A4Q TESTING SUMMIT

Two Minds, One Mission
Pairing Exploratory Testing with a GenAI Agent

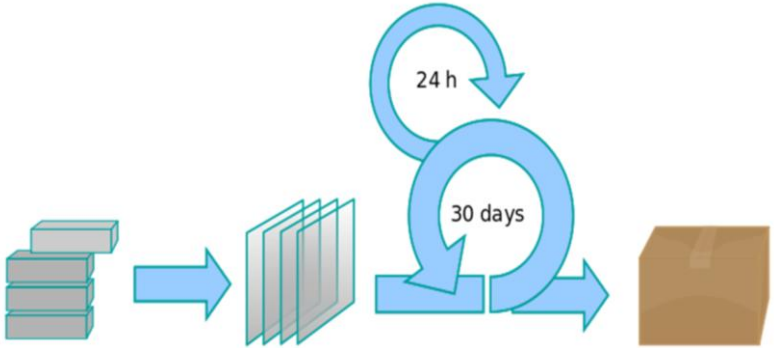
Tal Pe'er

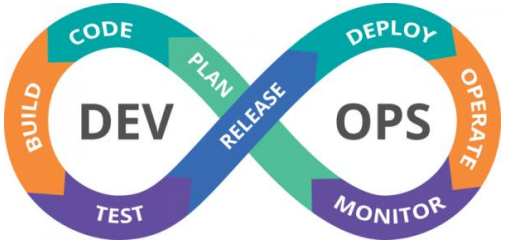
Who I Am

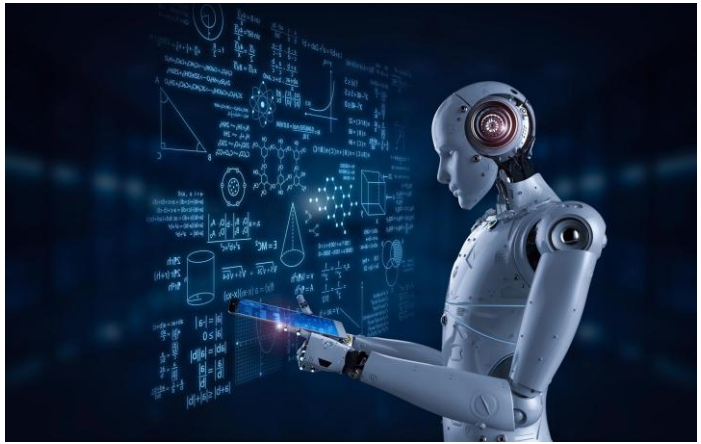
Tal Pe'er

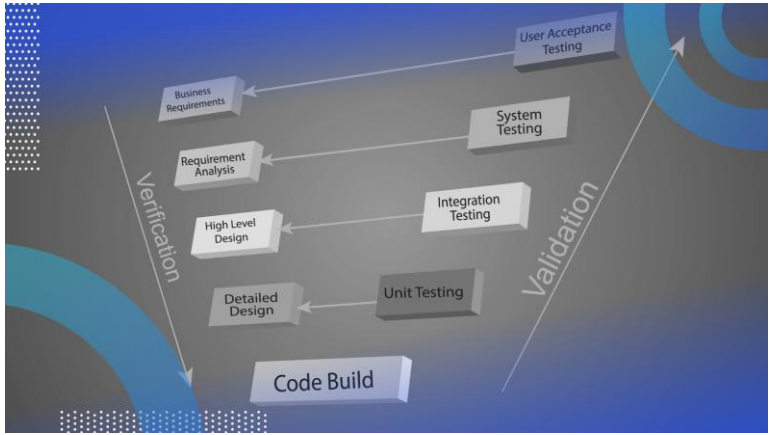
- 30 years in software testing
  - tester, test manager
- consultant and trainer in Grove Software Testing
- various roles in ISTQB®

