

Scaling a small team to build an A/B test flagging platform

LeadDev NYC 2025

Leo Romanovsky



DATADOG

From “Heroics” to Leverage



12 SDKs, 3 engineers,
monthly product asks



Each feature = 12x
reviews + 12x releases



Objective: Scale impact
without scaling
headcount

BEFORE

From duplication

Each SDK re-implements logic

Language specific reviews & QA

Drift over time

AFTER

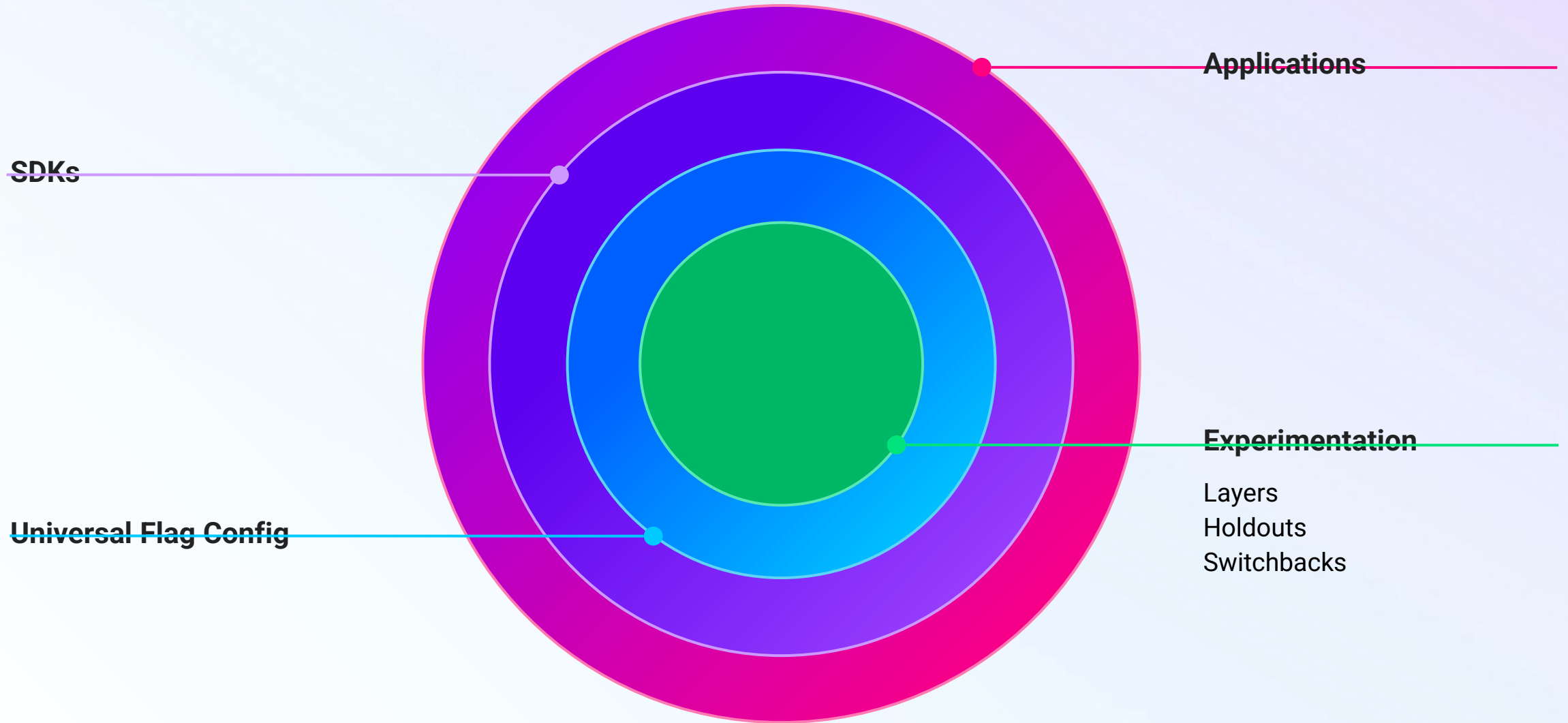
