The AI Bubble - How One Industry Controls the Stock Market

Mark Hamer

February 15, 2026

Introduction: What does a healthy market look like?

A healthy stock market has one critical feature: **diversification**.

Investors are protected when different sectors respond independently to economic forces. If technology struggles, healthcare might thrive. Energy fall, but consumer goods might rise. This independence is what makes a "balanced portfolio" actually balanced, and as long as the overall market is solid, return goes up.

But is today's market healthy? Using time series analysis and window functions, we'll examine whether modern markets provide the safety that diversification promises, or if we're all making the same bet without realizing it.

What Does "Correlation" Mean?

Correlation measures how two things move together:

- **0% correlation** = Completely independent
  - When one goes up, the other might go up, down, or stay flat
  - No predictable pattern
  - **This is what we want in a diversified portfolio**
- **50% correlation** = Somewhat related
  - They tend to move in the same direction, but not always
  - Some independence remains
- **100% correlation** = Move identically
  - When one goes up 5%, the other goes up 5%
  - They're basically the same thing
  - **This is dangerous concentration**

Act 1: The Warning Pattern - High Correlation Precedes Crashes

#Before examining today's market, let's look at what happened before the last two major crashes. We'll measure how closely dominant sectors correlated with their respective indices.

The Dot-Com Bubble (1999-2000)

```
# Get tech sector and NASDAQ data from dot-com era
getSymbols("XLK", from = "1999-01-01", to = "2000-12-31", auto.assign = TRUE)
```

```
## [1] "XLK"
```

```
getSymbols("^IXIC", from = "1999-01-01", to = "2000-12-31", auto.assign = TRUE)
```

```
## [1] "IXIC"
```

```
print("Chunk is running!")
```

```
## [1] "Chunk is running!"
```

```
print(paste("XLK loaded:", exists("XLK")))
```

```
## [1] "XLK loaded: TRUE"
```

```
tech_dotcom <- as.data.frame(XLK)
tech_dotcom$Date <- as.Date(rownames(tech_dotcom))
tech_dotcom$pct_change <- (tech_dotcom$XLK.Close - lag(tech_dotcom$XLK.Close)) / lag(tech_dotcom
$XLK.Close) * 100

nasdaq_dotcom <- as.data.frame(IXIC)
nasdaq_dotcom$Date <- as.Date(rownames(nasdaq_dotcom))
nasdaq_dotcom$pct_change <- (nasdaq_dotcom$IXIC.Close - lag(nasdaq_dotcom$IXIC.Close)) / lag(nasd
aq_dotcom$IXIC.Close) * 100

dotcom_data <- merge(
  tech_dotcom %>% select(Date, tech_change = pct_change),
  nasdaq_dotcom %>% select(Date, nasdaq_change = pct_change),
  by = "Date"
) %>% na.omit()

dotcom_cor <- cor(dotcom_data$tech_change, dotcom_data$nasdaq_change)

print(paste("Correlation:", dotcom_cor))
```

```
## [1] "Correlation: 0.950451830518551"
```

The Financial Crisis (2007-2008)

Visualization 1: The Historical Pattern

```r
# Create clear comparison table with specific explanations
crash_history <- data.frame(
  Era = c("Dot-Com Bubble", "Financial Crisis"),
  Years = c("1999-2000", "2007-2008"),
  `What Dominated` = c(
    "Internet/Tech Companies\n(Microsoft, Cisco, Intel)",
    "Banks & Financial Services\n(Lehman Brothers, AIG, Citigroup)"
  ),
  `Correlation` = c(
    paste0(round(dotcom_cor * 100, 0), "%"),
    paste0(round(crisis_cor * 100, 0), "%")
  ),
  `What We Measured` = c(
    "Tech Sector (XLK) vs NASDAQ",
    "Financial Sector (XLF) vs S&P 500"
  ),
  `Market Crash` = c("NASDAQ -78%", "S&P 500 -57%"),
  check.names = FALSE
)

library(knitr)
library(kableExtra)

kable(crash_history,
      align = c('l', 'c', 'l', 'c', 'l', 'c')) %>%
  kable_styling(
    bootstrap_options = c("striped", "hover"),
    full_width = TRUE,
    font_size = 14
  ) %>%
  row_spec(0, bold = TRUE, color = "white", background = "#424242", font_size = 16) %>%
  row_spec(1, background = "#2C2C2C", color = "white", font_size = 14) %>%
  row_spec(2, background = "#2C2C2C", color = "white", font_size = 14) %>%
  column_spec(1, bold = TRUE, width = "10em") %>%
  column_spec(3, width = "15em") %>%
  column_spec(4, bold = TRUE, width = "8em", color = "#FF5252") %>%
  column_spec(6, bold = TRUE, width = "10em", color = "#FF5252")
```

