How Silicon Valley and Energy Companies Must Both Win for AI to Scale

Jieul Jang, Senior Director @ Hanwha Qcells

Executive Summary

AI's growth depends equally on efficiency gains and renewable deployment

The Efficiency Reality

AI chips deliver **2–3x efficiency gains** per generation, but demand grows **10x faster**

The Energy Scale

Global AI: **176 TWh today** → **400–450 TWh by 2028**

The Speed Solution

Solar + storage: 18 months vs 7–10 years for gas

The Partnership Imperative

Silicon Valley + Energy sector partnership prevents AI power wall

The Paradox

Understanding the fundamental energy challenge facing AI

The AI Energy Reality

One AI Server Rack = 38 Tesla Charges Every Day

120kW

Power Consumption

Drains Tesla Model Y in > 1 hour

38

Daily Tesla Charges

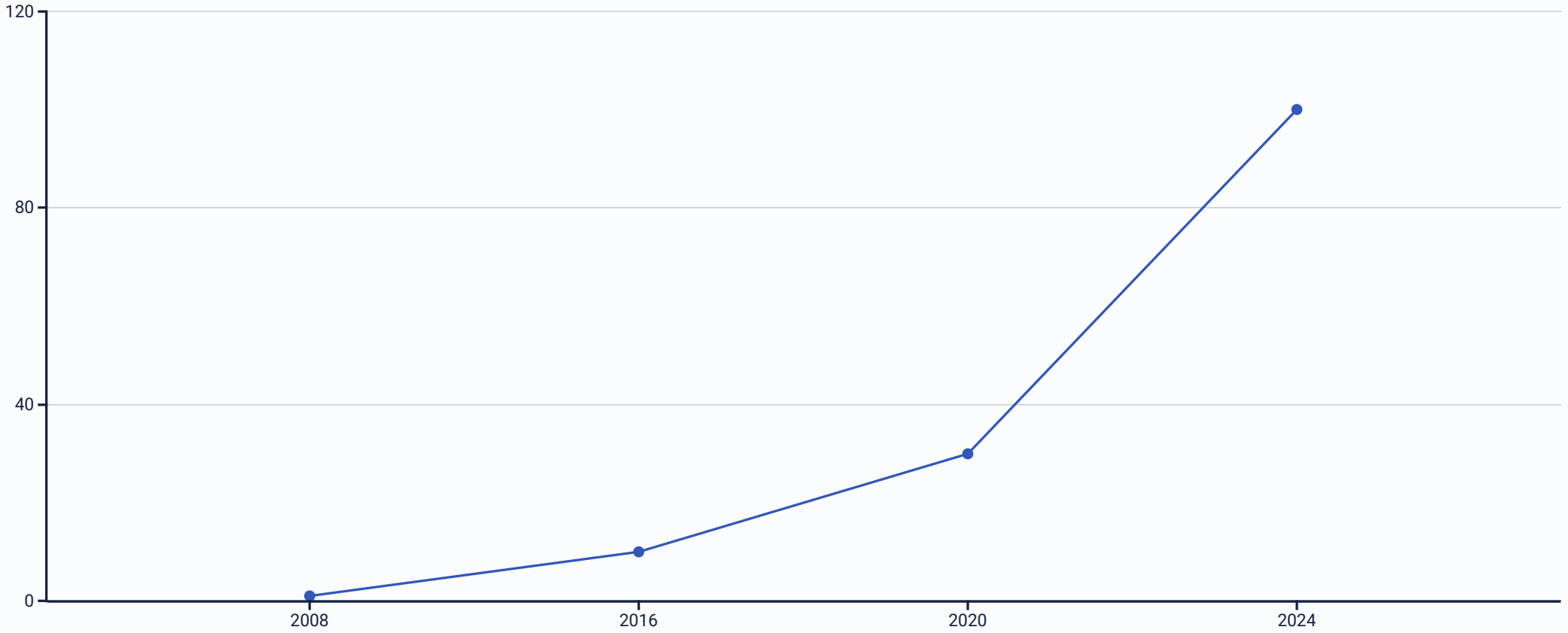
24 hours = 12,000 miles of driving

Even with Blackwell's 2.5x efficiency gain

The Efficiency Success Story

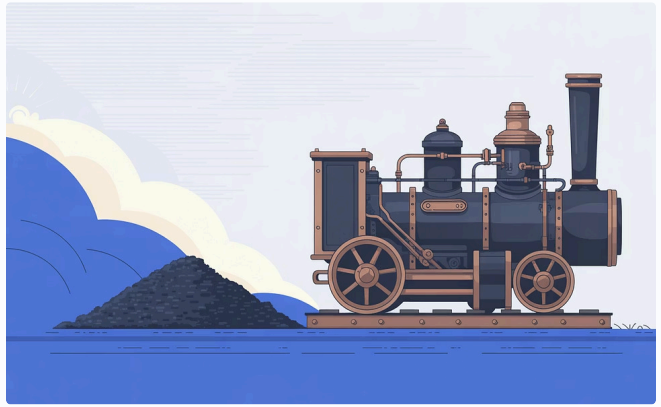
AI Chips: 100x More Efficient Since 2008

Moore's Law for AI is real and accelerating - using ~1% of the energy for the same work compared to 15 years ago.



