



AI in Test Automation

Exploring the Real Costs and Benefits of AI Tools

Test Bash Brighton

September 2024

Titus Fortner

Hi! I'm Titus



Software Development
Engineer in Test



Open Source Developer



Solutions Architect
Developer Experience Engineer
Developer Advocate



Optimize for Bottlenecks



What “The Money” Wants



Continuous Testing



	LEGACY TESTING	BEGINNING AUTOMATION	INTERMEDIATE AUTOMATION	CONTINUOUS TESTING
Org Structure	Dev & QA Separate	Separate but communicating	Tight Collaboration	QA merged into Dev
Pipeline	None	Early Stage CI	CI Server utilized for build, unit & functional tests	Code commit to deploy fully automated
Deploy Frequency	Scheduled (1-6 months)	Regular (weekly to quarterly)	Daily to Weekly	Anytime
Test Frequency	Release Only	Daily	On Build	On Commit



Testing: expertise at scale!

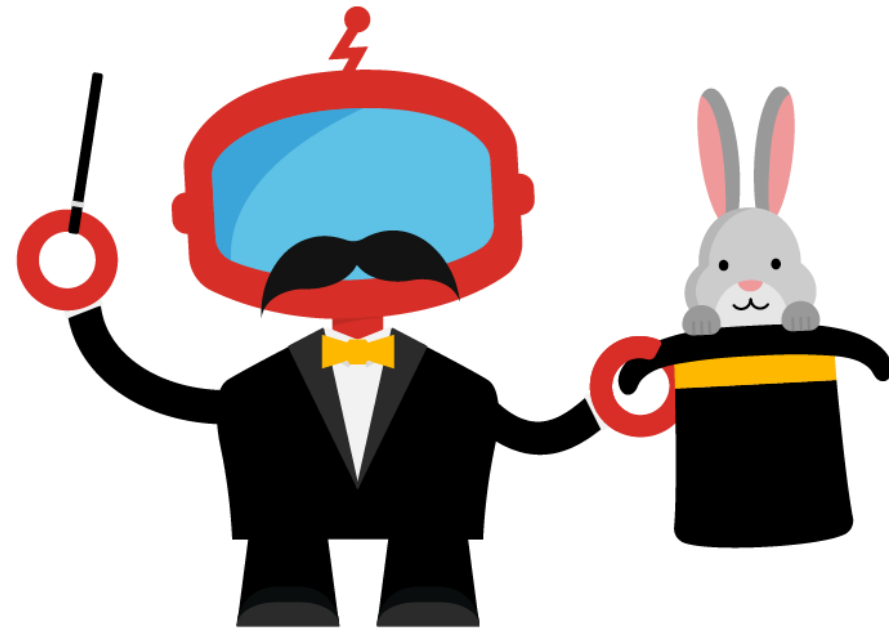


<https://titusfortner.com>



Magic Solutions

- Manual testing doesn't scale
- Good test automation is hard
- Tempting alternatives
 - Record & Playback
 - Codeless Testing
 - AI / ML / LLM





Testing: so easy,
even a developer could do it!



<https://titusfortner.com>

Shifting Left

- Excuse to fire QA Teams?
- Regression vs Acceptance Tests
- Context for Tests
 - Longevity
 - Volatility
 - Complexity





Optimize for the Bottleneck



Maintenance Costs



Creation Costs



Execution Costs



Types of Maintenance Costs

	Feature Correct	Feature Not Correct
Test Passing	Success!	Type II Error
Test Failing	Type I Error (flaky/brittle)	Real Issue (Reproduce/Report)



Type II Errors

- How do you validate your test?
- Does it do what you think it does?
- Are you *sure*?

RED^{BLVD}

GREENST

RED^{BLVD}

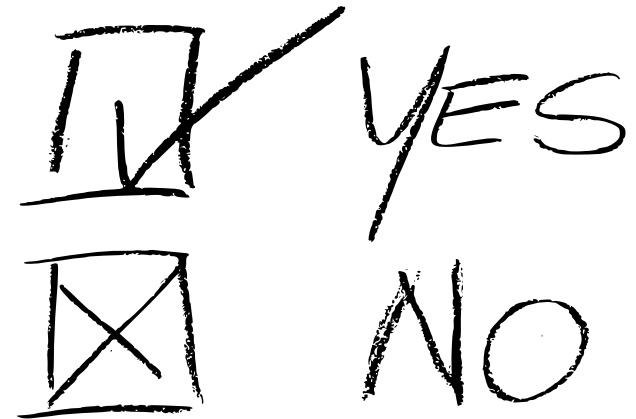
GREENST

REFACTOR^{AV}



Type I Errors

- Flaky Tests — intermittently fail
 - Known States
 - Independent Data
 - Race Conditions
- Brittle Tests — more likely to fail
 - Breaks from *legitimate* changes
 - Once failing always failing

