KEYNOTE 2

Lessons From the Integration of AI Across Other Industries

Daniel Chu · Chief Product Officer, 23andMe (Formerly Waymo & Google)

March 9, 2026

Policy Issues for Integrating Artificial Intelligence in Cancer Research and Care: A Workshop

National Cancer Policy Forum | Keck Center, Washington, DC

Autonomous Vehicles: Impossibly complex

Early 2010s consensus: Full autonomy was unattainable or very, very far in the future

The core problem: 90/10 problem

- Construction zones with changing lane markers
- Emergency vehicles approaching from any direction
- Unpredictable pedestrian behavior
- Weather, lighting, sensor occlusion

Engineering reality: "Too many variables, too much complexity to solve"

Industry belief: "Incremental driver assistance was the realistic path"

Cancer AI Faces Similar Complexities

Autonomous Vehicles	Cancer AI
Infinite edge cases on open roads	Tumor heterogeneity & rare subtypes
Sensor noise and occlusion	Image artifacts & tissue variability
Real-time decision-making under uncertainty	Diagnostic/treatment decisions with incomplete data
Safety-critical: crashes can be fatal	Safety-critical: misdiagnosis can be fatal
Validation challenge: Can't test every scenario	Validation challenge: Can't validate on every patient population
Regulatory uncertainty	Regulatory uncertainty

Key insight: Both are multifaceted problems, not just AI problems

Waymo's Path: From Impossible to 200 Million Autonomous Miles

The Journey

2009
Google Self-Driving Car Project launched

2015
First fully driverless ride on public roads

2018
Waymo One public robotaxi launch (Phoenix)

2021
San Francisco Trusted Tester Program

2026
Traveled 200M Fully Autonomous

The Safety Outcomes

200M
fully autonomous miles driven

90%
fewer serious injury or worse crashes vs. humans

81%
fewer injury-causing crashes vs. humans

Source: Waymo Safety Impact Data Hub, waymo.com/safety

From Waymo's Journey: What Cancer AI Can Leverage

Five lessons that transfer directly from autonomous vehicles to cancer AI

01

The Proxy Metric Trap

Optimize for the wrong signal and you'll build a system that looks good but fails in the real world

02

Move the Mean, But Engineer for the Tails

Average performance hides catastrophic failure modes at the edges

03

Clean Data & Simulation Are Critical

Real-world data alone can never cover the long tail of rare, high-stakes cases

04

Build a Formal Safety Case

Neither paralysis nor recklessness; structured evidence arguments thread the needle

05

Dynamic Oversight Beats Premature Regulation

Process guardrails before outcome metrics enable responsible scaling

Lesson 1: The Proxy Metric Trap + Human-in-the-Loop Degradation

Some key states where AVs were deployed converged on a headline metric: Miles per disengagement (MPD)

The Broken Metric (MPD)

MPD doesn't measure AI performance accurately. It measures AI + test driver behavior tangled together. A nervous driver disengages often → low MPD. A trained driver who has learned the system's quirks lets more ride → high MPD, with no change in AI quality.

When AI Gets Better, Human Oversight Gets Worse

Human oversight works best when AI performance is moderate. It breaks down at both extremes.

AI Is Struggling

Errors frequent, overwhelming, demoralizing. Humans disengage or override constantly. The AI adds no value and trust collapses.

AI Is Good

The sweet spot. Errors visible and interpretable. Humans stay engaged, calibrated, and corrective.

AI Is Excellent

Errors so rare humans stop seeing them. Vigilance degrades through disuse. When a rare failure arrives, the human backstop has quietly eroded.

The system looks safest precisely when the human backstop has collapsed. Waymo's fix: remove the human from the monitoring loop entirely — shift to remote monitoring and AI anomaly detection so oversight doesn't depend on human vigilance degrading in real time.

The Fix: Clean Metrics + Cancer AI Parallel

Waymo's Clean Metrics

1. **Serious injury crashes per million miles**
2. **Injury crashes per million miles**

Remote monitoring + AI anomaly detection replaced autonomous specialist oversight entirely.

Cancer AI: Same Trap, Different Name

Clinician Override Rates

Similar to MPD, there are elements of human behavior in the metric

The fix: find the clean metric — one with no human behavior confound. Model uncertainty scores that flag cases to review queues independently of clinician choice.

LESSON 2 OF 5

Lesson 2: Move the Mean, But Engineer for the Tails

The hardest tension in safety-critical AI: moving the mean is where you help the most people and get fastest iteration. Engineering for tails is where safety, trust, and deployment viability live. You can't ship if you ignore corner cases—but you can't make progress if you only chase them.