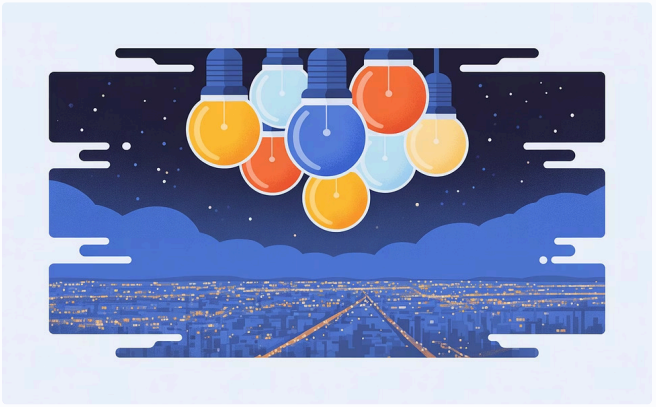
The Jevons Paradox

Why Better Efficiency Leads to MORE Energy Use, Not Less



Coal Engines 1860s

Efficiency doubled → Coal use increased 10x



LEDs 2000s

10x more efficient → Lighting energy use still grew

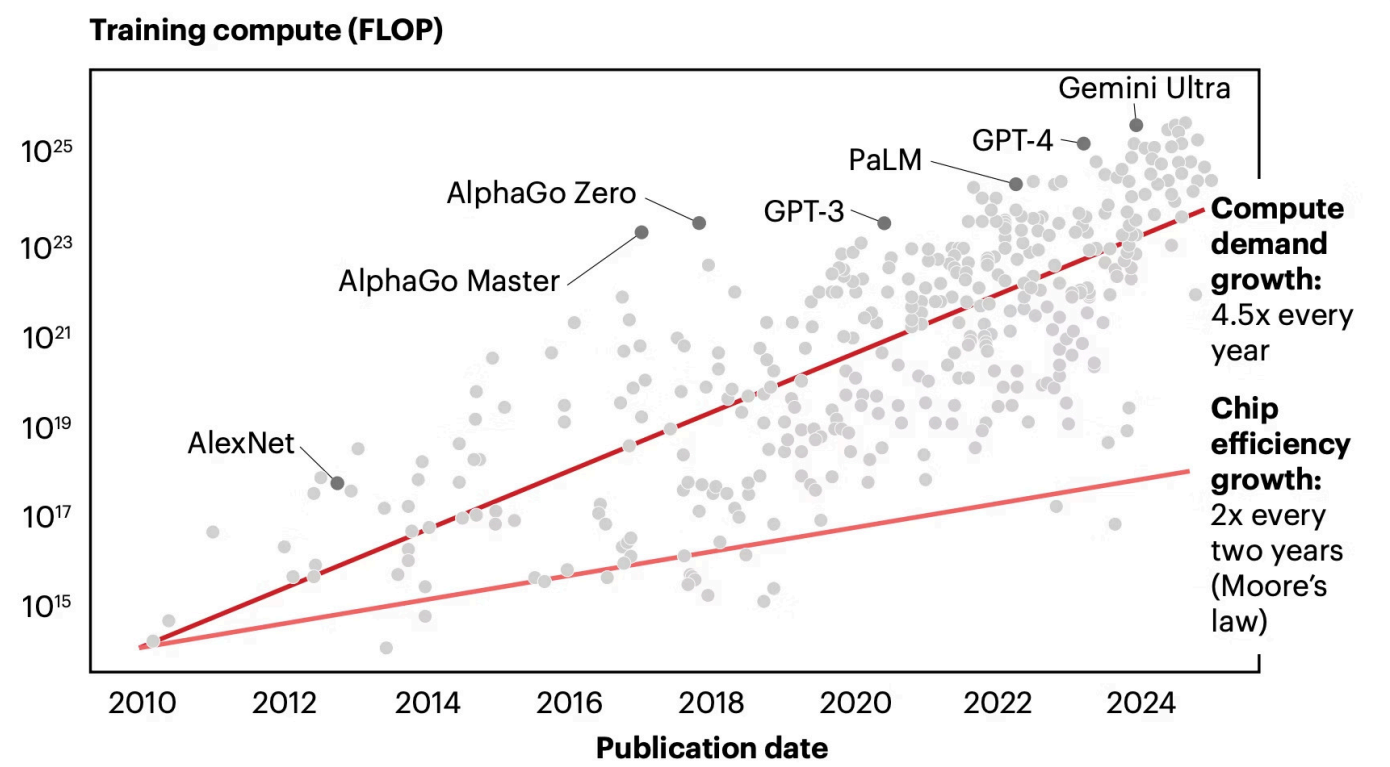
AI Today

99% more efficient → 3x more total energy

Core insight: Efficiency makes AI cheaper → Demand explodes

The Brutal Math

Efficiency Growing 2-3x, But Demand Growing 10x



Notes: Chip efficiency growth not shown to exact scale, with the rate of growth intended to be illustrative; FLOP=floating point operations, which are the number of calculations a system performs
Source: Epoch AI

What This Means

Even WITH Breakthrough Efficiency, We Need 2x Grid Capacity

No Efficiency Gains

10x energy need (impossible to meet)

Current Gains

3x energy need (crisis scenario)

Breakthrough Gains

2x energy need (still massive challenge)