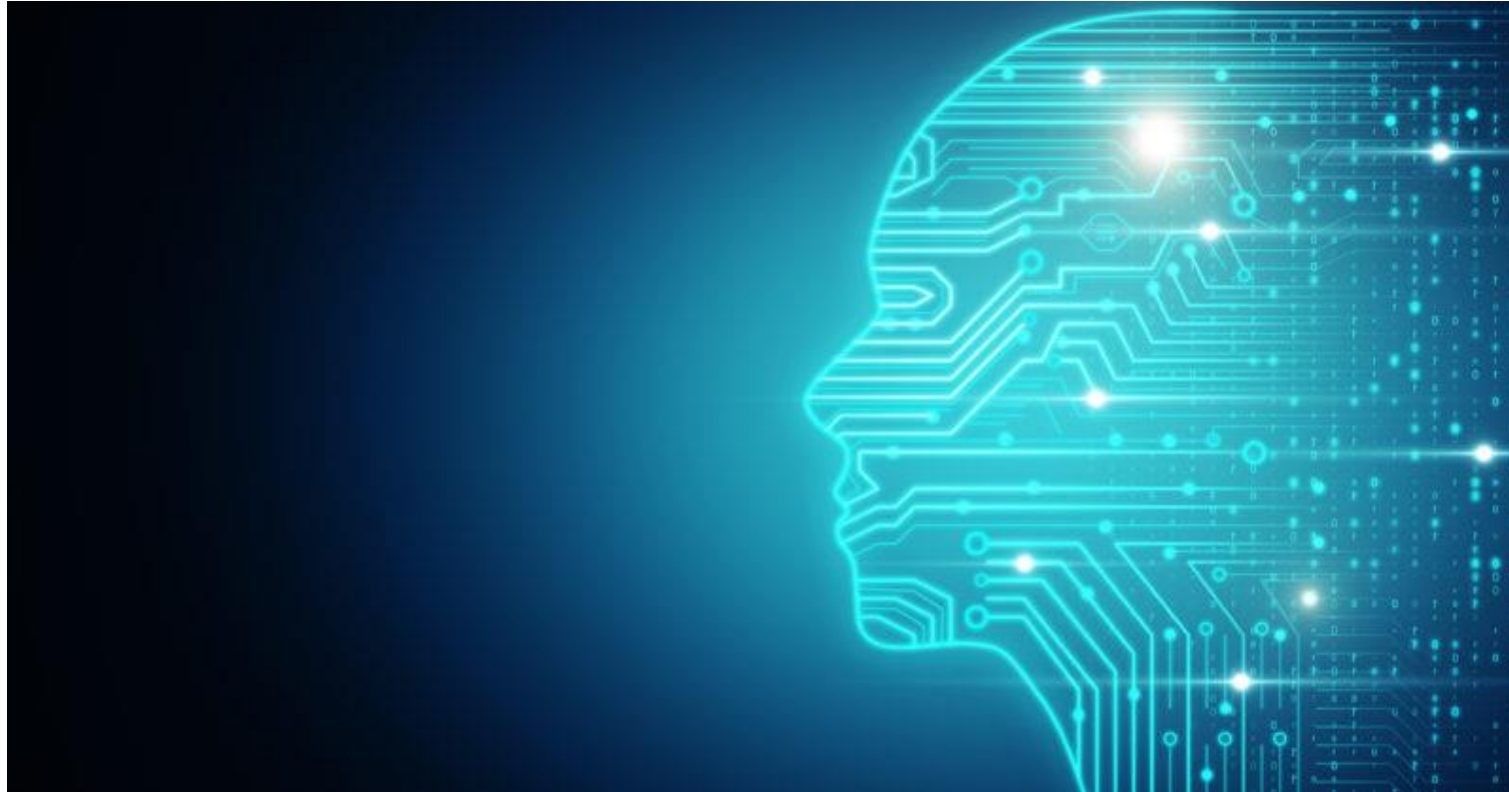


Real Issues

- How long to determine?
 - Error message
 - Screenshot
 - Video
 - Console Logs
 - Command Logs
- Review/Reproduce/Report
 - BDD Tooling



“Now with AI!”