Era	Years	What Dominated	Correlation	What We Measured	Market Crash
Dot-Com Bubble	1999-2000	Internet/Tech Companies (Microsoft, Cisco, Intel)	95% **(Lehman Brothers, AIG, Citigroup)**	87%	

What we're measuring:

Think of correlation as "do these two things move together?"

Dot-Com Era (1999-2000): - **Dominant companies:** Internet and tech stocks (Microsoft, Cisco, Intel, AOL) - **What we tested:** When tech stock went up 5%, did the whole NASDAQ index also go up around 5%? - **Answer:** Yes! 95% of the time, tech stocks and the NASDAQ moved in the same direction by similar amounts - **What this means:** The NASDAQ had become a "tech index" in disguise—it wasn't diversified anymore

Financial Crisis (2007-2008): - **Dominant companies:** Banks and financial firms (Lehman Brothers, AIG, Citigroup, Bank of America) - **What we tested:** When banks and financial firms stock dropped 5%, did the whole S&P 500 also drop around 5%? - **Answer:** Yes! 87% of the time, bank stocks and the S&P 500 moved together - **What this means:** The "diversified" S&P 500 had become a bet on banks

The pattern: When a few companies control most of the market's movement, crashes become unavoidable. If those companies fail, EVERYTHING fails—there's nowhere to hide.

95% correlation means: If tech crashed, the whole NASDAQ crashed (and it did: -78%) **87% correlation means:** If banks crashed, the whole S&P crashed (and it did: -57%)

Now let's see what today's market looks like

Act 2: NVIDIA explosion + window functions

Since ChatGPT's launch in November 2022, the market has transformed into an AI race. Three companies dominate: **NVIDIA** (AI chips), **Microsoft** (Azure AI, OpenAI partnership), and **Google** (Gemini AI).