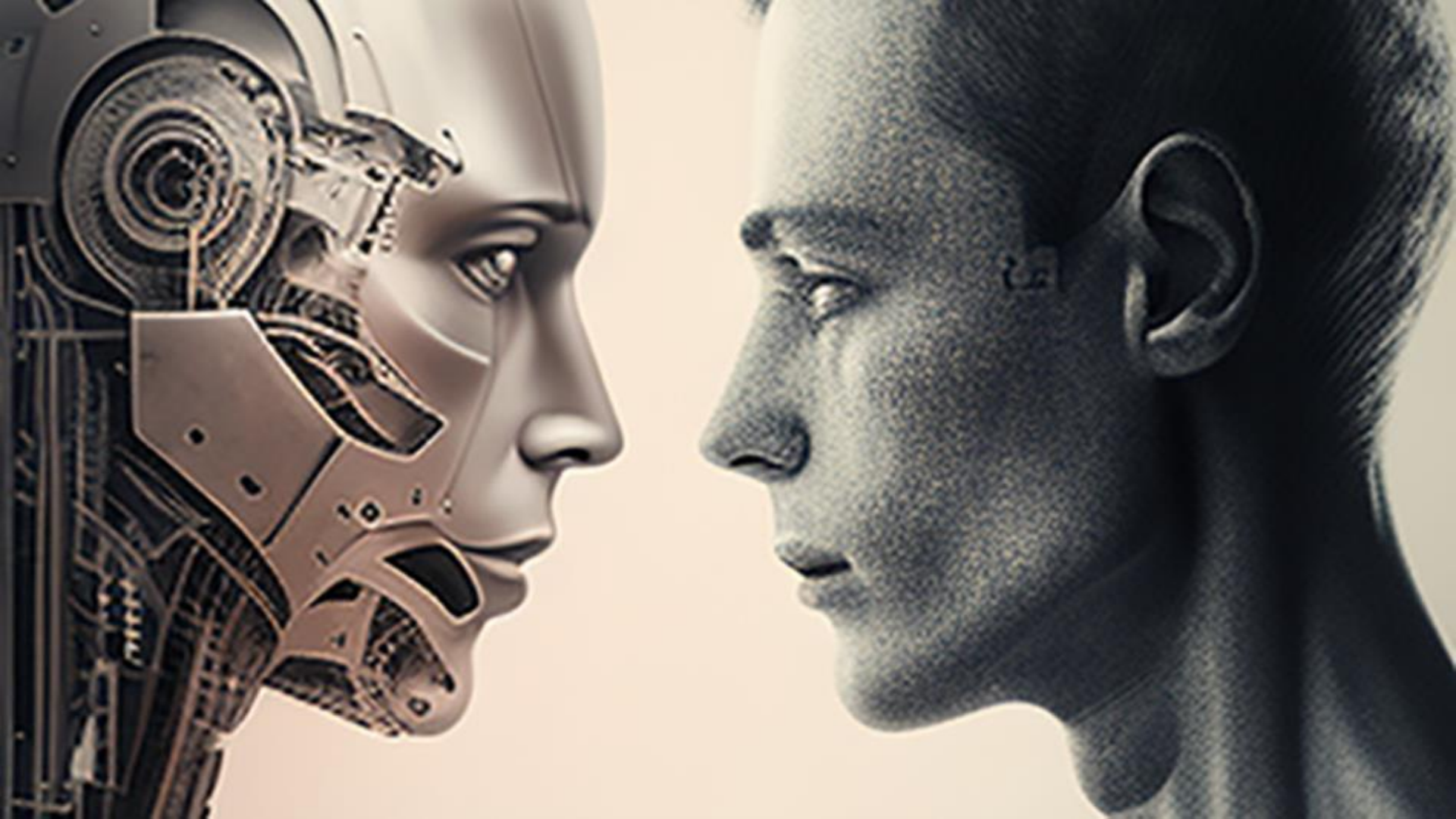
Modern Testing Landscape











Automation does not find bugs

TESTS DO

Automation Doesn't Find Bugs

Tests find bugs

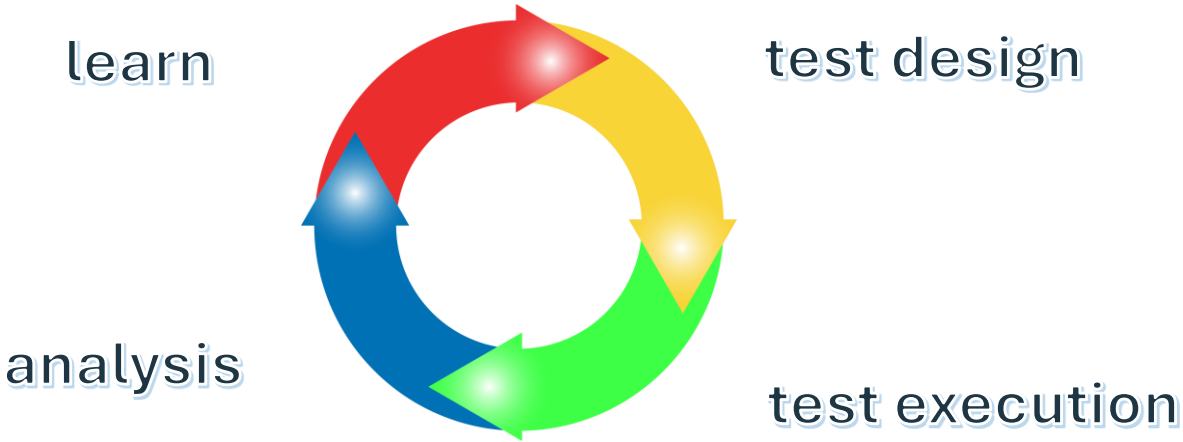
Automated tests	Manual Scripted	Exploratory	Fix Verification
9.3%	24.0%	58.2%	8.4%

Experiences of Test Automation, Ch 27, p 503, Ed Allen & Brian Newman

Exploratory Testing 101

James Bach: simultaneous learning, test design, and test execution

- no pre-designed test cases / scripts
- producing test ideas as a response to system's behaviour
- actively trying to find the 'cracks' in the software



Explore the system
(under test)
STAR TREK

Which AI agent to use?

Gemini

Claude

How do we split the work?

what a GenAI agent does?

- breadth of ideas
- rapid data generation
- structured notetaking
- assumption challenges

what a tester does?

- sets the charter and risk focus
- applies domain knowledge
- applies ethical judgement
- validate finding against oracles

Clear roles definitions

human driver; AI navigator

AI driver; human navigator

ping-pong pairing

shadow scribe

Designing effective sessions

- use charters to keep a structured but flexible session
  - a charter defines your mission statement
  - keeps human and AI aligned
- establish ethical guidelines
- set practical constraints and stop conditions
- keep the conversation professional
- use the AI agent in 3 phases
  - brainstorm before the session
  - navigator and scribe during the session
  - write reports after the session

Test charter example

Mission: Explore [feature/workflow] to discover risks related to [quality attributes].

Scope: [in-scope systems, integrations, data boundaries]

