Welcome slide. Set the stage: this presentation connects two seemingly unrelated worlds — digital advertising and quantitative trading — through a single powerful insight: you don't need to understand WHY something works. You just need to find the patterns that predict outcomes. That's correlation without causation.

THE CORE THESIS

> *"We don't start with models. We start with data.*
> *We don't have any preconceived notions.*
> *We look for things that can be replicated thousands of times."*
>
> — Jim Simons, Renaissance Technologies

META's Insight

You don't need to know WHY someone will buy. You just need the behavioral signals that predict they will.

Renaissance's Insight

You don't need to know WHY a stock will move. You just need the statistical signals that predict it will.

The core thesis bridges advertising and trading. Both META and Renaissance arrived at the same insight independently: causation is optional. Correlation is the signal. META predicts who will buy. Renaissance predicts what will move. Neither needs to know why.

CASE STUDY: TARGET

The Pregnancy Algorithm

What Happened

In 2010, Target's analytics team discovered they could predict pregnancy — and estimate due dates — from shopping patterns alone. No surveys. No baby registries. Just purchase data.

The Signals (none of them baby products):

- Unscented lotion purchases (2nd trimester)
- Zinc, calcium & magnesium supplements
- Extra-large bags of cotton balls
- Scent-free soap + hand sanitizer

90%

accuracy on predicting due dates

The Famous Punchline

An angry father stormed into Target demanding to know why his teenage daughter was receiving coupons for diapers and cribs.

Days later, he called back to apologize — his daughter was, in fact, pregnant. The algorithm knew before the family did.

TARGET CASE STUDY

1. Target wanted to capture new parents BEFORE competitors — before they started shopping for baby products.

2. Statistician Andrew Pole analyzed purchase data from women on the baby registry, working BACKWARD to find what they bought months before baby products.

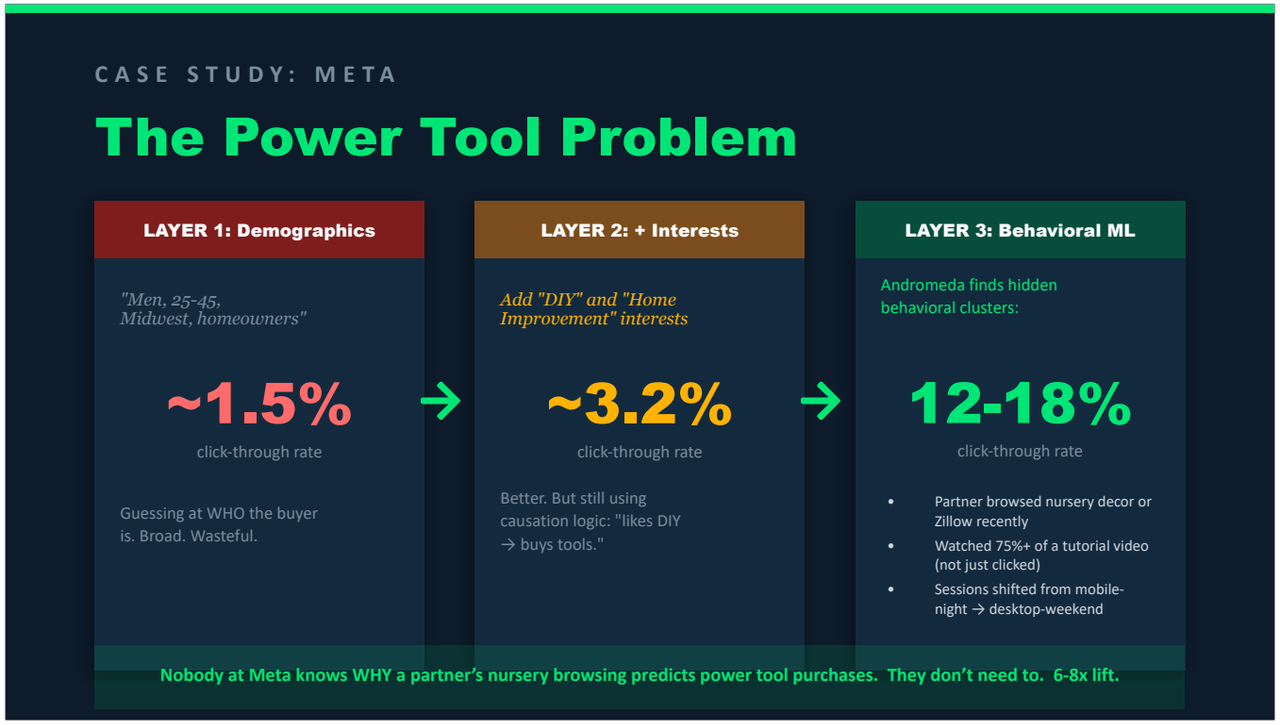
3. He identified ~25 products that predicted pregnancy. NONE were baby products: unscented lotion (smell sensitivity in 2nd trimester), zinc/calcium/magnesium supplements (first 20 weeks), extra-large cotton balls, scent-free soap.