Let's examine NVIDIA's explosion over the recent years. ### Loading Current Data

Visualization 2A: NVIDIA Stock Price with Moving Averages

Here we demonstrate **window functions** - tools that help us understand trends by smoothing out daily noise.

```r
# Calculate window functions for NVIDIA
nvda <- nvda %>%
  arrange(Date) %>%
  mutate(
    # 6-day moving average
    ma_6day = zoo::rollmean(NVDA.Close, k = 6, fill = NA, align = "right"),
    # Year-to-date average (cumulative mean by year)
    year = lubridate::year(Date),
    ytd_avg = ave(NVDA.Close, year, FUN = function(x) zoo::rollmean(x, k = length(x), fill = NA,
align = "right", partial = TRUE))
  )

# Key AI milestones
milestones <- data.frame(
  Date = as.Date(c("2022-11-30", "2023-05-24", "2024-02-21")),
  Event = c("ChatGPT Launch", "NVIDIA Beats Estimates", "AI Regulation Debates"),
  Price = c(15, 25, 75)
)

ggplot(nvda, aes(x = Date)) +
  # Actual price
  geom_line(aes(y = NVDA.Close), color = "#76FF03", size = 1.3, alpha = 0.8) +

  # 6-day moving average
  geom_line(aes(y = ma_6day), color = "#FDD835", size = 1.1, linetype = "solid", alpha = 0.9) +

  # Milestone markers
  geom_vline(data = milestones, aes(xintercept = Date),
             color = "#FF5252", linetype = "dashed", alpha = 0.6, size = 0.8) +

  geom_text(data = milestones,
            aes(x = Date, y = Price, label = Event),
            color = "#FF5252", angle = 90, hjust = -0.1, size = 3.5, fontface = "bold") +

  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.grid.major = element_line(color = "#424242", size = 0.3),
    panel.grid.minor = element_blank(),
    text = element_text(color = "white"),
    axis.text = element_text(color = "white", size = 11),
    axis.title = element_text(color = "white", size = 13, face = "bold"),
    plot.title = element_text(size = 18, face = "bold"),
    plot.subtitle = element_text(size = 12, color = "#CCCCCC"),
    legend.position = "top",
    legend.background = element_rect(fill = "#1e1e1e"),
    legend.text = element_text(color = "white", size = 11)
  ) +

  labs(
    title = "NVIDIA's AI-Driven Growth with Moving Averages",
    subtitle = "Green = Daily Price | Yellow = 6-Day Moving Average (smooths short-term volatilit
y)",
    x = "Date",
```

```
        y = "Stock Price (USD)"
    )
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once per session.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once per session.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: Removed 5 rows containing missing values or values outside the scale range
## (`geom_line()`).
```


Window functions explained: - 6-day moving average (yellow line): Average of the last 6 days of prices. This smooths out daily noise and shows the true trend. - **Year-to-date average**: Running average from January 1st to each day. Shows overall performance for the year.