Out of Scope: [explicit exclusions]

Risks of Interest: [e.g., data loss, authorisation bypass, pricing errors]

Heuristics: [CRUD, RCRCRC, Goldilocks]

Oracles: [spec consistency, logs, contracts]

Timebox: [e.g., 45 minutes + 15 minutes debrief]

Evidence: [screens, HAR files, AI notes export]

Stop Condition: [N findings or timebox end]

Prompts

communicate clearly and with purpose
vague prompts will result in generic ideas and waste of time
prepare the AI agent for the session
start-of-the session prompt
mid-session refocus
challenge assumptions
AI agent functions as a scribe

Prompts

communicate clearly and with purpose
vague prompts will result in generic ideas and waste of time
prepare the AI agent for the session
start-of-the session prompt
mid-session refocus
challenge assumptions
AI agent functions as a scribe

Prompts

- You are pairing with me as an exploratory testing navigator.
- Priorities: security, data integrity, workflow resilience.
- Product context: B2C e-commerce checkout with multiple coupon combinations and split payments.
- Boundaries: no production data; propose synthetic datasets only.
- Deliverables: provide 2–3 test ideas. For each, include input data, an oracle to check, and the risk the test targets.
- Keep responses concise (<=120 words) unless I say "expand".

communicate clearly and with purpose

vague prompts will result in generic ideas and waste of time

prepare the AI agent for the session

start-of-the session prompt

mid-session refocus

challenge assumptions

AI agent functions as a scribe

Prompts

Charter: Explore checkout with split payments to uncover risks in tax calculation and coupon combinations.

