Choosing the right model
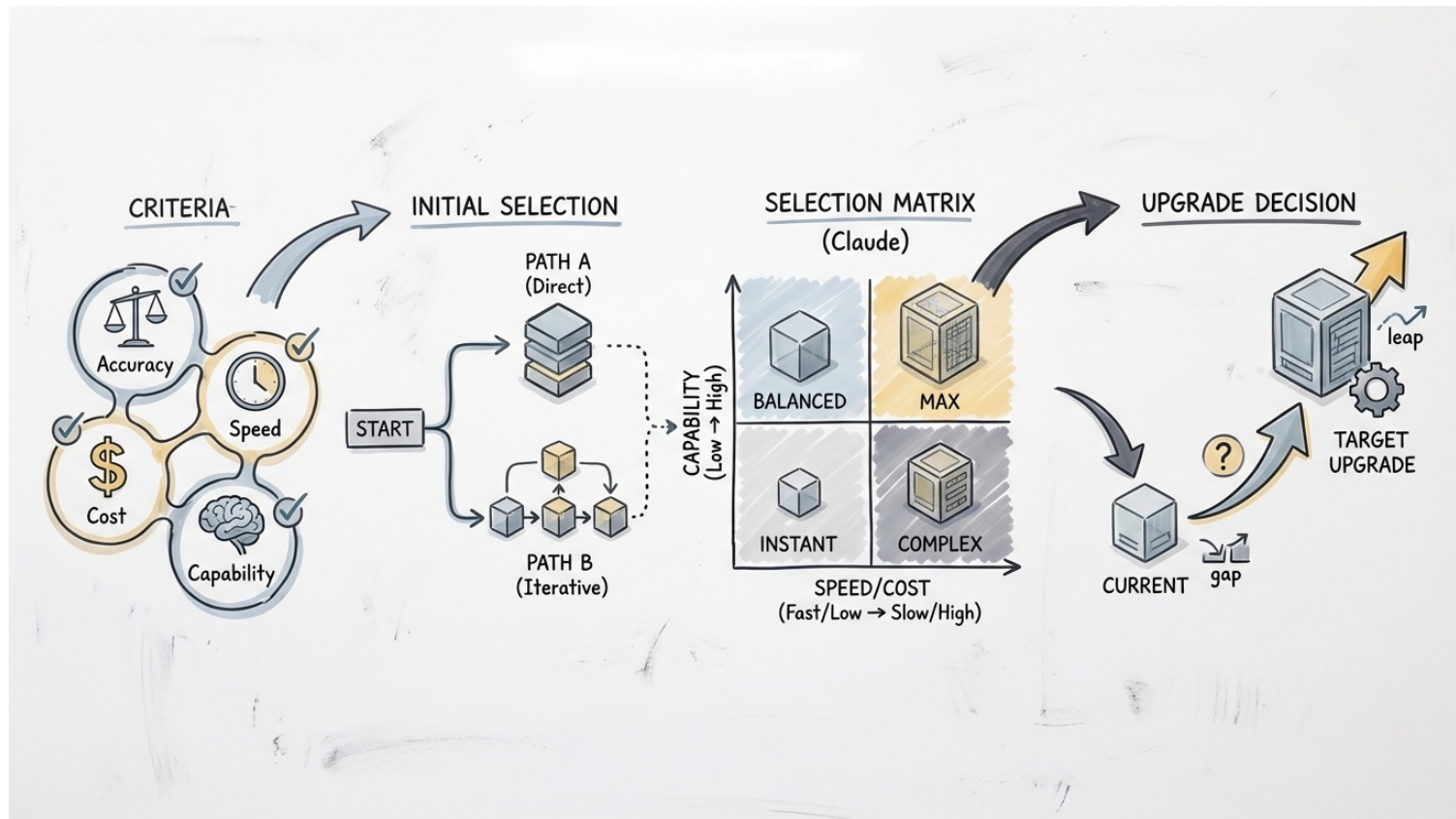
Choosing the right model

Agenda

✓ Introduction

★ Establish Key Selection Criteria

◆ Approaches to Initial Model Selection

→ Claude Model Selection Matrix

• Decide on Model Upgrades

■ Key Takeaways

Introduction

✓ Balance model capabilities with speed and cost.

★ Choose optimal Claude model for your application.

