

# Test-Driven Development for Technology Policy

Neal Parikh, Ginny Fahs, Brandie Nonnecke  
Aspen Tech Policy Hub  
The Aspen Institute

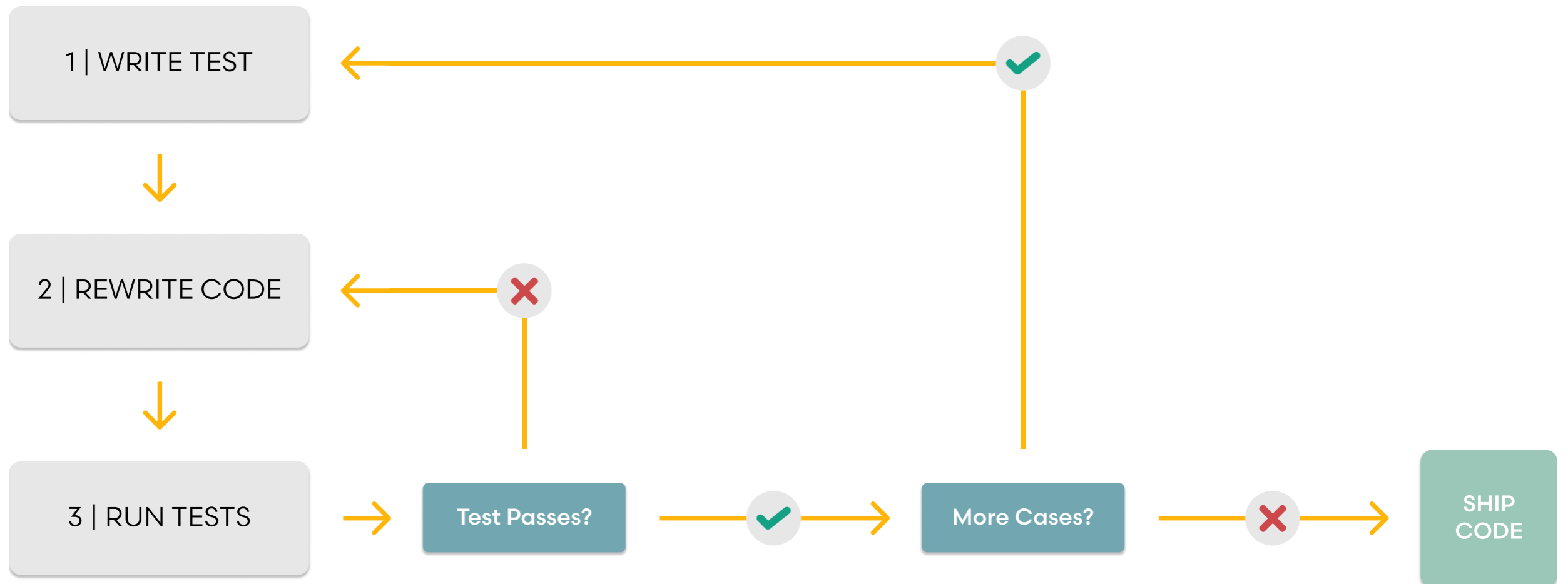
November 2019

# Technology Policymaking

- Goal: Help make technology policy more robust by
  - Ensuring policy actually addresses problem of interest
  - Avoiding unintended consequences
  - Reducing ‘whack-a-mole’ problem where subsequent drafts of policy introduce conflicts with previously agreed upon goal
- Adapt computer science and engineering methods used to build and deploy robust, reliable software to policymaking

# Software Tests

What's Being Tested	Test
Normal US number, format 1	(650) 723-2300 is accepted
Normal US number, format 2	650-723-2300 is accepted
US number in invalid format	65-07-232-300 is rejected
UK number	+44-20-7925-0918 is rejected
Special number	911 is rejected
Non-numeric input	"ABCDEF" is rejected