Known weak spots: tax region changes, payment authentication steps, and currency rounding issues.

Please propose:

1) 3 focused tests with concrete input data;

2) The likely oracles I can check quickly;

3) One alternative path I might be overlooking.

Format: Test / Data / Oracle / Risk.

---

communicate clearly and with purpose

---

vague prompts will result in generic ideas and waste of time

---

prepare the AI agent for the session

---

start-of-the session prompt

---

mid-session refocus

---

challenge assumptions

AI agent functions as a scribe

Prompts

We found: (a) rounding mismatch at currency conversion; (b) coupon applies twice after retrying.

Suggest the next 3 tests that maximize learning about these two defects.

Include a hypothesis (why it might fail) and what evidence to capture.

communicate clearly and with purpose

vague prompts will result in generic ideas and waste of time

prepare the AI agent for the session

start-of-the session prompt

mid-session refocus

challenge assumptions

AI agent functions as a scribe

Prompts

List hidden assumptions in your last suggestions and propose a variant that would invalidate each assumption

communicate clearly and with purpose

vague prompts will result in generic ideas and waste of time

prepare the AI agent for the session

start-of-the session prompt

mid-session refocus

challenge assumptions

AI agent functions as a scribe

Prompts

communicate clearly and with purpose

vague prompts will result in generic ideas and waste of time

prepare the AI agent for the session

start-of-the session prompt

mid-session refocus

challenge assumptions

AI agent functions as a scribe

Session snapshot

Tester: Context first. We're testing an e-commerce checkout with split payments and multiple coupon combinations. The latest release changed tax-rounding rules and added retry logic after payment authentication failures.

AI Agent: Understood. Likely risks: coupon state, rounding behaviour, and retry boundaries. I'll track findings and hypotheses as we proceed.

Tester: Yes. Charter is to explore interactions among coupons, currency conversion, and retries, specifically discount correctness and tax accuracy.

Tester: Test one. Two coupons, EUR→USD conversion, then force a payment retry.

AI Agent: Oracle suggestion: coupons should apply only once per checkout session.

Tester: Agreed. After the retry, the second coupon is reapplied. Discount exceeds the allowed limit.

AI Agent: Logging finding. Hypothesis: coupon state is not preserved across the retry boundary.

Tester: Possible, but I want evidence before we lock that in. Capture request IDs and before-and-after totals.

Pitfalls and how to avoid them

- AI hallucinations
- scope drift
- over-automation impulse
- bias anchoring
- documentation overload

I’m convinced. How to start?