To Leverage

Shared language core

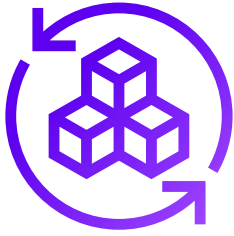
Extensible configuration

Cross-language validation suite

From flags to experiments



A/B testing: A democracy of ideas



Reduce business risk



Accelerate learning



Invest in proven results

Enter Feature Flagging - the Safety Net

Feature Flags: Seatbelts for Product Launches

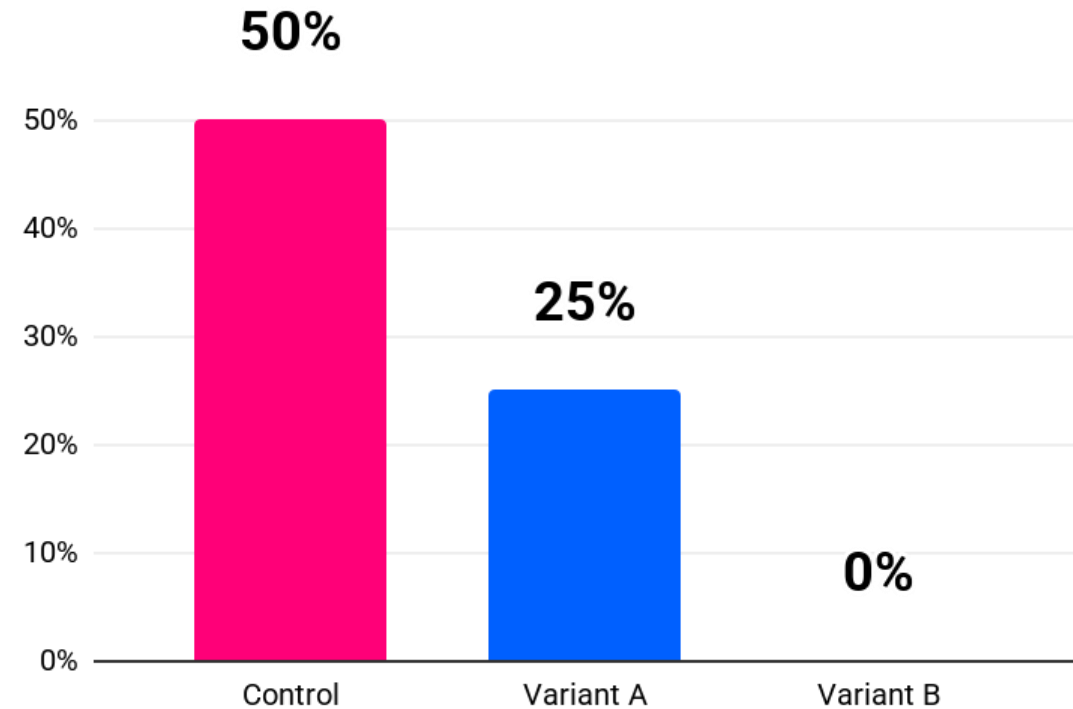
Show new features to small audiences