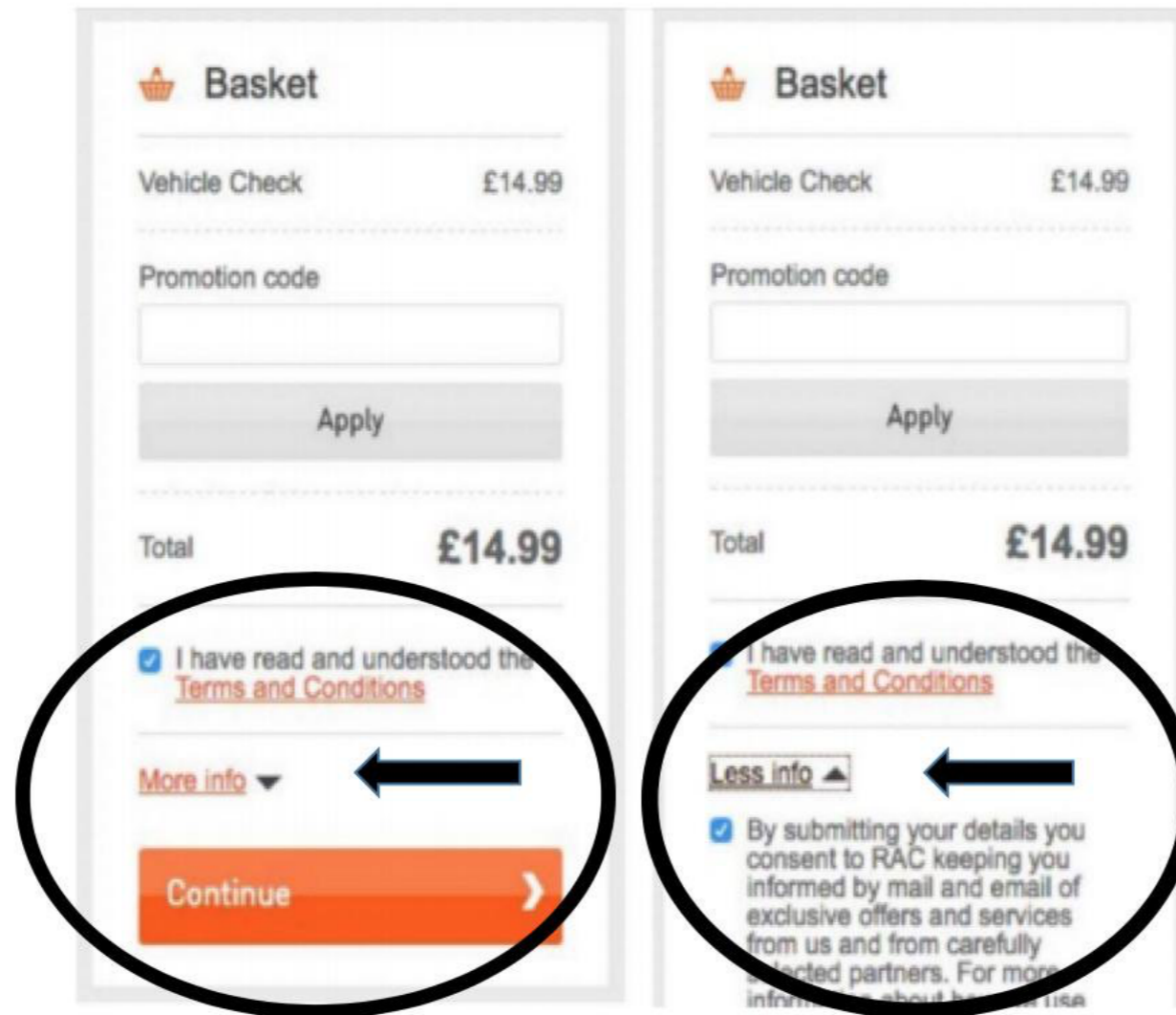
# Testing for Policy

- “Test” = snippet of text that describes concrete situation the policy is **or is not** supposed to address
- “Code” = policy language being developed (legal or less formal language)
- “Running the tests” = policy staff looking through a test suite and comparing to current policy language
- Write tests **first** (feels counterintuitive at first)

# Example

- Goal of policy: ban use of “dark patterns” (manipulative online marketing) by “large online platforms” with over 100 million monthly users
- Deceptive Experiences to Online Users Reduction (DETOUR) Act introduced in 2019

# Testing DETOUR Act



**Potential Test:** don't allow hiding opt-in / consent language in dropdown box

# Testing DETOUR Act



**Potential Test:** don't allow fake hair to make people tap an ad on mobile device



# Testing DETOUR Act

ADVERTISEMENT

 **Cooking** Recipes. Advice. Inspiration.

Subscribe for \$1.25 a week.

[SUBSCRIBE](#)

**Potential Test:** allow normal online banner ad for NYT Cooking

# Designing Tests

- Policy testing requires creativity, as does software testing
- But there are patterns in types of tests that are useful
- Tests can be reused across policies and the policymaker becomes faster with experience

## Test Category

## Questions

## Test Examples

### Numeric Thresholds

- How does the policy change as numbers in the policy are adjusted?
- What happens to entities that pass in/out of the threshold?

Covering online platforms with over 1M vs 100M vs 1B active users

### Subcategory

What are all the distinct subcategories of a category referenced in the policy?

For content moderation policy, nudity could include pornography, historical photos, photojournalism, nudes in art

### Valid Practices

If some behavior is being banned, what is some similar but allowed behavior?

Consider valid examples of online marketing or benign A/B testing in addition to dark patterns

### Demographics

Does the policy make sense as you vary the attributes of the people/companies covered?

- Demographics (age, gender, income, location), access to internet
- Market cap, funding stage, sector

### Jargon

How do different people (engineers, domain experts, laypeople) interpret special jargon in the policy?

DETOUR Act includes term “behavioral experiment”, which could be interpreted differently by engineers, users, social scientists

# Benefits

- Easier for technologists to contribute domain expertise via tests rather than policy language
- Empowers a wider range of people to contribute to the policymaking process
- Many others...

Questions?