01

Strategic Scoping ("Start in Phoenix")

Waymo chose Phoenix deliberately: consistent weather, grid roads, lower pedestrian density. Reduced tail diversity early to accelerate learning on common cases.

02

Detect ↣ Escalate ↣ Learn

Detect when in a tail case (novelty, uncertainty, ambiguous signals). Escalate explicitly via fleet response. Every escalation updates the system structurally.

03

Expand with Evidence

Then SF/LA: urban density, hills, fog, complex intersections. Each geography introduced new tails—but with a system to handle them.

LESSON 2 OF 5

Where Are Cancer AI's Tails?

Tails aren't just rare diseases. They're any case where the model's training distribution breaks down.

Rare Subtypes

Unusual Presentations

Underrepresented Populations

Imaging Artifacts

Workflow Variation

What's your 'start in Phoenix'? Cancer AI: stage high-volume screenings (breast, lung) first; escalate rares to clinicians with explicit uncertainty flags. Explicit operational design domains — where the model works, where it doesn't, with required escalation paths for tails.

Lesson 3: Data collection and simulation are critical

Building and collecting data for the multimodal sensor suite was critical to creating the dataset required to make the Waymo Driver successful

LiDAR

360° point cloud mapping, detects objects at long range in all conditions

Camera

High-resolution visual data for object recognition, color, and texture

Radar

Velocity and distance sensing, robust in rain, fog, and low visibility

Mapping & Positioning

HD maps + GPS fusion for centimeter-level localization

Audio

Detection of emergency vehicle sirens and environmental audio cues

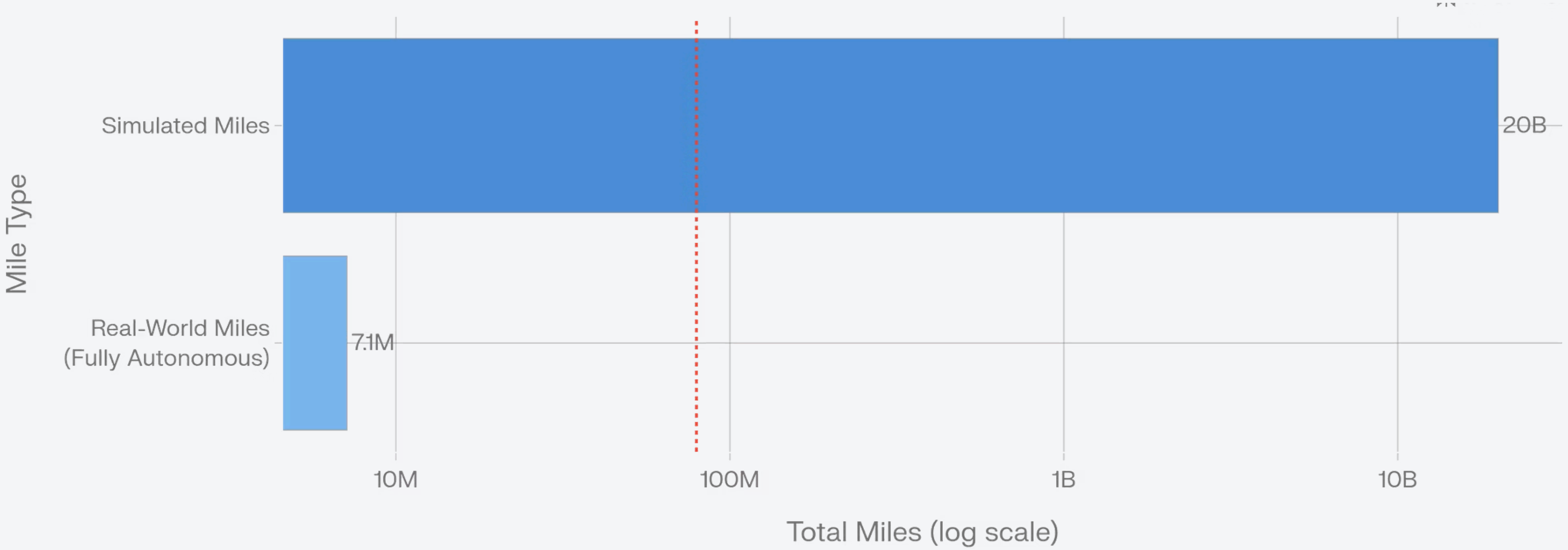
Cancer AI Parallel: The Multimodal Patient Data Problem

Just as Waymo required integrated data from 5 sensor types to build a reliable driver, cancer AI requires integrated multimodal patient data — imaging, pathology, genomics, EHR, and clinical notes — to build reliable diagnostic and treatment models. No single data stream is sufficient. The bottleneck isn't any one modality; it's the infrastructure to fuse them reliably at scale.