The Two-Front Battle

Part 1: Understanding what's happening on both the technology and energy fronts

The Tech Front - What's Actually Happening

Silicon Valley IS Delivering Efficiency Gains

Nvidia Blackwell

2.5x performance/watt improvement over previous generation

Google TPUv5

3x improvement in AI training efficiency

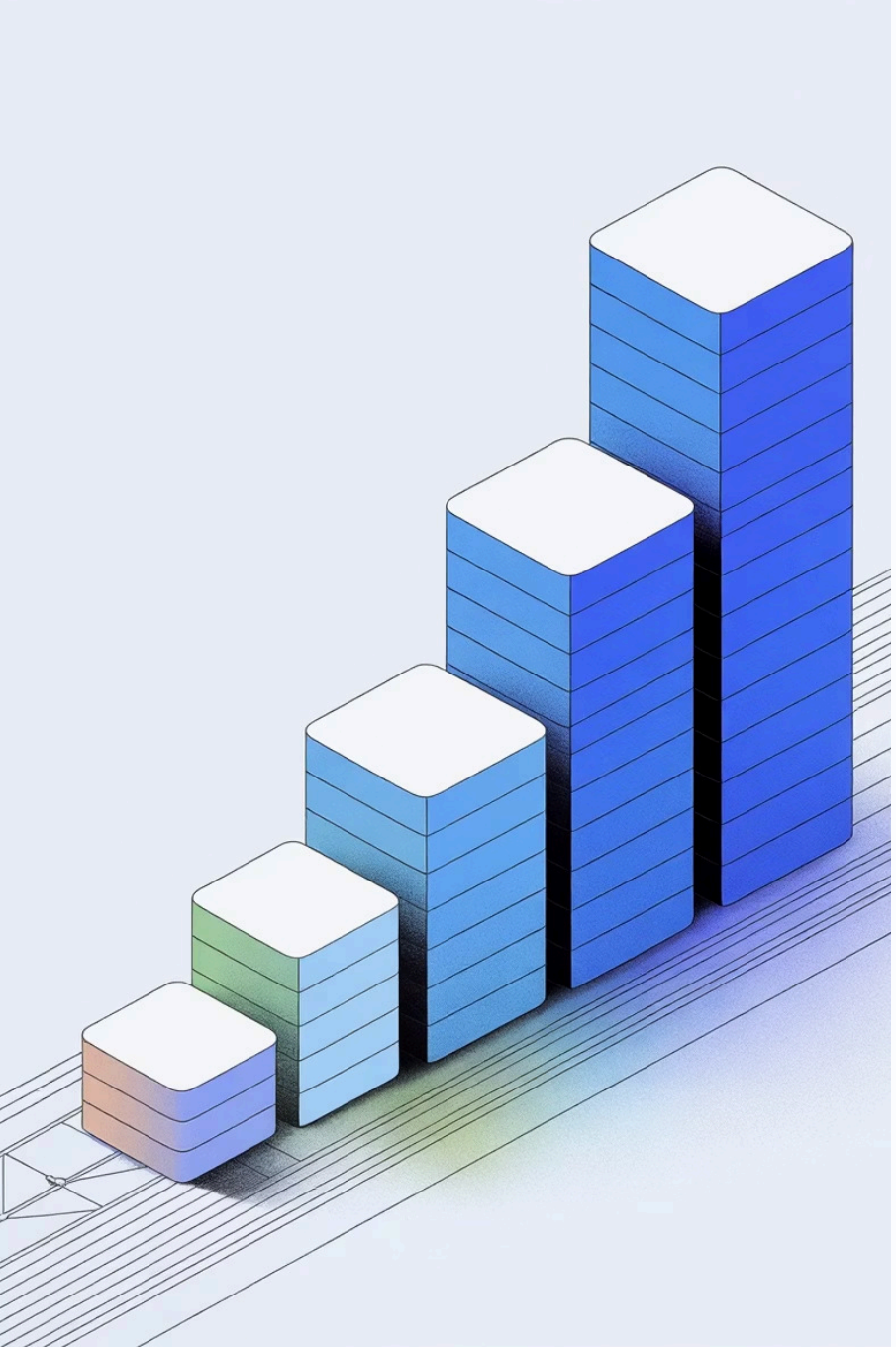
Liquid Cooling

40% PUE improvement in data center operations

Edge Computing

20-30% reduction in central load for certain workloads

Real progress, just not enough to offset 10x demand growth



But Models Keep Growing

Each Efficiency Gain Enables LARGER Models, Not Lower Power

GPT-3

175 billion parameters

GPT-4

1.7 trillion parameters

GPT-5 (Estimate)