Rollback instantly if something breaks

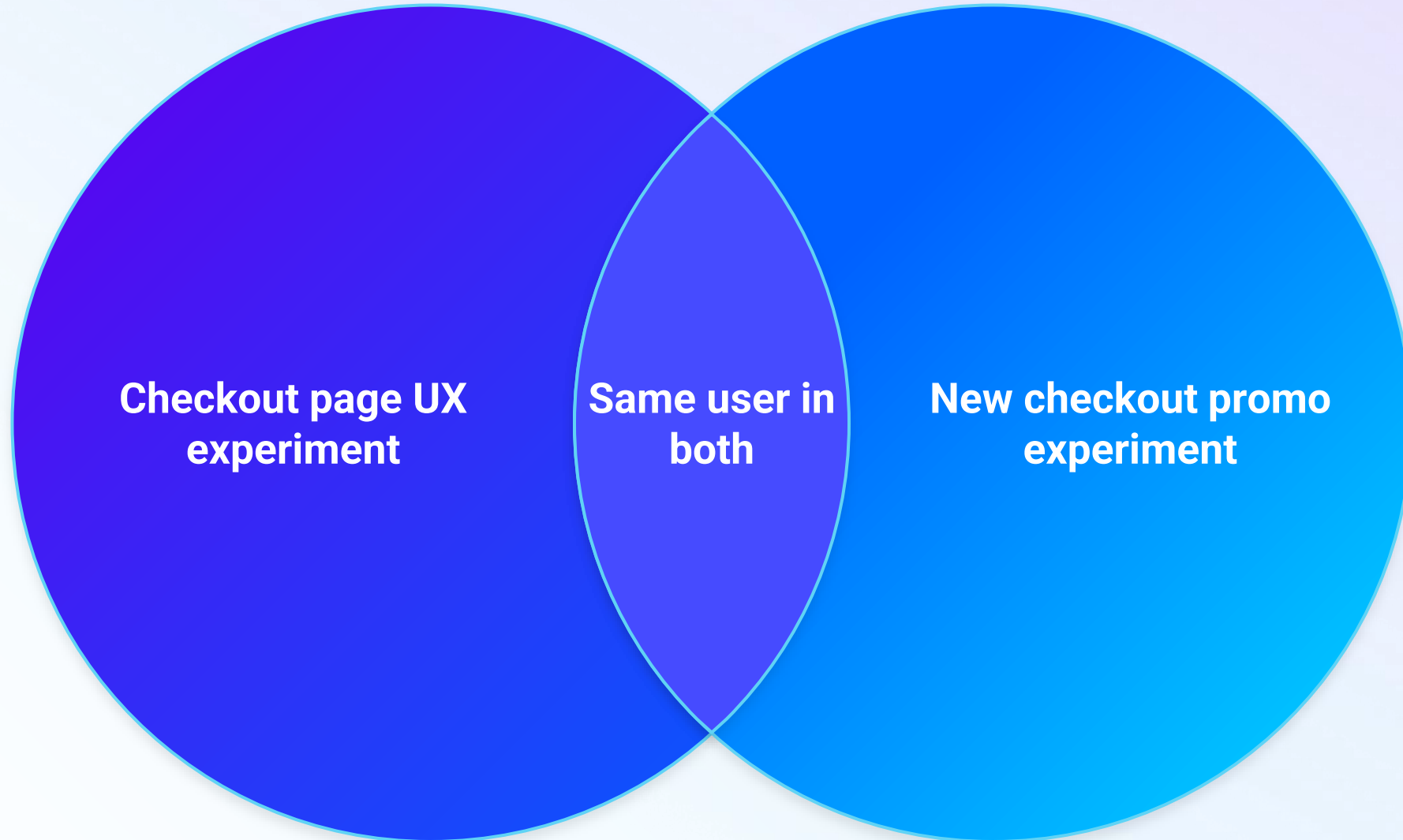
Evaluate consistently across all platforms

The Spanish Checkout Incident

Exposures



Layers Collide: When Two Experiments Fight



Now multiply that by.. 12

12 SDKs walk into a bar



**Java, Python, JS, Swift, Kotlin,
Ruby, Go, PHP, .NET,..**

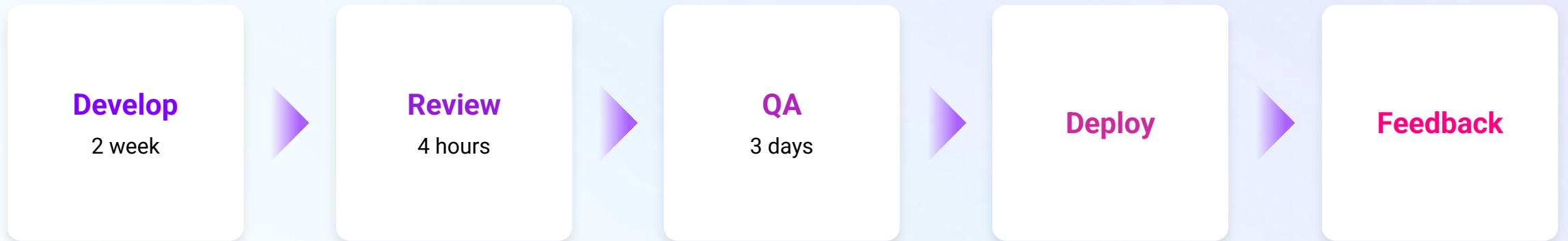


Ruby on Rails, NextJS, Django,...