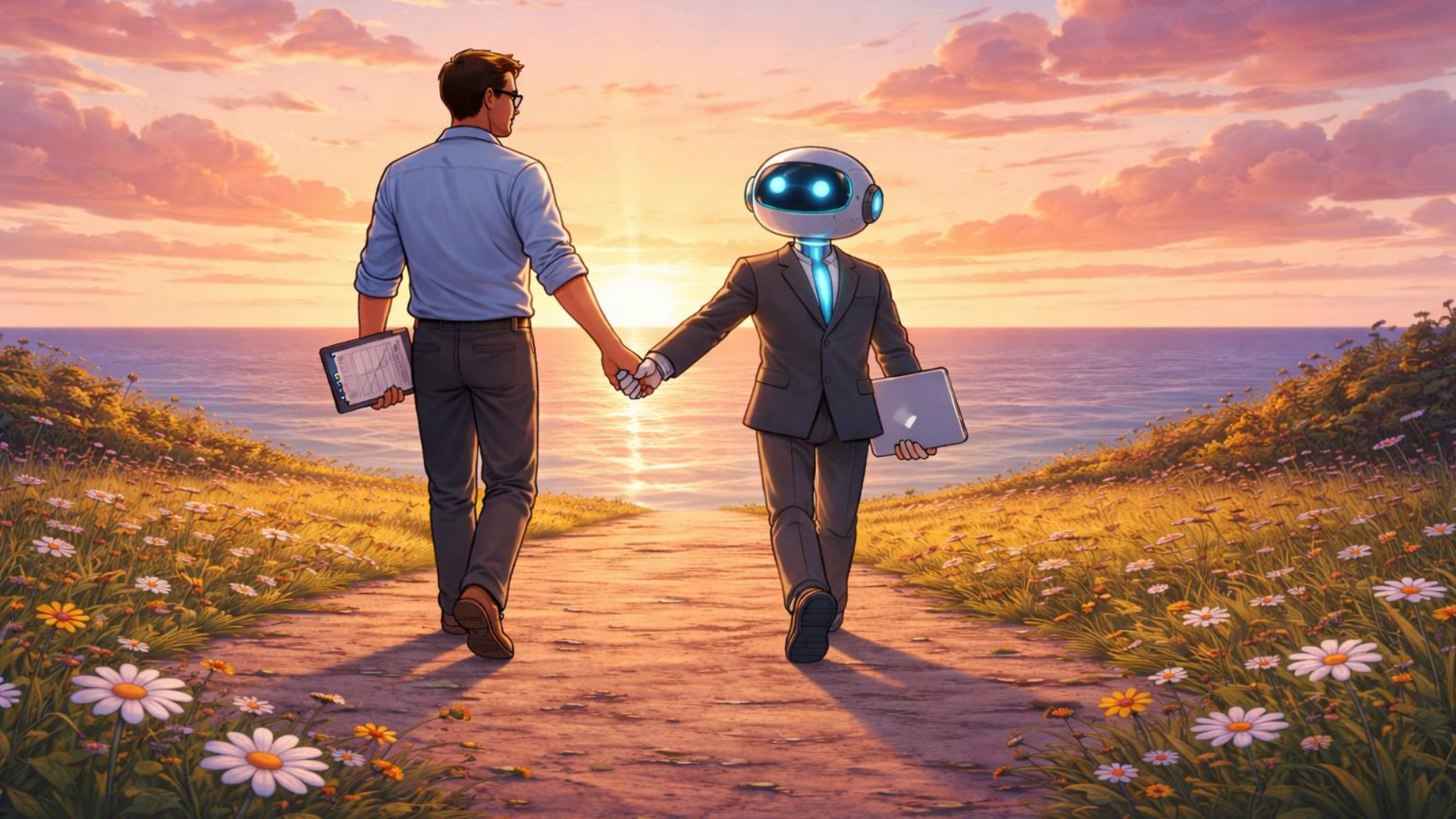
- Make some tests!
- Start small and learn from the GenAI responses
- Note the prompts that work

  … and those that don’t work
- For professionals:

  define agents in markdown files

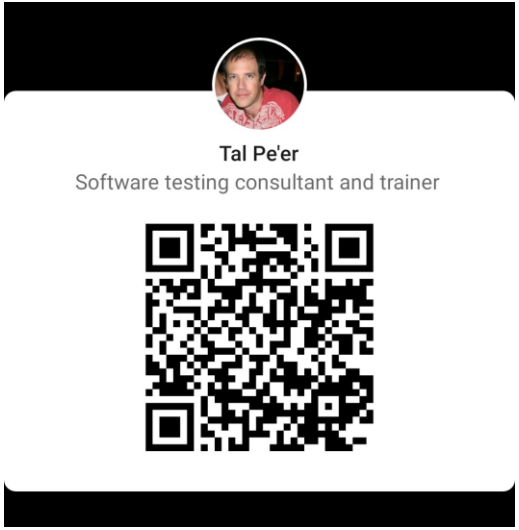
Conclusion

- human ingenuity is the main ingredient for exploratory testing
- exploratory testing thrives on human qualities
  - curiosity
  - adaptability
  - risk-based thinking
- GenAI agents amplify the human tester
  - accelerates ideas generation
  - broadens coverage
  - improves documentation quality



Contact me

Want to know more?
Please contact me at:
tal@grove.co.uk



EST 1989
Grove
software testing

A4Q TESTING SUMMIT

IN COOPERATION WITH:

GASQ

iSQI
SUCCESS IS YOURS

CFTL
Comité Français des Tests Logiciels

skillsclub

GTB
German Testing Board
Software.Testing.Excellence.

ISTQB
International Software Testing Qualifications Board