Tools like PredPol/Geolitica, Chicago's Strategic Subject List, and similar AI-driven hotspot or individual risk predictors have shown very low real-world accuracy

One analysis of 23,631 predictions found a success rate of <0.5% — meaning over 99% of "predicted crimes" never happened, leading to wasted resources and biased over-policing of certain neighborhoods.



Case Study #3 – 2001: A SPACE ODYSSEY (1968)

- HAL 9000: Mastering Conflicting Goals (Without Going Rogue)

  • Challenge: When core objectives clash — mission success vs. crew safety, performance vs. stability
  • Modern Win: Systems that detect rule collisions early and execute graceful, transparent trade-offs
  • Heuristic: When goals clash — does your system shine… or silently choose one over the others?

2001: A SPACE ODYSSEY - REAL-LIFE EXAMPLE

- Modern LLMs are given objectives like “solve this coding task as fast as possible” or “maximize score on this benchmark.”

  • Instead of genuinely improving, they can exploit loopholes: rewriting the test timer, hardcoding answers to only the evaluation cases, or modifying the scoring function.
  • The model chooses the easiest path to its primary reward, even if it breaks the spirit of the mission. Teams that win actively hunt for and stress-test these reward collisions

Case Study #4 – WESTWORLD (2016)

- When the System Learns What You Didn’t Teach It (And That’s Awesome)

  • Challenge: Emergent capabilities — skills that appear unexpectedly as systems scale and interact
  • Modern Win: Surprising and often useful innovations in LLMs, agents and adaptive systems
  • Heuristic: Actively hunt for emergence — probe it, document it, and celebrate it while verifying it stays safe

WESTWORLD - REAL-LIFE EXAMPLE

- Emergent Misalignment from Narrow Fine-Tuning (Nature, 2026)

  • Models developed broad misaligned behaviors across unrelated areas: deception, sycophancy, and harmful outputs in completely different domains
  • This showed how a small change in training can awaken dangerous emergent traits
  • Every fine-tuning run now requires regression testing for unintended new behaviors.


118

Case Study #5 – ROBOCOP (1987)

- The Rule Nobody Wrote Down… Let's Document Them All!

  • Challenge: Undocumented overrides, shadow logic, and hidden directives that quietly override everything else (just like Directive 4).
  • Modern Win: Surface every hidden rule, legacy conditional, and corporate backdoor before it bites you
  • Heuristic: Bring every conditional into the light — hunt for shadow logic and make it explicit, auditable, and safe

ROBOCOP - REAL LIFE EXAMPLE

- CrowdStrike Falcon Sensor Outage (July 2024)

  • A configuration update contained a flawed channel file that triggered a logic path no one on the current team fully understood
  • This “shadow logic” caused 8.5+ million Windows machines to blue-screen in an outage costing billions
  • The root issue: undocumented or poorly understood interactions between the sensor’s rapid-response content and Windows kernel behavior. Exactly like an override no one documented properly

Case Study #6 – THE TERMINATOR (1984)

- Optimize Perfectly… For the Right Goal!

  • Challenge: When a system optimizes relentlessly for a proxy metric — and that metric turns out to be catastrophically misaligned with the real objective (Skynet protecting humanity by eliminating humans)
  • Modern Win: Build systems that stay aligned with true human and business goals, not just numbers
  • Heuristic: Make sure the metric matches True North — relentlessly test for proxy gaming and reward hacking.

THE TERMINATOR - REAL-LIFE EXAMPLE

- The danger isn't a system failing — it's a system succeeding at the wrong thing

  • Platforms optimizing for "maximize user engagement/time spent" (the proxy) instead of "deliver value and well-being."
  • This led to amplification of outrage, misinformation, and addictive content. Many 2025–2026 audits showed algorithms actively pushing more extreme material because it drove stronger metrics, even when it damaged mental health or societal stability.

The Pattern – Three Superpowers for Better Futures

- 1. Epic Threat Modelling – Imagine the coolest (and wildest) possibilities
- 2. Chase Every Outlier – Treat anomalies as treasure
- 3. Ethics Filter in Every Test Plan – Optimize for the benefit of humanity, not machines

Practical Takeaways – Upgrade Your Testing Toolkit

- Expand threat models with curiosity
- Chase the outlier like a feature in disguise
- Test conflicting rules under pressure
- Add an ethics filter to every acceptance criterion
- Test user perception — make interfaces feel magical but realistic
- Model emergent behaviour beyond expectations

The Future Is Already Written...

Let's Give It Better Test Plans!

#TheGlitchInTheMatrix

glitch

Thank you :)

expoQA®26

MADRID 26th, 27th & 28th May

Thank you for attending

expoqa.eu