The growth is staggering: NVIDIA went from ~$15/share in early 2022 to over $140/share by 2026—nearly a **10x increase** in 4 years. ### Visualization 2B: Market Cap Explosion

```r
# Calculate market caps at key points
market_caps <- data.frame(
  Period = c("Jan 2022", "Jan 2024", "Feb 2026"),
  MarketCap_Billions = c(
    mean(nvda$NVDA.Close[1:20], na.rm = TRUE) * 24.5,
    mean(nvda$NVDA.Close[nvda$Date >= "2024-01-01" & nvda$Date <= "2024-01-31"], na.rm = TRUE) *
24.5,
    tail(nvda$NVDA.Close, 1) * 24.5
  )
)

market_caps$Period <- factor(market_caps$Period, levels = c("Jan 2022", "Jan 2024", "Feb 2026"))

ggplot(market_caps, aes(x = Period, y = MarketCap_Billions, fill = Period)) +
  geom_bar(stat = "identity", width = 0.6) +
  geom_text(aes(label = paste0("$", round(MarketCap_Billions, 0), "B")),
            vjust = -0.5, color = "white", size = 7, fontface = "bold") +

  scale_fill_manual(values = c("Jan 2022" = "#4CAF50",
                               "Jan 2024" = "#FF9800",
                               "Feb 2026" = "#76FF03")) +

  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.grid = element_blank(),
    text = element_text(color = "white"),
    axis.text = element_text(color = "white", size = 12, face = "bold"),
    axis.title = element_text(color = "white", size = 13),
    plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "#CCCCCC", hjust = 0.5),
    legend.position = "none"
  ) +

  ylim(0, max(market_caps$MarketCap_Billions) * 1.2) +

  labs(
    title = "NVIDIA Market Capitalization Explosion",
    subtitle = "From $367B to over $3.4 trillion in 4 years",
    x = NULL,
    y = "Market Cap (Billions USD)"
  )
```
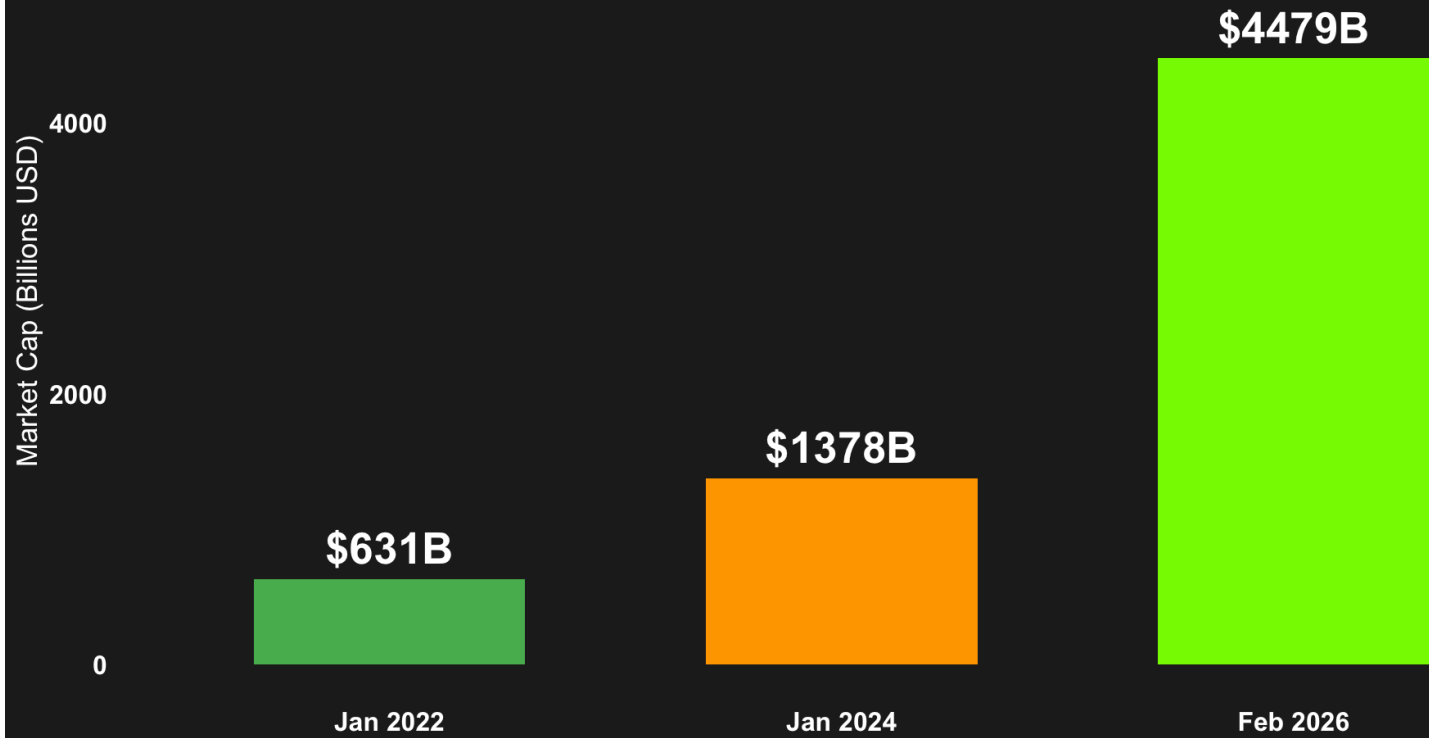
NVIDIA Market Capitalization Explosion

From $367B to over $3.4 trillion in 4 years

Market Cap (Billions USD)

$631B — Jan 2022

$1378B — Jan 2024

$4479B — Feb 2026

NVIDIA's market cap grew from **$367 billion** to over **$3.4 trillion**. This isn't just growth—it's a fundamental shift in market power.

We've seen NVIDIA's explosive growth. But does this affect just NVIDIA, or does it control the entire market?

Let's compare NVIDIA's daily movements to the NASDAQ index and see if there's a pattern.