4. The algorithm predicted due dates with ~90% accuracy, letting Target time marketing precisely.

5. Famous story: angry father in Minneapolis confronted Target about his teenage daughter getting baby coupons. Called back days later to apologize — she was due in August.

6. Target learned to MIX baby coupons with random products to hide the targeting. Revenue grew from $44B to $67B (2002-2010).

7. BRIDGE: Nobody at Target knew WHY unscented lotion predicted pregnancy. The correlation was the signal. This is exactly how ML trading works.



META CASE STUDY

1. Selling DeWalt power tools. Demographics only (men, 25-45, Midwest) = ~1.5% CTR. Wasteful.

2. Add DIY interests = ~3.2% CTR. Better but still causation logic.

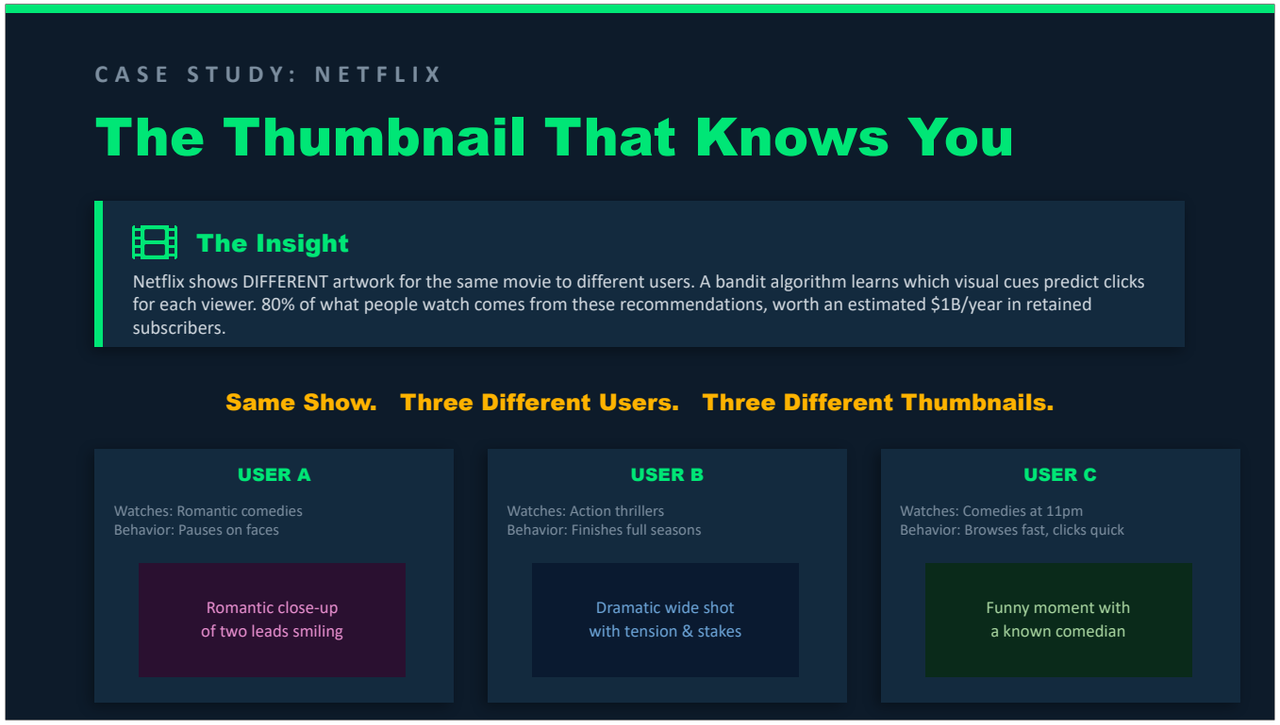
3. Meta's Andromeda AI (10,000x larger ML models) finds hidden behavioral clusters: partner browsing nursery/Zillow + watching 75%+ of tutorials + sessions shifting to desktop-weekends = 12-18% CTR. 6-8x lift.