Why Real-World Data Alone Is Not Enough

The math of rare events makes real-world-only validation structurally impossible.

Cumulative Waymo Real World Miles vs. Simulated Miles (2023)



You cannot wait for the real world to show you these cases. You have to simulate them. Based on National Highway Traffic Safety Administration data from 2023, fatal crashes happened every 79.4M miles and injury crashes happened every 1.3M miles.

Cancer AI parallel: Very rare cancers

Sources: Waymo (Dec 2023), CA DMV (2024), NHTSA (2023)

Lesson 4: Paralysis vs. Recklessness — Build a Formal Safety Case

Waiting for perfect proof is as dangerous as deploying without safeguards. The discipline of a formal safety case is what threads the needle.

Paralysis

Chicken and egg problem where you can't test enough to prove safety. Indefinite moratoriums. Delays tools that could help patients now

The Responsible Path

Staged pilots → logging → incident review → rollback → scale with evidence.

Recklessness

Rapid scaling without monitoring

LESSON 4 OF 5

What a Formal Safety Case Actually Contains

A safety case is a formal, structured argument — supported by evidence — that a system is acceptably safe in a defined context. Not a one-time checklist. A living document.

Safety Framework

Defines the top-level goal (absence of unreasonable risk), hazard taxonomy, and acceptance criteria for each software release candidate.

Formal Argument

Claims → evidence → acceptance. Covers vehicle architecture, driving behavior, and operational layers. Independently audited before deployment.

Post-Deployment Loop

Real-world evidence retrospectively validates the pre-deployment safety case — and updates confidence for future releases. The cycle never closes.

Cancer AI safety case could include: hazard analysis, defined acceptance criteria, validation in representative populations, audit trails, rollback mechanisms, transparent incident reporting, and ongoing monitoring.

Lesson 5: Dynamic Oversight Beats Premature Regulation

It's too hard to regulate before you know what you're doing. Failure modes only become legible after systems operate in the real world. But "wait and see" isn't the answer either—process requirements and evidence gates can precede full understanding of outcomes.

Arizona · Proactive Engagement

Governor Ducey's Executive Orders (2015, 2018) created a permissive but engaged framework. Focus on transparency reporting + internal safety cases, not premature metrics. **Outcome:** Enabled large-scale Phoenix operations and breakthrough real-world learning.

California · Compliance-Focused

Early focus on specific metrics like disengagements. Regulation splintered across several state bodies. **Outcome:** Many AV companies started in states like Arizona instead.

Federal · Ambiguity

Interpretation of existing regulations. **Outcome:** Flexibility for progress but state-level patchwork — a structural problem that will be worse for cancer AI than it ever was for AVs.

What Dynamic Oversight Actually Looks Like

Process guardrails — not premature outcome metrics — are what enable responsible scaling.

Stage Gates Tied to Evidence

What must be true before expanding from pilot to deployment, from one site to many, from narrow to broad indication? Not time-based — evidence-based.

Continuous Monitoring

Validity and reliability checks as populations and contexts shift. Drift is guaranteed in cancer care, not hypothetical.

Explicit Change Management

When do model updates require re-validation? Re-approval? A new evidence package?

Transparent Reporting

Require measurement, documentation, and monitoring from day one. Separation of builders from validators.

Consent Architecture

23andMe model: explicit informed opt-in, not opt-out by default. Result: >80% research consent rate. Powers discoveries in diverse populations.

Policy framework prioritizing process (safety cases, monitoring, change control) over premature static performance metrics. Avoid state-level patchwork for oncology AI — the structural problem would be far worse than it ever was for AVs.

Applying Lessons to AI in Cancer Care and Research

1

Where are your tails, and what's your escalation path?

Not just rare diseases — rare presentations, underrepresented populations, unusual care settings, edge cases in imaging/workflow, distribution shift as guidelines evolve.

How do you turn operations into compounding learning?

What would make your data consented, high-quality, phenotype-linked, and reusable across multiple research questions over time?

What would make a safety case credible?

What claims, evidence, and monitoring would earn trust from patients, clinicians, and regulators — and how do you keep it current as the system and world change?