10-50 trillion parameters

We use efficiency to do more, not use less

The Demand Explosion

From 0 to 200 Million ChatGPT Users in 2 Years

100M

Users in 2 Months

Fastest adoption in history

200M+

Users by Year 2

Every company adding AI features

100x

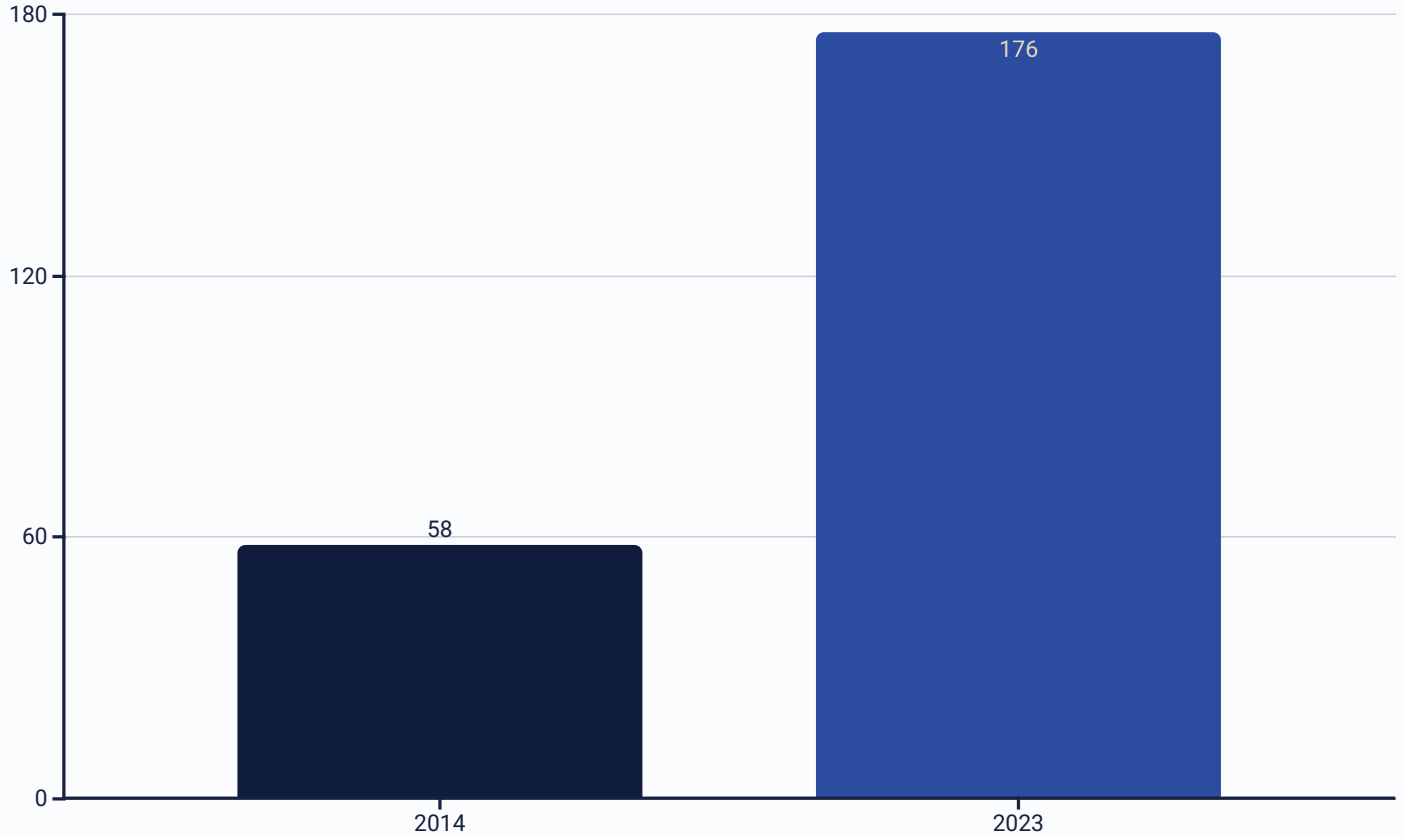
Power Jump

Text → Image → Video compute cost

Democratization drives exponential consumption growth

Current State - The Starting Line

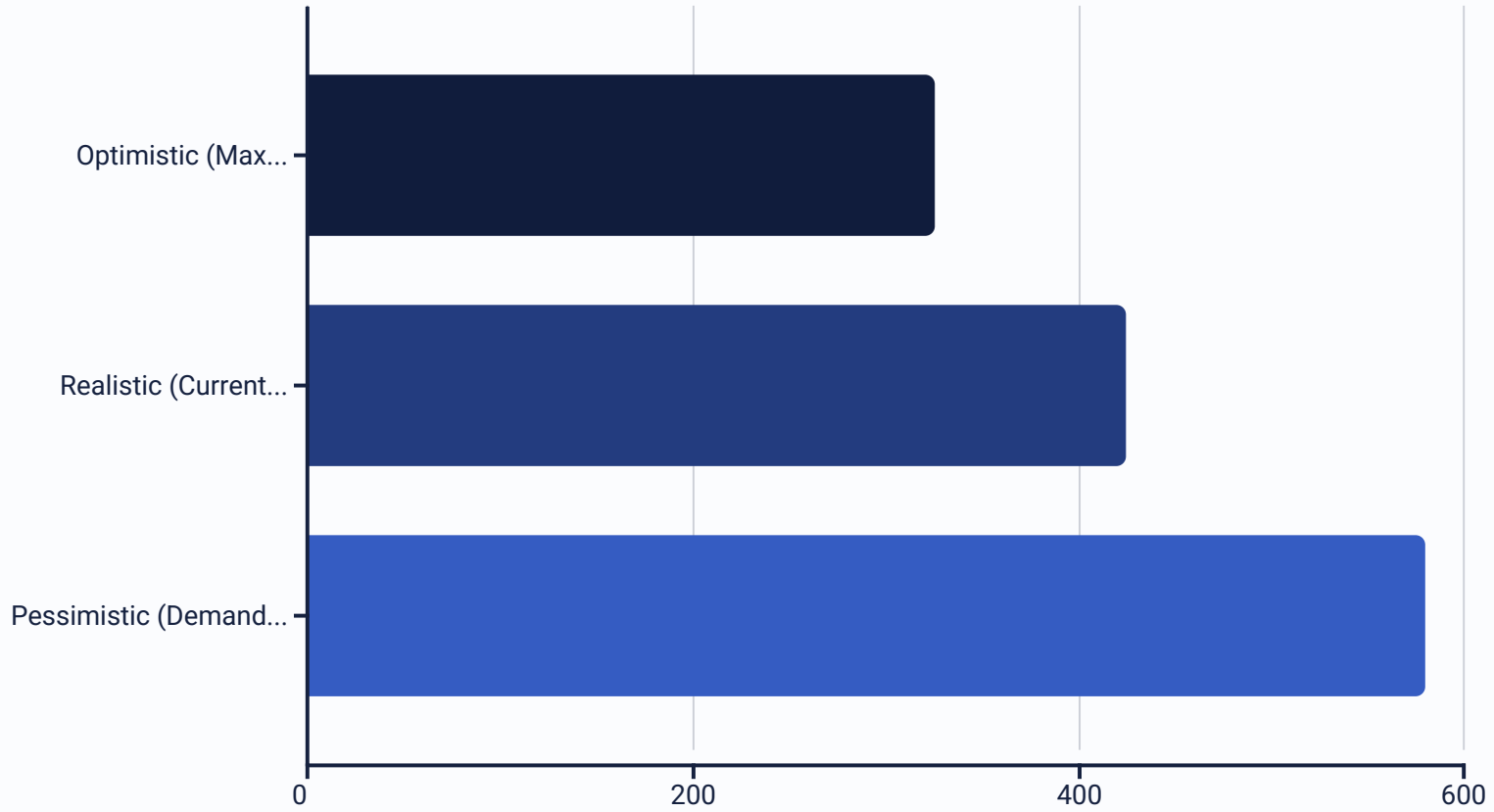
176 TWh Today = Larger than New York State's Entire Electricity Use



Current AI consumption represents **4.4% of U.S. electricity** - this is AFTER a decade of efficiency gains

The 2028 Reality

Most Realistic Scenario: 400-450 TWh



Even the optimistic scenario equals adding **Texas-sized electricity demand** to the grid