4. Nobody knows WHY. Likely: he's moving into a new house and planning a workshop. But the algorithm just sees the pattern.

5. Meta's guidance: 'Go broad. Feed the algorithm. Let the machine figure it out.'

6. Real stats: Advantage+ = 22% higher ROAS, 13% lower cost per catalog sale, 28% lower CPC.

7. BRIDGE: Same as Renaissance — they don't know WHY weather in Kansas correlates with soybean futures. They just trade the pattern.



NETFLIX CASE STUDY

1. 17,000+ titles. Users browse 60-90 seconds before leaving. The thumbnail is the #1 click predictor.

2. For one show, Netflix generates dozens of thumbnail variants and uses contextual bandit algorithms to learn which works per user.

3. Rom-com watchers who pause on faces respond to warm close-ups. Action watchers who finish seasons respond to dramatic wide shots. Late-night comedy browsers respond to bright recognizable moments.

4. Netflix doesn't know WHY pause-on-faces correlates with warm-lit thumbnails. The correlation IS the signal.

5. 80% of watched content comes from algorithmic recs. Worth ~$1B/year in retention.

6. BRIDGE: Every thumbnail is a micro-trade. Millions per day. Tiny statistical edge that compounds. Same as Medallion's 50.75% across millions of trades.

RENAISSANCE TECHNOLOGIES

The Greatest Investing Track Record in History

Jim Simons (1938–2024)

- Mathematician & former NSA codebreaker
- PhD from Berkeley at age 23
- Founded Renaissance Technologies in 1982
- Launched the Medallion Fund in 1988
- Hired physicists, astronomers, linguists
- Zero finance backgrounds — by design
- *"Wall Street experience is frowned on"*

66%
avg annual return (gross)

90,129x
net return 1988–2022

50.75%
win rate — millions of trades

Jim Simons bio and the three headline stats. Key: he hired scientists, not traders. The 50.75% win rate is the most important number — barely better than a coin flip, but across millions of trades it generated $100B+ in profits.

HOW META DOES IT

Behavioral Prediction at Scale

Andromeda AI Engine (2025)

- ML models 10,000x larger than previous system
- Scans billions of candidate ads per second
- Maps users & ads in high-dimensional space
- Learns from behavioral sequences, not profiles
- Real-time recalibration every impression
- $4.52 return per $1 spent (avg)

Behavioral Signals Tracked

- Watch time, scroll speed, dwell time
- Product page views & add-to-carts
- Cross-platform activity patterns
- AI chat conversations (Dec 2025+)
- Device usage & session timing
- Content return & re-engagement

HOW META DOES IT: Andromeda replaced the old ad delivery system in late 2024. ML models 10,000x larger. It maps users and ads in high-dimensional mathematical space using embeddings. Learns from behavioral SEQUENCES not static profiles. As of Dec 2025, Meta AI chat conversations are also used for ad personalization. Key stat: $4.52 return per $1 spent on average.