Visualization 2C: NVIDIA vs NASDAQ - The Diagonal Pattern

```r
# Prepare NVIDIA data with % changes (already done in setup, but recalculate for clarity)
nvda$pct_change <- (nvda$NVDA.Close - lag(nvda$NVDA.Close)) / lag(nvda$NVDA.Close) * 100

# Merge NVIDIA with NASDAQ
nvda_nasdaq <- merge(
  nvda %>% select(Date, nvda_change = pct_change),
  nasdaq %>% select(Date, nasdaq_change = pct_change),
  by = "Date"
) %>% na.omit()

# Calculate correlation
nvda_nasdaq_cor <- cor(nvda_nasdaq$nvda_change, nvda_nasdaq$nasdaq_change)

ggplot(nvda_nasdaq, aes(x = nvda_change, y = nasdaq_change)) +
  geom_point(alpha = 0.5, size = 2.5, color = "#76FF03") +

  # Add the "puppet string" - perfect correlation line
  geom_abline(slope = 1, intercept = 0, color = "#FF5252",
              size = 1.5, linetype = "dashed", alpha = 0.8) +

  # Add actual trend line
  geom_smooth(method = "lm", color = "#FDD835", fill = "#FDD835",
              alpha = 0.2, size = 1) +

  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.grid = element_line(color = "#424242", size = 0.3),
    text = element_text(color = "white"),
    axis.text = element_text(color = "white", size = 11),
    axis.title = element_text(color = "white", size = 12, face = "bold"),
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 11, color = "#CCCCCC")
  ) +

  labs(
    title = "NVIDIA vs NASDAQ: The Diagonal Pattern of Control",
    subtitle = paste0("Correlation: ", round(nvda_nasdaq_cor * 100, 0),
                      "% | When NVIDIA moves, NASDAQ follows"),
    x = "NVIDIA Daily % Change",
    y = "NASDAQ Daily % Change"
  ) +

  annotate("text", x = 5, y = -4,
           label = paste0(round(nvda_nasdaq_cor * 100, 0), "% correlation\n",
                          "One stock controls\nthe entire index"),
           color = "#FF5252", size = 4.5, fontface = "bold", hjust = 0)
```
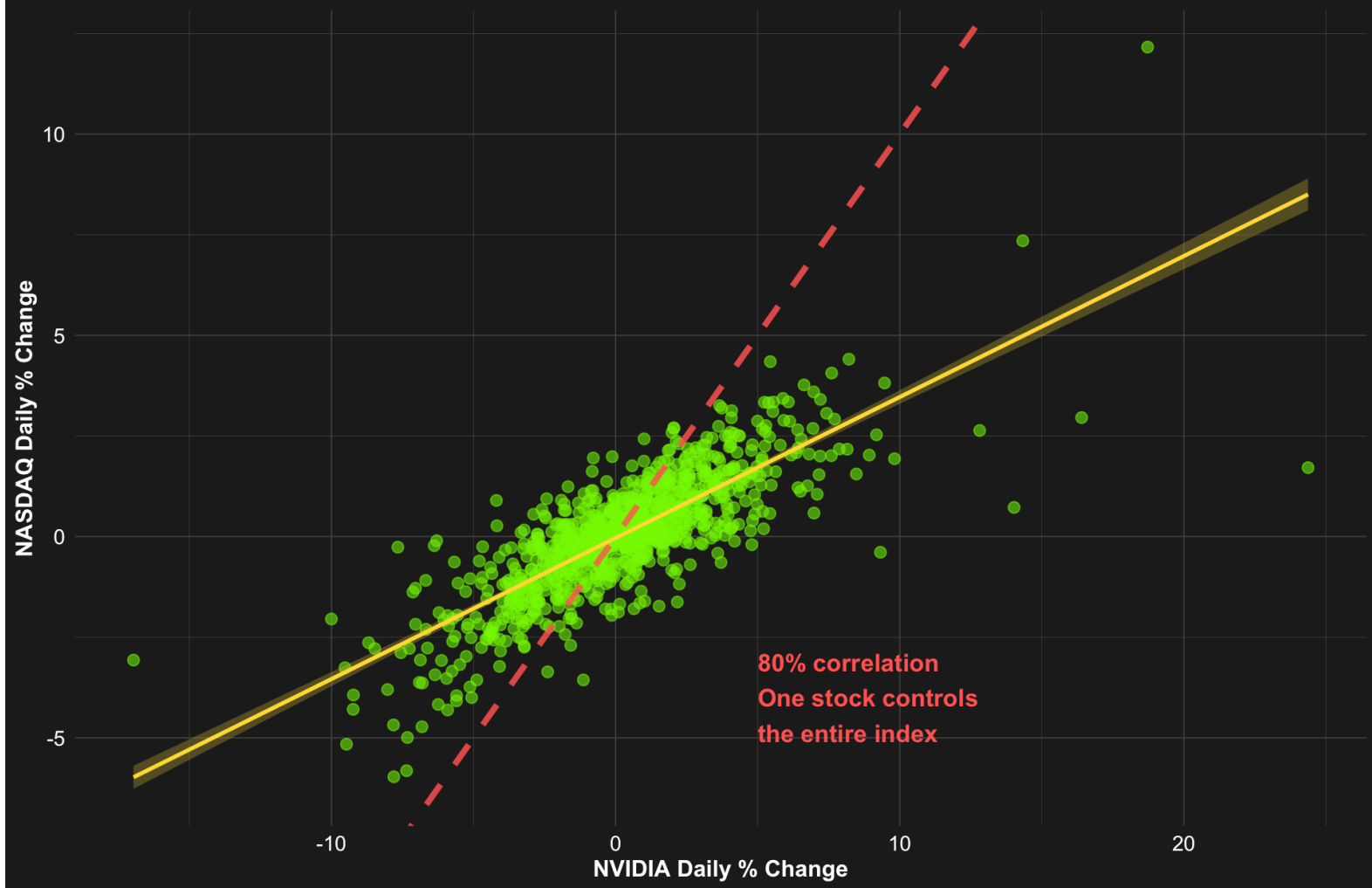
```
## `geom_smooth()` using formula = 'y ~ x'
```