Why Speed Matters More Than Ever

Part 2: The critical timeline mismatch between AI needs and energy solutions

The Infrastructure Reality

Tech Moves in Quarters, Energy Moves in Decades

0.5

New AI Model

6 months development cycle

2

New Chip Generation

2 years from design to production

1.5

New Solar Farm

18 months from decision to operation

8.5

New Gas Plant

7-10 years total timeline

10

New Nuclear Plant

10+ years development and construction

The Turbine Crisis

Gas Turbines: 5-7 Year Wait Just for Equipment

Order Today

Get turbine delivery in 2030-2032

Construction Phase

Additional 30 months for plant construction

Online Date

Earliest operation: 2032-2035

AI needs power by 2027, not 2035

Solar's Speed Advantage

18 Months from Decision to Electrons

Permitting Phase

3-6 months for regulatory approvals and site preparation

Construction Phase

9-12 months for installation and commissioning

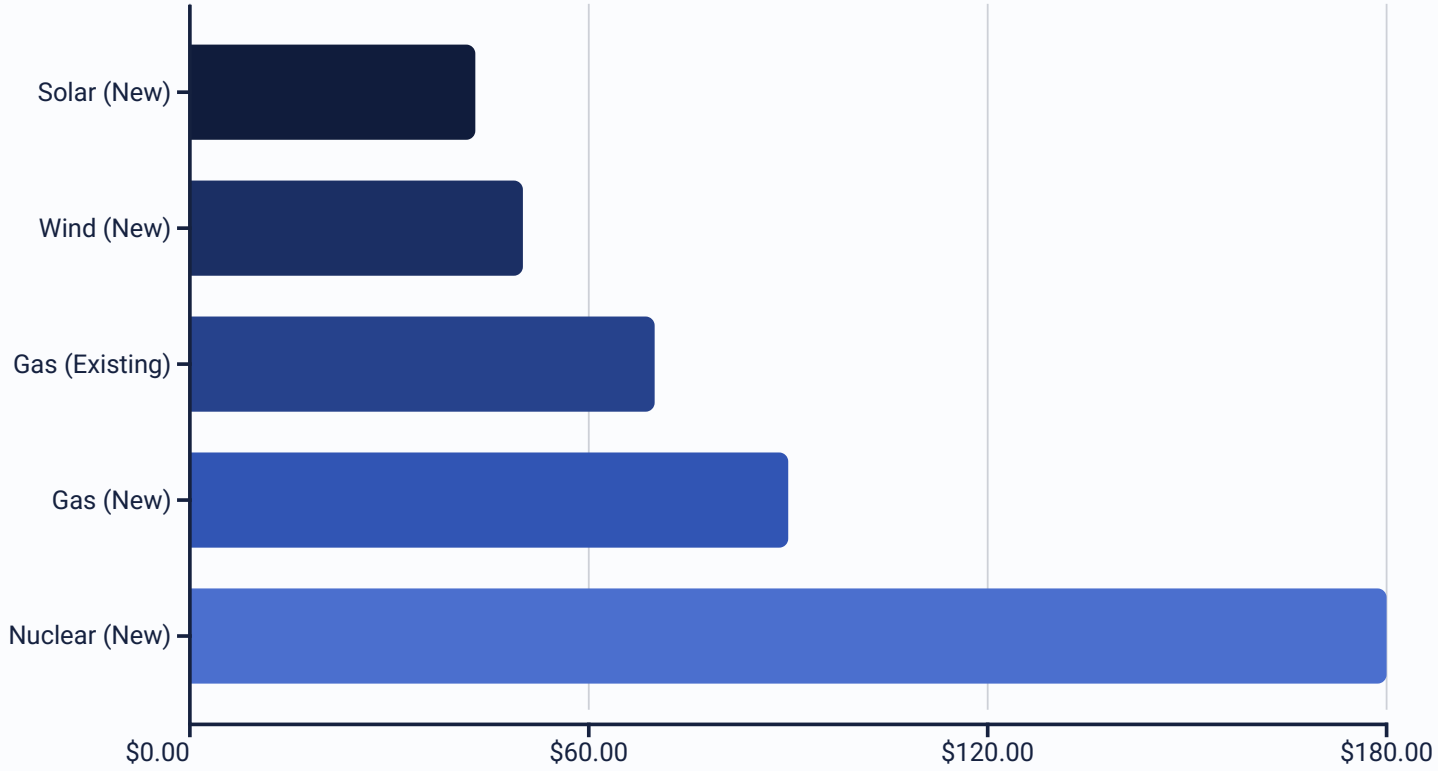
Grid Connection

3 months for interconnection and testing

The ONLY option that matches AI's timeline demands

The Cost Reality

Even With Efficiency, Cheaper to Build New Solar Than Run Existing Coal



Economics and speed align perfectly for renewable deployment

What Tech Companies Are Actually Doing

Part 3: Real-world strategies from hyperscale companies

The Hyperscaler Strategy

Buy Everything: Renewable PPAs + Nuclear + Efficiency

Amazon Web Services

34 GW renewables contracted + nuclear SMR investments for comprehensive portfolio approach

Microsoft

23 GW renewables + Three Mile Island nuclear restart partnership

Google

24/7 carbon-free energy by 2030 across all operations and data centers

Meta

18 GW renewables contracted + on-site generation pilots

The Innovation Sprint

VC Money Flooding Into BOTH Efficiency and Generation

$8B	$2B	$4.2B	$2.8B
Chip Efficiency	**Cooling Technology**	**Energy Storage**	**Grid Technology**
Investment in next-generation processors and AI accelerators	Advanced liquid cooling and immersion cooling systems	Long-duration batteries and grid-scale storage solutions	Smart grid software and transmission optimization