THE META PLAYBOOK

BEFORE: Demographics

"Women, 25-34, living in suburbs, interested in fitness"

- Manual audience selection
- Static interest categories
- Guessing at causation
- Low conversion rates

AFTER: Behavioral ML

"Go broad. Feed the algorithm. Let the machine figure it out."

- Broad targeting, ML optimizes
- Real-time behavioral sequences
- Pure pattern recognition
- 22% higher ROAS vs manual

THE META PLAYBOOK: The shift from demographics (guessing at causation) to behavioral ML (pure correlation). Before: manually selecting audiences based on who you THINK will buy. After: go broad, let the algorithm find patterns in behavioral data. 22% higher ROAS. 13% lower cost per catalog sale. 28% lower CPC. This is the advertising industry's version of what Renaissance did to finance.

INSIDE RENAISSANCE

The Data Machine

> *"There's no data like more data." — Renaissance Technologies unofficial motto*

50,000	150 Gbps	40 TB/day	1960s→now
compute cores	global connectivity	new data ingested	tick data archive

Traditional Market Data

- Tick-by-tick prices from 1960s onward
- Every trade order, including uncompleted
- Commodity exchanges & futures tables
- Currency data back to the 1800s

Alternative / Unconventional Data

- Weather patterns & satellite imagery
- Shipping logs & flight departure frequency
- News archives (WSJ back decades)
- Astrological / lunar cycle data

DATA MACHINE: 50K cores, 150Gbps connectivity, 40TB/day ingested. Sandor Straus started collecting intraday data when everyone else used daily closes — 20-year head start. They collected EVERYTHING: weather, satellite, shipping, lunar cycles. Philosophy: anything that might correlate with anything is worth testing.

INSIDE RENAISSANCE

The Mathematical Arsenal

The core breakthrough: applying CODEBREAKING techniques (Hidden Markov Models) from the NSA to financial data — the same math used for speech recognition at IBM.

Hidden Markov Models

Model market as hidden states. Predict regime transitions (bull/bear) without seeing the regime directly.

Baum-Welch Algorithm

Finds unknown HMM parameters. Created by Lenny Baum, Simons' first recruit from the NSA.

Kernel Methods

Maps data into higher dimensions, revealing hidden correlations invisible in normal space.

Mean Reversion

Buy unusually low opens, sell unusually high. Thousands of small bets = high Sharpe ratio.

Brownian Motion

Model random asset behavior over time. From Einstein's 1905 paper. Used to price the noise.

Vector Embeddings

Assets mapped as vectors. Similar assets cluster. Enabled the single unified model.

MATH ARSENAL: NSA codebreaking + IBM speech recognition = financial trading. HMMs model hidden market regimes. Baum-Welch (created by Simons' NSA colleague) learns the parameters. Kernel methods find non-linear correlations in high dimensions. Mean reversion is the bread-and-butter. Brownian motion prices the noise. Vector embeddings (decades before modern AI) enabled the single unified model.

INSIDE RENAISSANCE

The Execution Engine

150K-300K	~2 days	12.5x	50.75%
trades per day	avg holding period	typical leverage (up to 20x)	win rate

The 3-Step Signal Discovery Process

1	FIND	Scan for anomalous patterns in historical pricing data. Use automated systems — not human hypotheses — to discover correlations.
2	VALIDATE	Must be statistically significant, non-random, consistent over time. Discard 99%+ of signals. Only extreme confidence deployed.
3	EXECUTE	Deploy as 'tradeable effect' into the single unified model. No human can interfere once live. Cost model self-corrects.