Each language = own quirks

Scale work, not people



Turning repetition into leverage

Make the Machine Do the Boring Stuff



AI drafts repetitive
code



Engineers review, not
rewrite



Focus shifts to design
and correctness

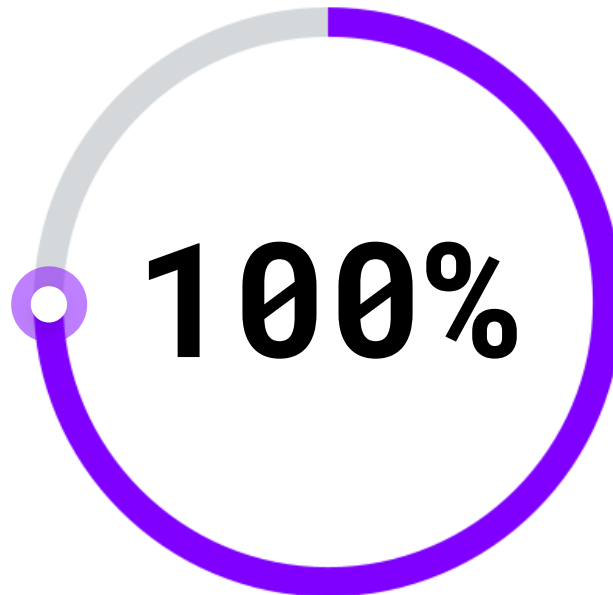
If It's Boring, Automate It



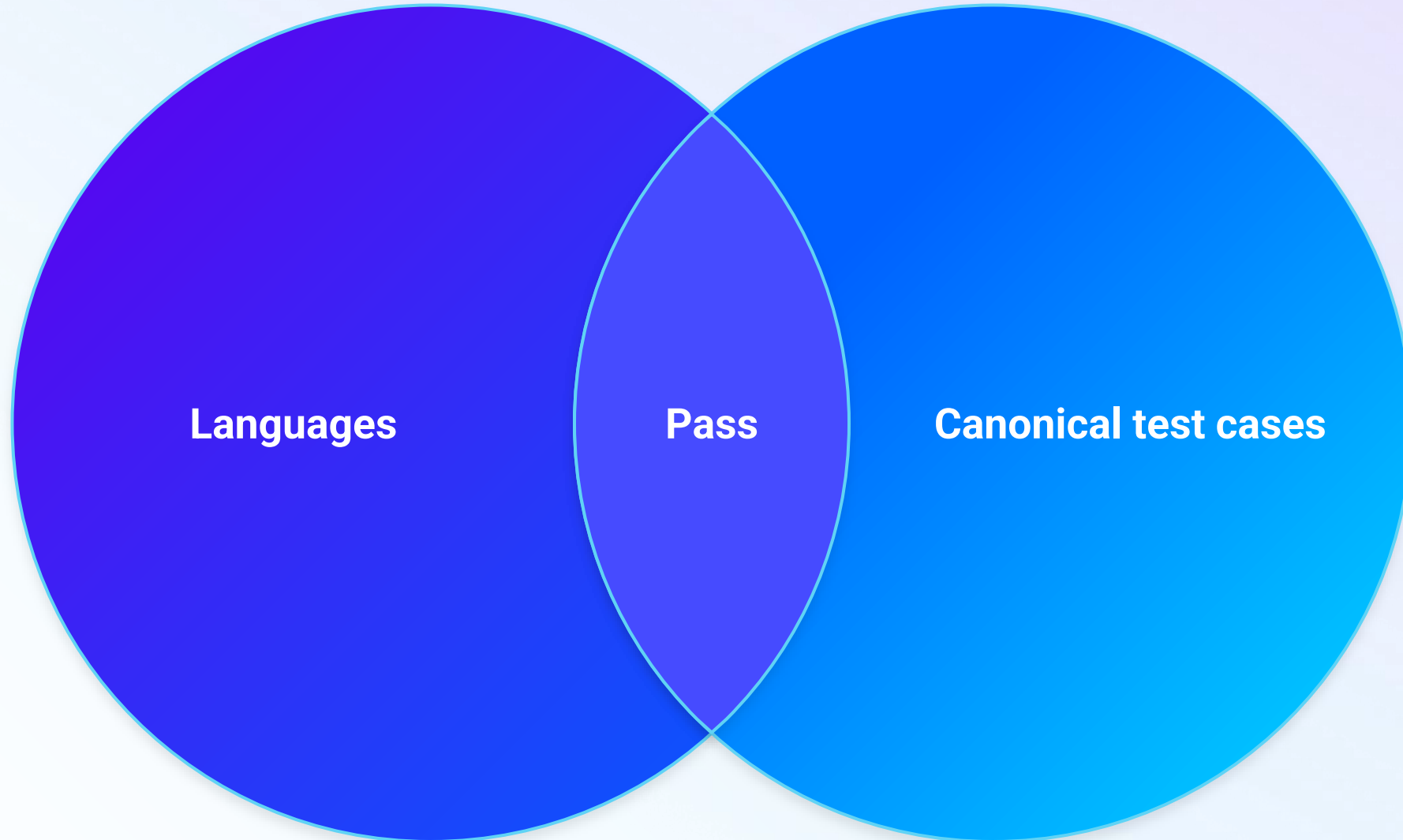
Don't trust, verify

Cross-language validation suite

Coverage



Every language must agree



Backend-Driven: The UFC

The Universal Flag Configuration (UFC)

Backend-driven features that assemble themselves



Single canonical configuration

Defines every flag, rule, and variant



SDKs share a schema

They all know how to build logic from the same building blocks.



New capabilities don't wait on SDK releases

Rollout new evaluation types globally.



Backend-drive

Consistent correctness

UFC in Action

Reusable Building Blocks

Every SDK Understands

Immediately Globally Available

```
"flags": {
  "checkout-new-design": {
    "entityId": 162,
    "key": "checkout-new-design",
    "enabled": true,
    "variationType": "BOOLEAN",
    "totalShards": 10000,
    "variations": {
      "true": {
        "key": "true",
        "value": true
      },
      "false": {
        "key": "false",
        "value": false
      }
    }
  },
  "allocations": [
    {
      "key": "allocation-36649",
      "doLog": false,
      "startAt": "2025-06-25T19:03:29.537Z",
      "endAt": "9999-12-31T00:00:00.000Z",
      "splits": [
        {
          "variationKey": "true",
          "shards": [0, 5000]
        },
        {
          "variationKey": "false",
          "shards": [5000, 10000]
        }
      ]
    },
    {
      "key": "allocation-36650",
      "doLog": false,
      "startAt": "2025-06-25T19:03:29.537Z",
      "endAt": "9999-12-31T00:00:00.000Z",
      "splits": [
        {
          "variationKey": "true",
          "shards": [0, 5000]
        },
        {
          "variationKey": "false",
          "shards": [5000, 10000]
        }
      ]
    }
  ],
  "targetingRules": [
    {
      "attribute": "country",
      "operator": "IN",
      "values": ["US", "CA"]
    },
    {
      "attribute": "appVersion",
      "operator": "SEMVER_GREATER_OR_EQUAL",
      "values": ["3.2.0"]
    }
  ]
}
```