Two-front war requires two-front investment across the entire stack

Time-Shifting - The Smart Compromise

Running AI Training When the Sun Shines

40%

Workload Shifting

Google shifts this percentage of workloads to high-renewable periods

20%

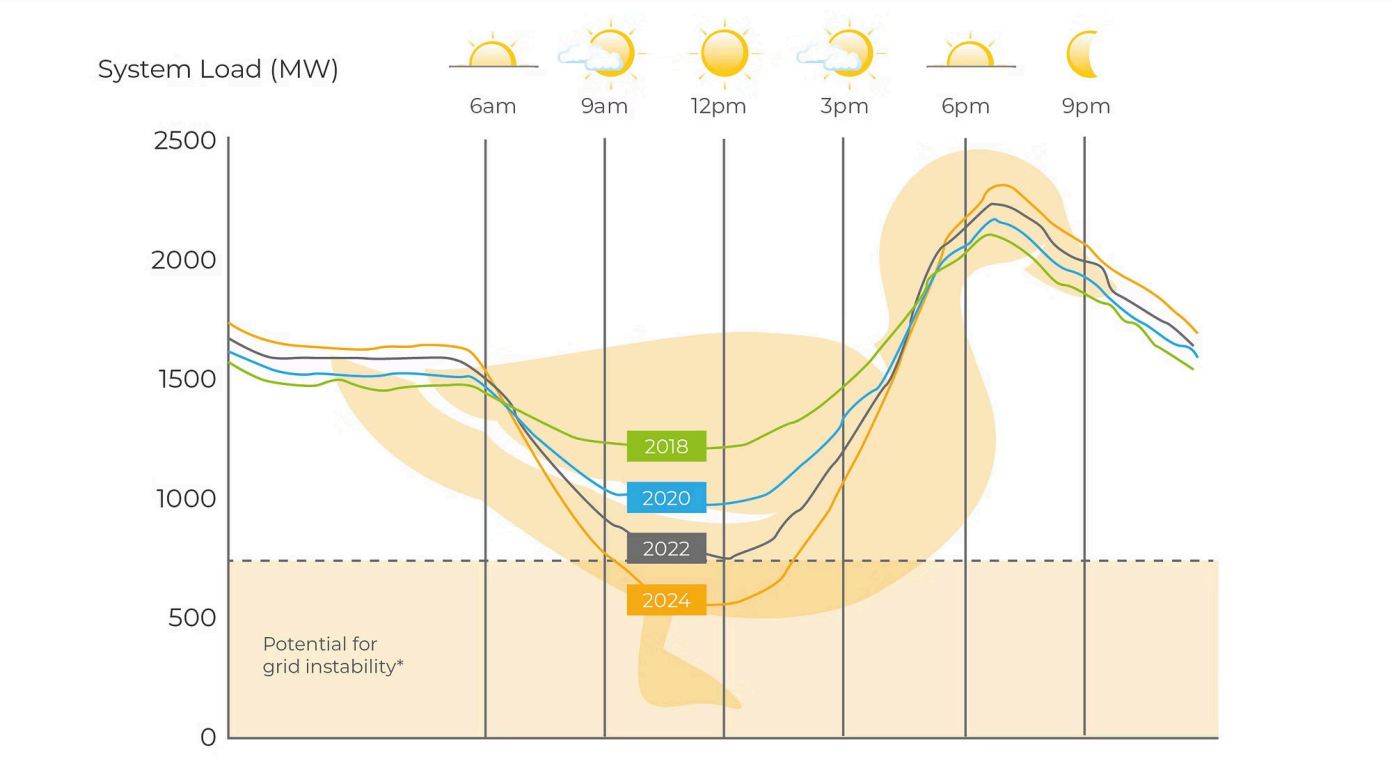
Energy Reduction

Achievable reduction with no new capacity investment required

But this strategy only works for training, not real-time inference

The Duck Curve Problem Remains

AI Inference Runs 24/7, Solar Doesn't



Peak Solar Generation

12:00 PM - Maximum renewable output during midday hours

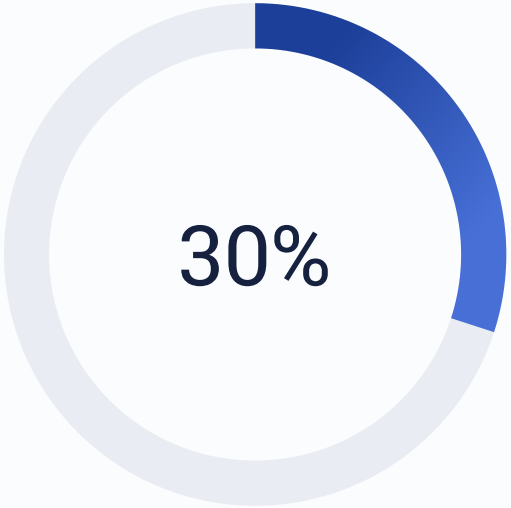
Peak AI Inference Demand

6:00-10:00 PM - When users are most active, solar is unavailable

Batteries needed but current 4-hour duration isn't sufficient for this gap

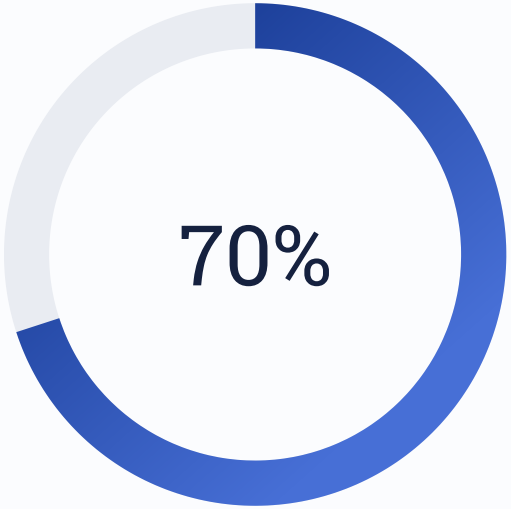
Why Distributed Isn't Enough

Edge Computing Helps but Doesn't Solve the Problem



Maximum Load Reduction

Potential reduction with aggressive edge deployment



Still Centralized

Training and large model inference remain in hyperscale data centers

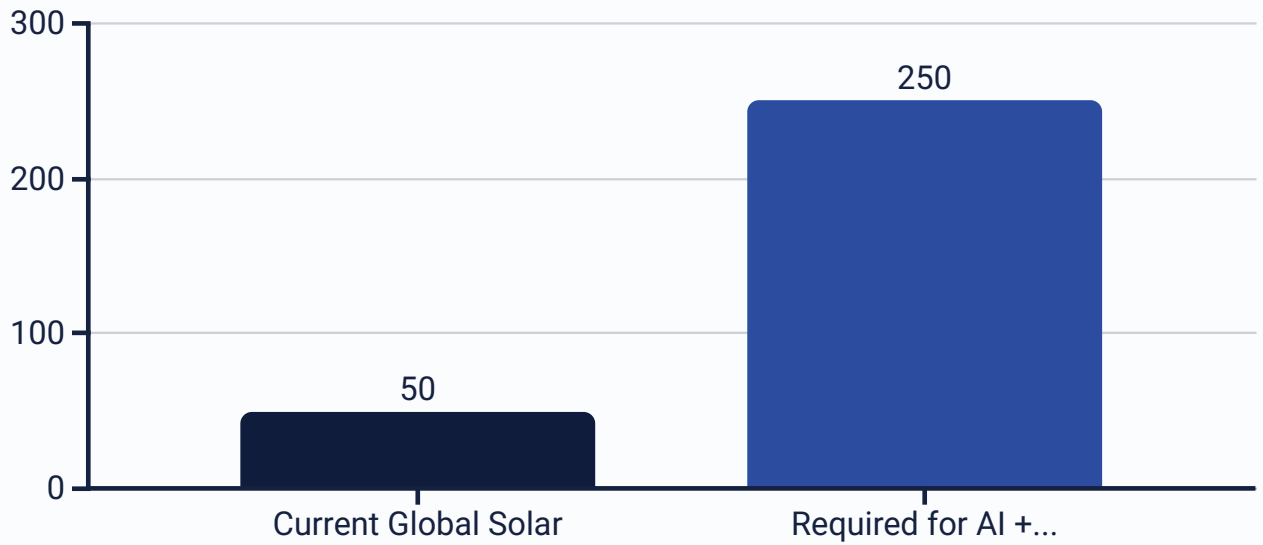