The Bottom Line

- Throughput
- Confidence
- Cost

Impact on Test Creation

- Test Case Management
 - Disaster Scenarios
 - Likely Scenarios
- Code Generation
 - Edge Cases
 - Test Data
- Obfuscation





... Additional Considerations

- Confidence Manager
- Test the Tests
- Beware Slick Demos
- Focus on Maintainability
 - Populate templates
 - Example code
 - Iterative improvements





Impact on Test Execution

- When is a test suite finished?
- Scalable
- Efficient
- Performant



Impact on Test Maintenance

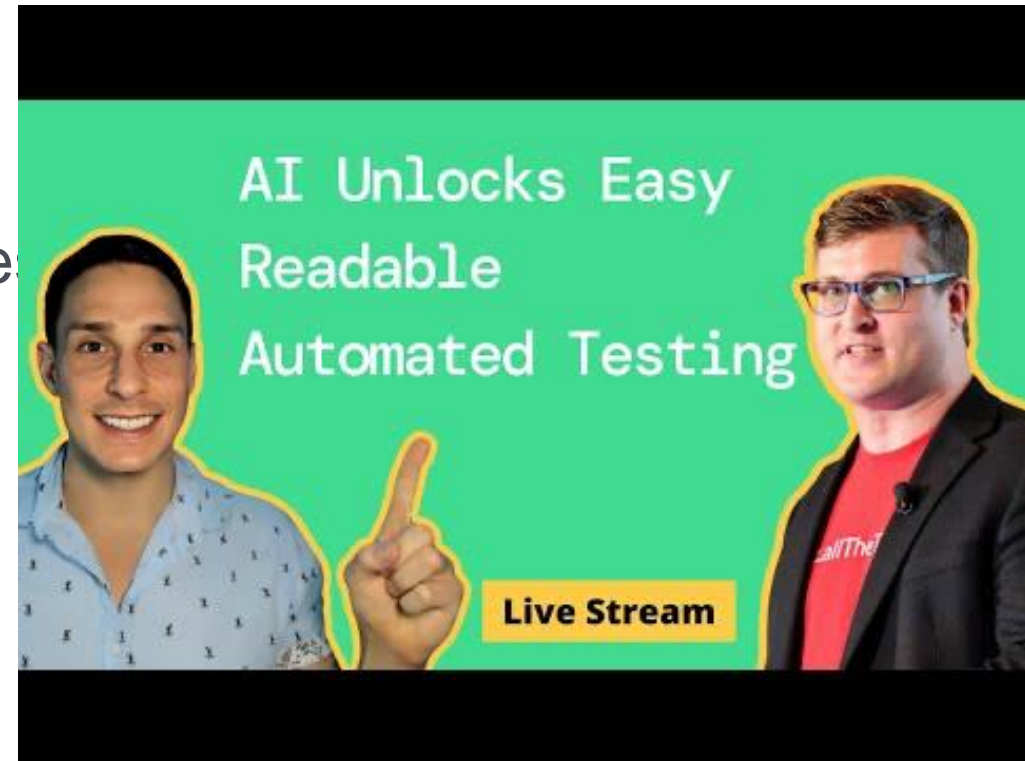
- Type II Errors
 - Hallucinations
 - Self-Healing
- Type I Errors
 - The Halting Problem
- Real Bugs
 - Fast & actionable information





Test Automation Experience LiveStream

- Reverse BDD Tooling
 - Readability without Regular Expressions
- Declarative Code
- Imperative Reports



<https://www.youtube.com/watch?v=kq04gZGaUVk>

1-Oct-24

@titusfortner

21

Developing Selenium with AI



5 Programming Languages

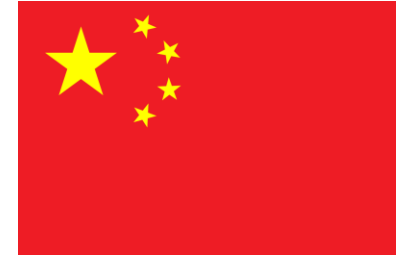
- Translating methods
- Translating classes
- Matching conventions
- Common APIs





4 Documentation Languages

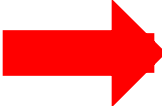



- Automatic PR translations
- Mixing Human and AI updates
- Chat GPT not good enough (yet?)
- Special Services better
 - Expensive
 - Effort





Building and Releasing

- Multiple days
- I don't know Bash at all
- Bazel can be challenging
- BUT...
 - Hidden assumptions
 - Breaking changes
 - Obscure workarounds

- GitHub Action 
- Rake Task 
- Ruby Code 
- Bash Script 
- Bazel Command 
- Python Code



Questions