◆ Guide structured decision-making based on requirements.
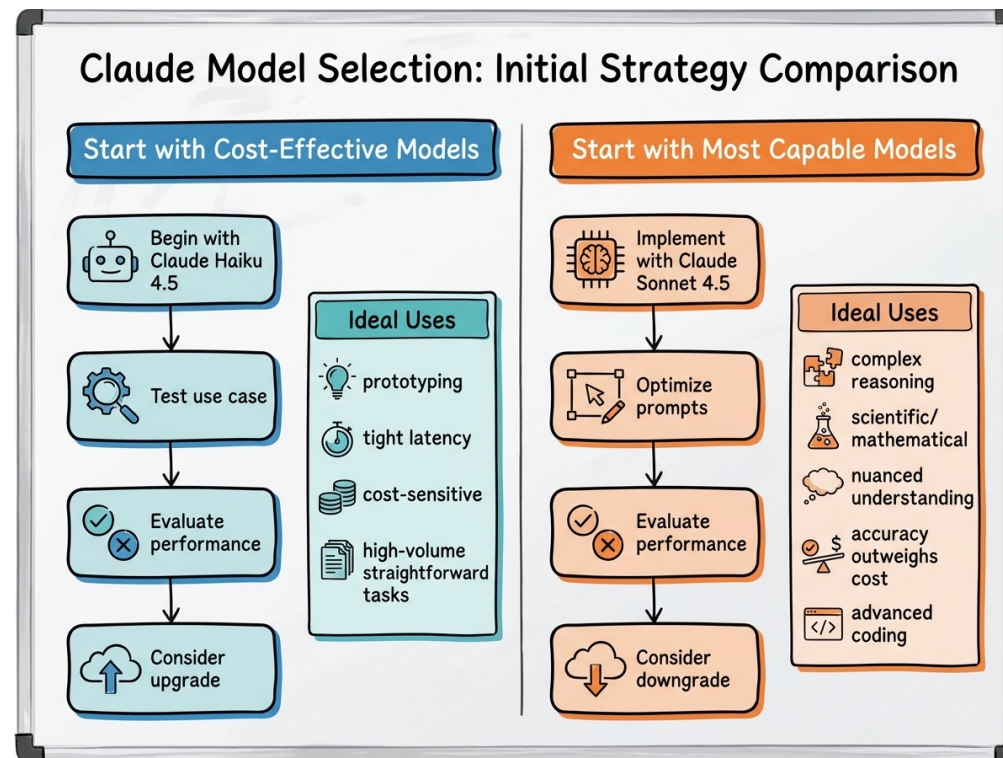
Establish Key Selection Criteria
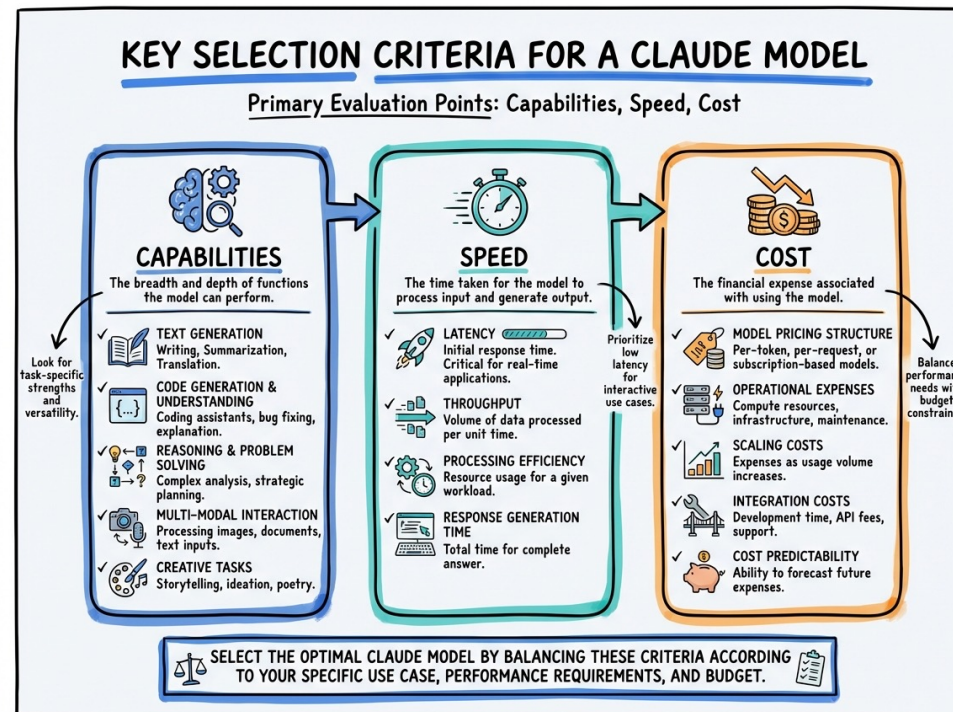
Infographic for Establish Key Selection Criteria