Edge computing reduces but doesn't eliminate the fundamental challenge

The Renewable Imperative

Part 4: The scale and speed of renewable deployment required

The Deployment Gap

Need 250 GW/year, Building 50 GW/year



Current Global Deployment

50 GW/year solar capacity additions worldwide

Required Acceleration

250 GW/year needed for AI plus broader electrification

5x acceleration required - the biggest infrastructure challenge in human history

Hanwha's Full Stack Response

Manufacturing to Software

Manufacturing

8.4 GW capacity with 3-month delivery capability

Development

6.3 GW in PPAs with 18-month deployment timeline

Storage

$1.4B projects solving the duck curve problem

Software

Geli AI optimization platform for maximum efficiency

The Storage Evolution

From 4-Hour to 12-Hour Storage by 2027

2024: 4-Hour Standard

Current Li-ion systems provide basic peak shifting

2026: 8-Hour Systems

LFP technology extends duration capabilities

2027: 12-Hour Storage

Iron-air batteries enable overnight coverage

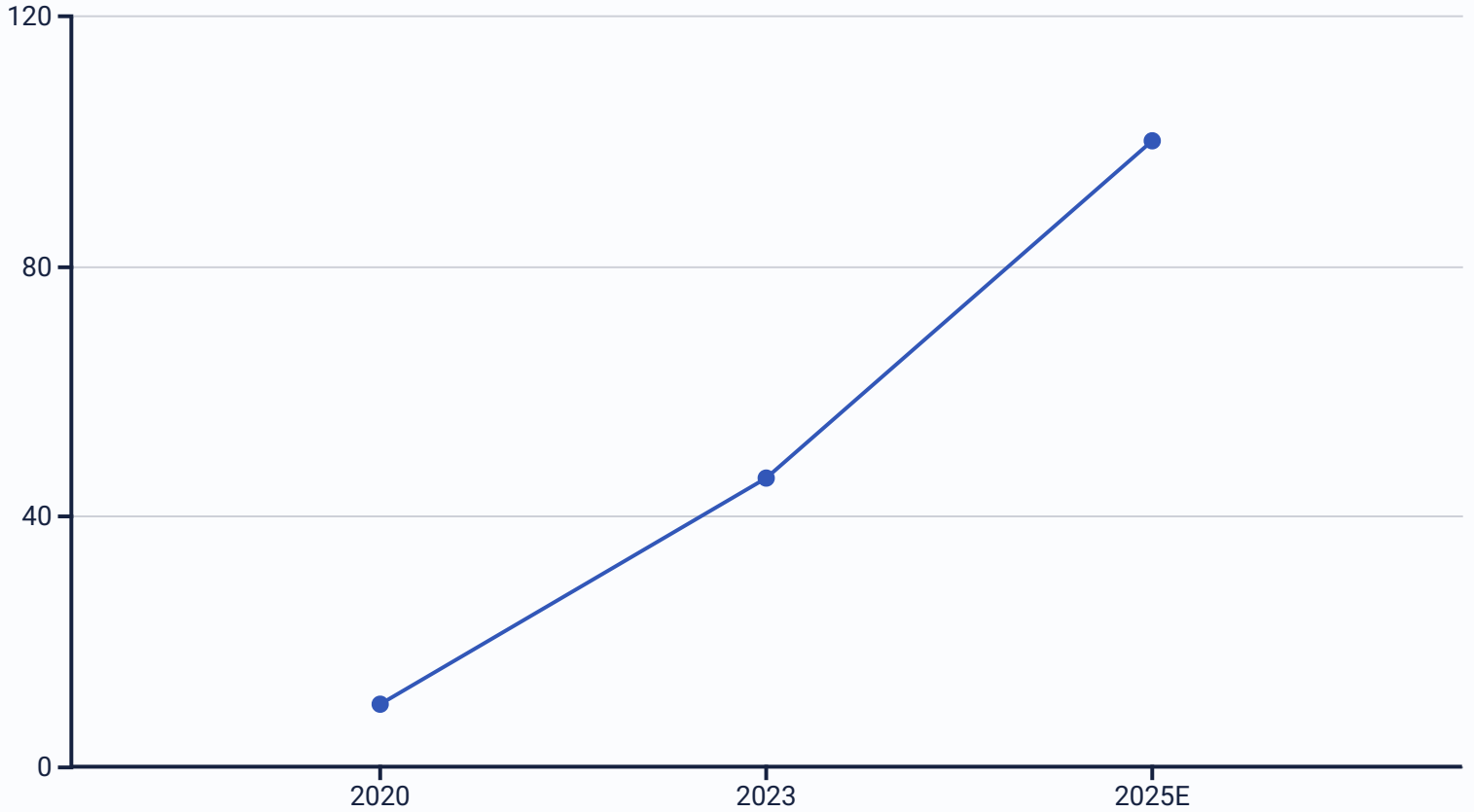
2030: Seasonal Storage?