Known "Tradeable Effects"

Mean Reversion · Trend Following · Economic Release Patterns · Day-of-Week Seasonality · Pairs/Cointegration ("Déjà Vu")

EXECUTION: 150K-300K trades/day, ~2 day hold, 12.5x leverage. The 3-step process: FIND anomalies automatically (no human hypotheses), VALIDATE ruthlessly (discard 99%+), EXECUTE with no human interference. Secret weapon was the transaction cost model — they were best at estimating trade costs, not finding signals. Known strategies: mean reversion, trend following, economic release patterns, day-of-week seasonality, pairs trading.

INSIDE RENAISSANCE

Why Nobody Can Copy It

> Post-2008, the entire hedge fund industry tried to become Renaissance. They hired PhDs, built data centers, developed ML models. Result: ~15% returns. Good. Not Renaissance good.

Scientists, Not Traders

Physicists, astronomers, linguists, speech recognition experts. "Wall Street experience is frowned on." They sought astronomers for their understanding of low signal-to-noise problems.

One Model, Not Teams

At Citadel, teams compete for capital. At Two Sigma, strategies get spun off. At Renaissance, EVERYTHING stays in ONE unified model. Improvements anywhere help everywhere.

Open Source Internally

Everyone had access to the full source code — even admin staff. Compensation tied to the overall model, not individual strategies. Collaboration instead of competition.

Stay Small On Purpose

Capped at $10-15B for 20+ years. Returns all profits every 6 months. Kicked out ALL external investors in 2005. Optimizing for returns per dollar, not total AUM.

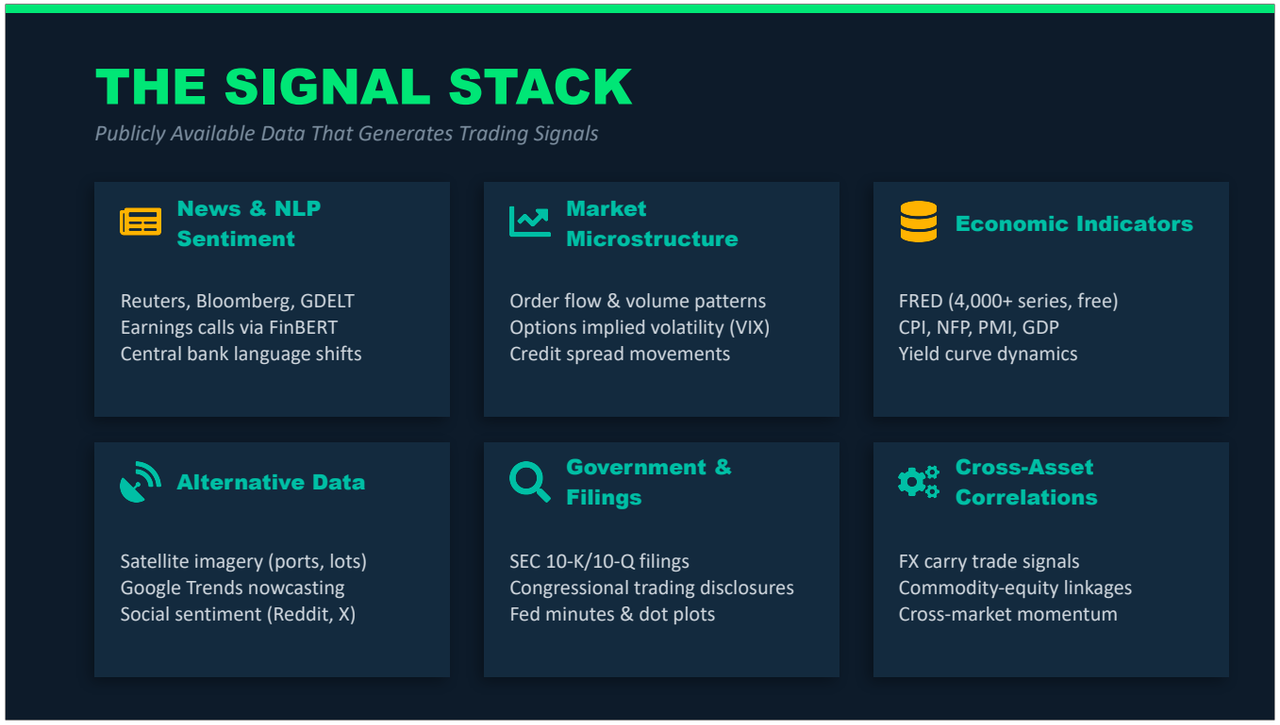
WHY NOBODY CAN COPY: Copycats thought the edge was technology. It’s structure. Scientists not traders. One unified model (not competing teams). Open source internally (radical collaboration). Staying small on purpose (returns per dollar, not AUM). The lesson for us: you don’t need their infrastructure to apply their PHILOSOPHY.