Establish Key Selection Criteria

✓ Define precise model capabilities for application needs.

★ Determine required response speed and processing latency.

◆ Establish clear financial constraints and cost per token.

Approaches to Initial Model Selection
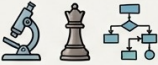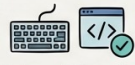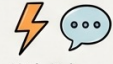
Infographic for Approaches to Initial Model Selection

Approaches to Initial Model Selection

✓ Begin with cost-effective models, e.g., Claude Haiku 4.5.

★ Prioritize rapid iteration and economical development.

◆ Test thoroughly for suitability and initial performance.

→ Upgrade only if specific capability gaps arise.

Claude Model Selection Matrix

CLAUDE MODELS: VISUAL COMPARISON MATRIX (Handwritten Style)

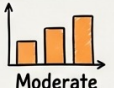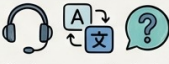|	RECOMMENDED USE	INTELLIGENCE LEVELS (Capabilities)	SPEED (Response Time)	COST (Relative Pricing)	EXAMPLE USE CASES
Opus 4.5	Complex R&D, Deep Analysis, Strategic Planning.	Highest Capability (Deep Reasoning, Nuance).	Slower (Thorough Processing).	Premium. Highest.	Scientific research, advanced problem-solving, policy creation.
Opus 4.1	High-Stakes Tasks, Long-Form Content, Creative Work.	Very High (Strong Contextual Understanding).	Moderate-Slow (Balanced Depth).	High. High.	Legal document review, novel writing, complex coding.
Sonnet 4.5	Balanced Everyday Tasks, Coding Assistants, Content Gen.	High (Solid Logic, Efficient Skill).	Fast (Quick & Responsive).	Moderate. Moderate (Cost-Effective).	Blog post writing, data extraction, standard coding support.
Haiku 4.5	High-Volume, Quick Responses, Real-time Chat.	Moderate (Fast, Efficient, Specialized).	Fastest (Near-Instant).	Lowest. Lowest (Economical).	Customer service bots, live translation, simple Q&A.

Infographic for Claude Model Selection Matrix

Claude Model Selection Matrix

✓ Use Claude Sonnet 4.5 for complex agents, coding.

★ Select Sonnet 4.5 for superior tool orchestration.

◆ Choose Claude Opus 4.5 for maximum intelligence.