Long-duration solutions for multi-day periods

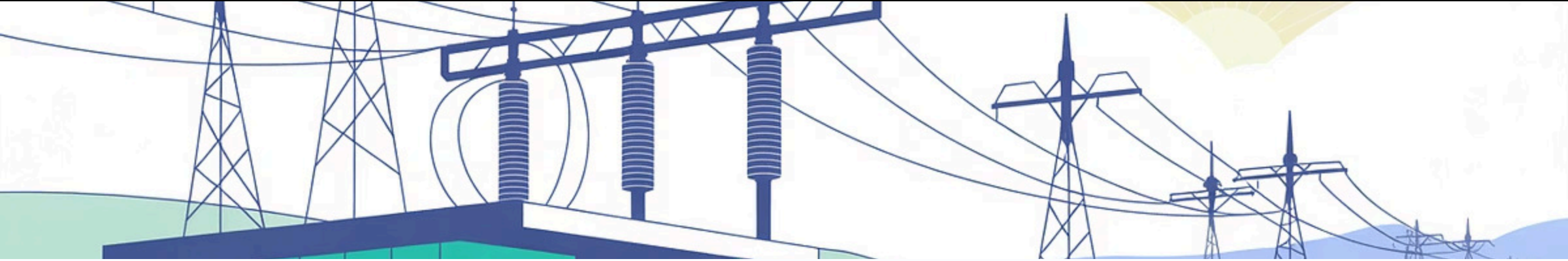
Finally solving the overnight AI inference problem?

Virtual PPAs Scale

46 GW Contracted in 2023, 100 GW Expected by 2025



Corporations bypassing slow utilities with direct renewable procurement



Grid Enhancement Reality

$180B Needed for Transmission Alone

Transmission Lines

$180B for new high-voltage connections

Substation Upgrades

$50B for capacity and smart grid integration

Smart Grid Technology

$30B for advanced control and monitoring systems

Even with massive renewable deployment, the grid itself needs fundamental upgrades to handle distributed generation and AI's concentrated loads.

The Path Forward

Part 5: Actionable strategies for different stakeholders

Three Uncomfortable Truths

What Silicon Valley Needs to Accept

Efficiency Won't Save Us

Jevons Paradox means efficiency gains increase total consumption, not reduce it

Nuclear Too Slow

10+ year development timeline doesn't match AI's 2-3 year growth cycles

Gas Is the Slow Option

7-10 year deployment makes natural gas slower than renewable alternatives

Only massive renewable deployment can scale at the required speed

For Entrepreneurs - The Real Opportunities

Build for the Efficiency-Demand Gap

1

12+ Hour Storage Solutions

Technologies that can bridge the overnight gap when solar isn't available

2

Grid Flexibility Software

AI-driven systems that optimize renewable energy distribution and consumption

3

Hybrid Renewable Controllers

Smart systems managing solar, wind, and storage as integrated units

4

Advanced Cooling Technology

Solutions for 200kW+ server racks that dramatically reduce cooling energy

For Corporate Leaders - Strategic Imperatives

Lock in Energy Now to Win the AI Race

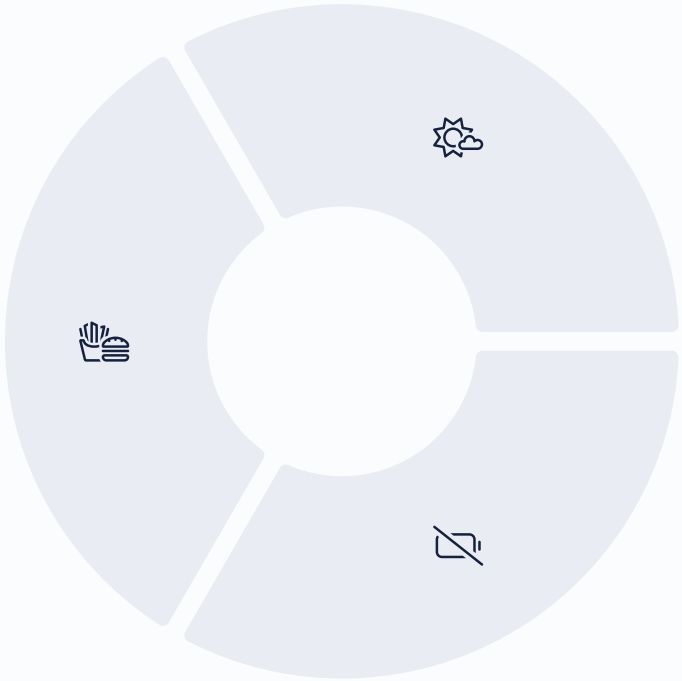
- **Secure PPAs**
  Lock in long-term renewable power purchase agreements before prices rise
- **On-site Generation**
  Evaluate distributed solar and storage for critical data centers
- **Co-location Strategy**
  Plan data center locations near renewable energy sources
- **Time-shifting**
  Implement workload scheduling to match renewable availability

For Investors - The Double Play

Bet on BOTH Efficiency and Capacity

Efficiency Tech

40% allocation to chips, cooling, and optimization software



Renewable Generation

40% allocation to solar, wind, and deployment companies

Enabling Technology

20% allocation to storage, grid tech, and integration

Single bets on either efficiency or capacity will fail. The winning strategy requires balanced exposure to both sides of the energy equation.

The Closing Reality

The Race We Must Win Together

Technology Track

Deliver 3x efficiency by 2030 ✓ (on track)

Silicon Valley is executing on chip improvements, cooling advances, and architectural innovations

Energy Track

Deploy 3x capacity by 2030 ✗ (behind)

Renewable deployment must accelerate from 50 GW to 250 GW annually

Tech is doing its part. Energy must match the pace.
The companies solving BOTH sides win the decade