FROM ADS TO ALPHA

The Same Philosophy, Different Markets

Dimension	META Advertising	ML Trading
Input Data	Behavioral signals, clicks, scroll patterns, watch time	News sentiment, economic indicators, price action
Method	Pattern recognition in high-dimensional space	Statistical arbitrage across thousands of signals
Prediction	Who will convert (buy, sign up, engage)	What will move (price direction, volatility)
Key Insight	No need to know WHY they buy	No need to know WHY it moves
Edge	Millisecond optimization at massive scale	Tiny statistical edge compounded across millions

The bridge slide. Shows the direct parallel between META advertising and ML trading across every dimension. Same philosophy, same math, different markets.



The six categories of publicly available data we can use: NLP sentiment, market microstructure, economic indicators (FRED is free with 4,000+ series), alternative data, government filings, and cross-asset correlations.

THE ML PIPELINE

From Raw Data to Automated Buy/Sell Signals

01	INGEST	Aggregate news feeds, economic data, market prices, and alternative data into a unified data lake
02	PROCESS	NLP sentiment scoring, feature extraction, normalization. Turn unstructured data into numerical features
03	DETECT	ML models scan for statistical anomalies and correlations across thousands of feature combinations
04	VALIDATE	Backtest against 20+ years of history. Discard 99%+ of signals — deploy only statistically significant ones
05	EXECUTE	Generate automated buy/sell signals. Position sizing based on conviction score and portfolio risk

THE ML PIPELINE: Five steps from raw data to trading signals. INGEST: aggregate all data sources. PROCESS: NLP scoring, feature extraction. DETECT: scan for anomalies across thousands of combinations. VALIDATE: backtest 20+ years, discard 99%+ of signals. EXECUTE: generate automated buy/sell with risk-adjusted sizing. This mirrors Renaissance's 3-step process but adapted for our implementation.

THE OPPORTUNITY

Building Our ML Correlation Layer

Phase 1: Foundation

- FRED API integration (free, 4,000+ series)
- News sentiment via NLP (FinBERT)
- Historical backtesting framework
- Signal discovery engine

Phase 2: Intelligence

- Real-time data streaming (WebSocket)
- Multi-signal correlation detection
- Automated anomaly alerts
- Portfolio-level signal aggregation

Phase 3: Automation

- Auto buy/sell signal generation
- Risk-adjusted position sizing
- Continuous model retraining
- Alternative data integration

We independently arrived at the same insight that made Renaissance the most successful quant fund ever.

Our three-phase build plan. Phase 1: FRED + NLP + backtesting. Phase 2: real-time streaming + multi-signal detection. Phase 3: automated buy/sell signals. The closing callout ties it back to Renaissance.

CORRELATION
IS ALL YOU NEED.

The explanation comes later — if ever.

Key Source: *The Man Who Solved the Market* by Gregory Zuckerman

Close strong. 'Correlation is all you need' — echo the Simons philosophy. Reference the Zuckerman book as the anchor source. Leave them with the key insight: the explanation comes later, if ever.