**Humans + AI = Continuous
Feedback**

The Rubber Duck Army



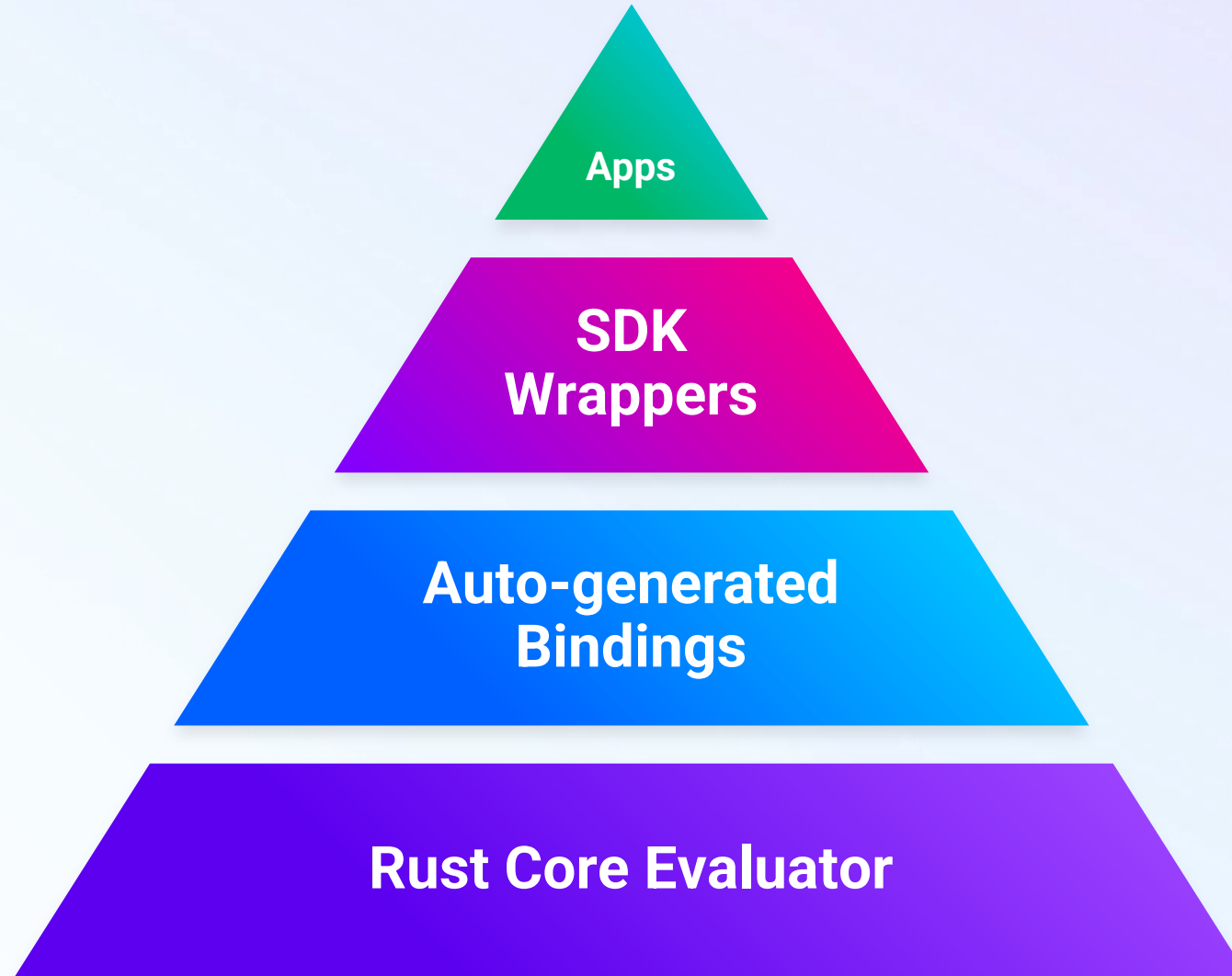
**Engineers correct,
feed examples back**



**Each correction
improves the next
cycle**

Architectural Leverage

One Core to Rule Them All



Humans x AI = Leverage

75% Duplicate code eliminated

8x Faster SDK releases

0 Cross language drift

A11 Engineers focusing on impact, not repetition

Foundations that Make AI Worth It



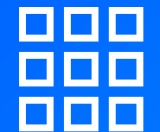
Automate Repetition



Validate across Platforms



Feedback loops compound



Gain leverage with shared Architecture

Thank you

Leo Romanovsky
Staff Engineer, Datadog



DATADOG