→ Apply Opus 4.5 to complex, specialized tasks.
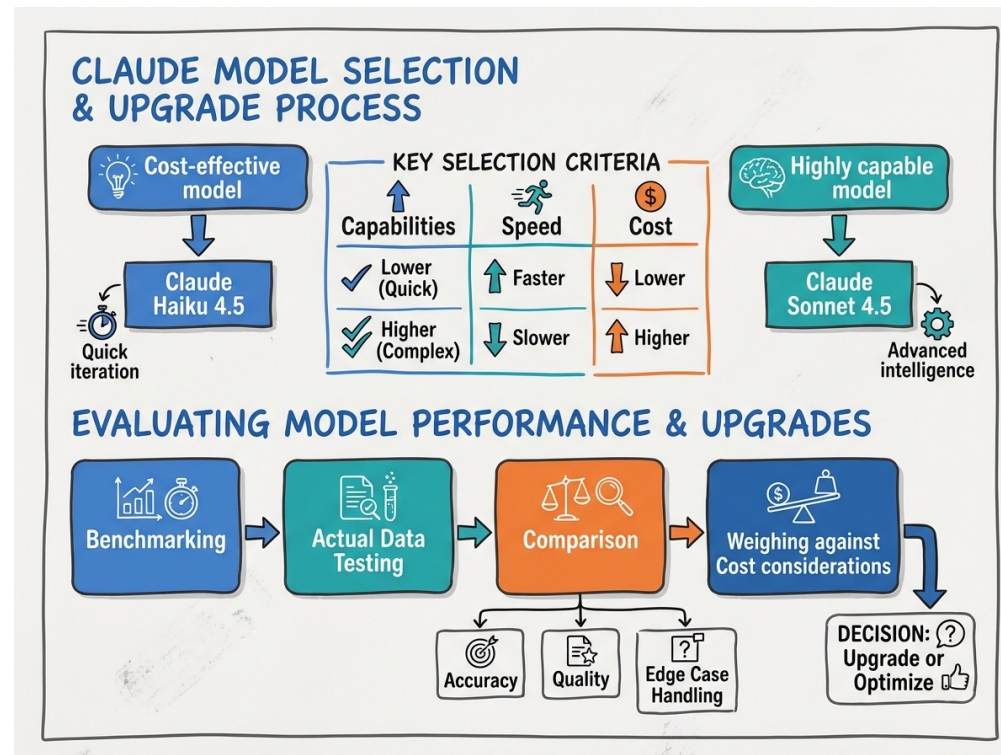
Decide on Model Upgrades
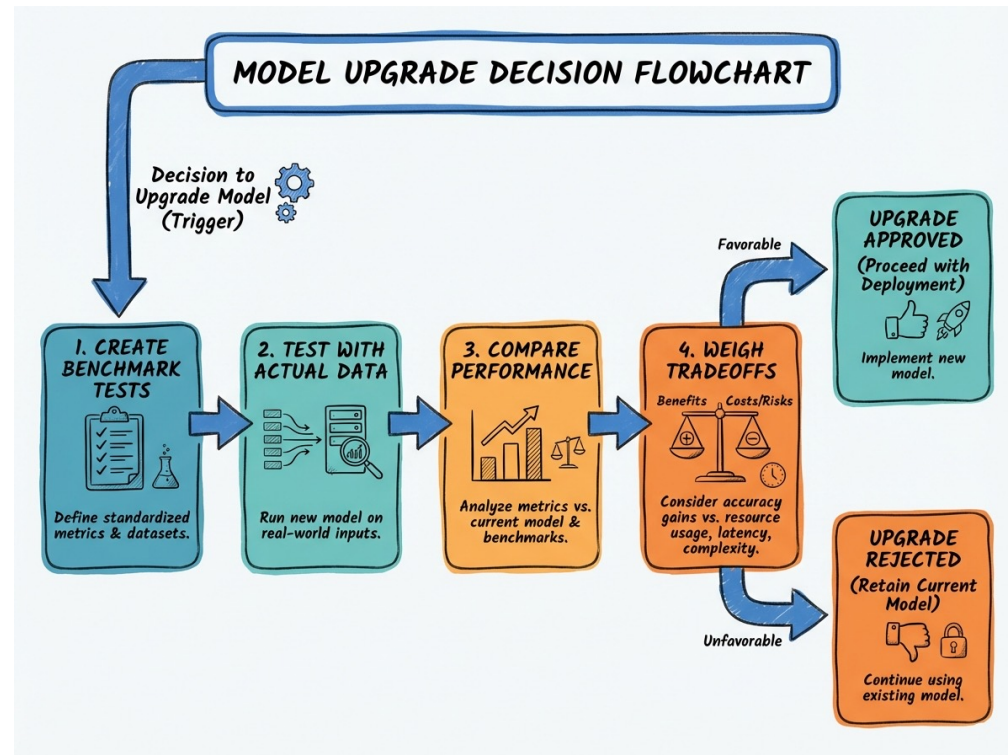
Infographic for Decide on Model Upgrades

Decide on Model Upgrades

✓ Develop specific benchmark tests for your application.

★ Conduct testing using actual prompts and real-world data.

◆ Compare model performance across accuracy and quality.

→ Evaluate handling of edge cases for robust decisions.

Key Takeaways

Infographic for Key Takeaways

Key Takeaways

✓ Balance model capabilities, speed, and cost factors.

★ Start with cost-effective (Haiku) or highly capable (Sonnet).

◆ Conduct rigorous benchmarking with actual data.

→ Compare accuracy, quality, edge cases, and total cost.