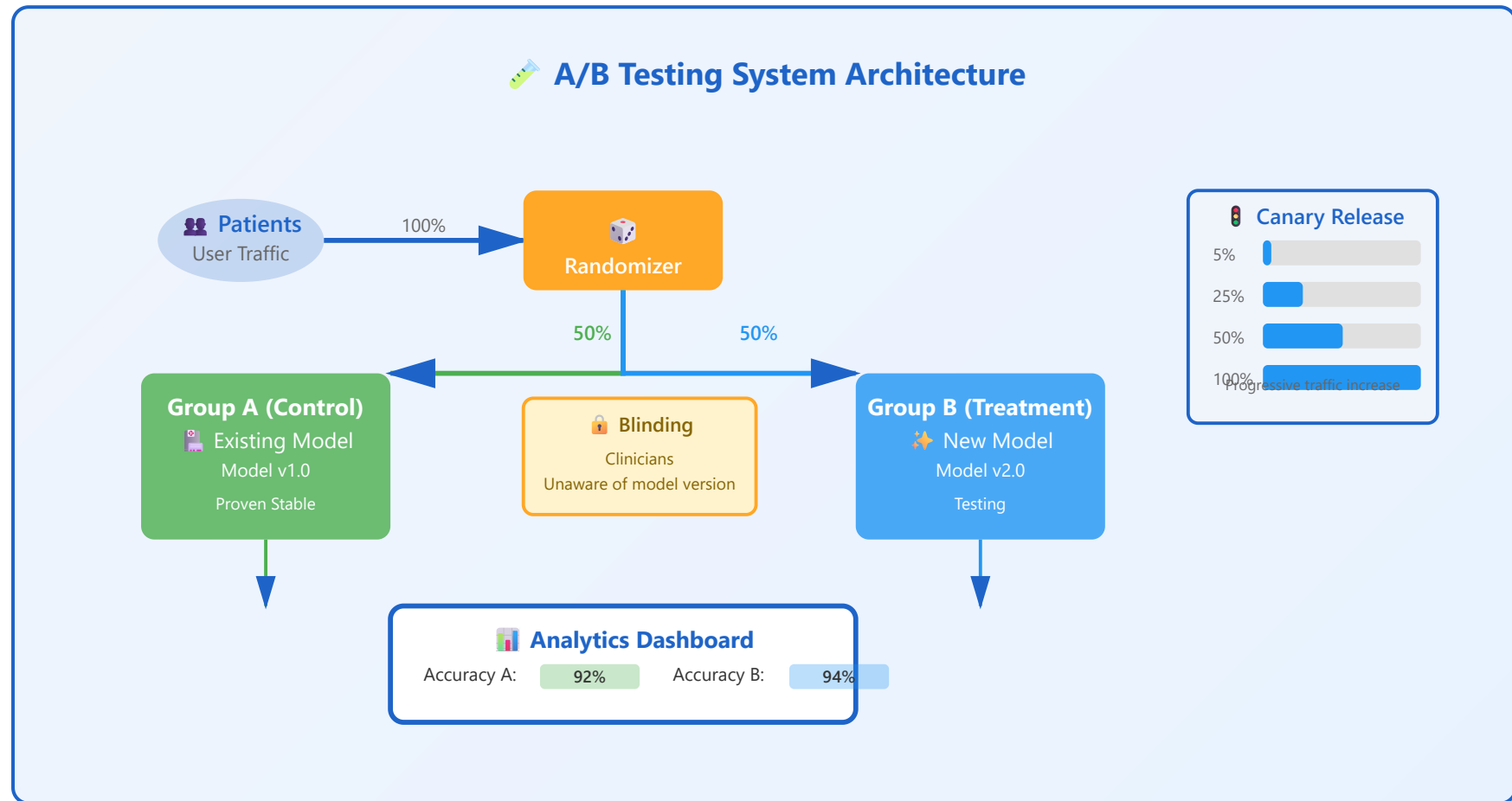


A/B Testing in Production (A/B Testing in Production)



Test Design

- Control Group (A): Existing model (50% traffic)
- Treatment Group (B): New model (50% traffic)
- Randomization: Random assignment per patient

- Blinding: Clinicians unaware of model version



Evaluation Metrics

- Primary Metrics: Diagnostic accuracy, AUC
- Secondary Metrics: Inference time, user satisfaction
- Safety Metrics: False positive/negative rates
- Business Metrics: Cost efficiency



Traffic Split

- Canary Release: 5% → 25% → 50% → 100%
- Gradual Rollout: Gradually increase traffic
- Geographic Split: Regional split testing
- Risk-based Split: Start with low-risk patients