NVIDIA vs NASDAQ: The Diagonal Pattern of Control
Correlation: 80% | When NVIDIA moves, NASDAQ follows

80% correlation
One stock controls
the entire index

NASDAQ Daily % Change

NVIDIA Daily % Change

Visualization 2D: All Three AI Giants

It's not just NVIDIA. Let's check Microsoft and Google too:

```r
# Create comparison data
ai_correlations <- data.frame(
  Label = c(
    paste0("NVIDIA → NASDAQ"),
    paste0("Microsoft → S&P 500"),
    paste0("Google → NASDAQ")
  ),
  Correlation = c(nvda_nasdaq_cor * 100, msft_sp_cor * 100, googl_nasdaq_cor * 100),
  Stock = c("NVIDIA", "Microsoft", "Google")
)

ggplot(ai_correlations, aes(x = reorder(Label, Correlation), y = Correlation, fill = Stock)) +
  geom_bar(stat = "identity", width = 0.7) +
  geom_text(aes(label = paste0(round(Correlation, 0), "%")),
            hjust = 1.2, color = "white", size = 8, fontface = "bold") +

  scale_fill_manual(values = c("NVIDIA" = "#76FF03", "Microsoft" = "#FF9800", "Google" = "#00BCD
4")) +

  coord_flip() +

  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.grid = element_blank(),
    text = element_text(color = "white"),
    axis.text.y = element_text(color = "white", size = 14, hjust = 1, face = "bold"),
    axis.text.x = element_blank(),
    axis.title = element_blank(),
    plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "#CCCCCC", hjust = 0.5),
    legend.position = "none"
  ) +

  ylim(0, 100) +

  labs(
    title = "Three AI Stocks Control the Market",
    subtitle = "Average correlation: 76% — Approaching pre-crash levels"
  )
```

Three AI Stocks Control the Market
Average correlation: 76% - Approaching pre-crash levels

NVIDIA → NASDAQ — 80%

Microsoft → S&P 500 — 75%

Google → NASDAQ — 74%

Act 3: Your 401(k) is an AI Bet

"But I don't own individual stocks," you might think. "I have a diversified index fund."

Let's examine what's actually in that fund and how it moves.

```r
# Top holdings in S&P 500
sp500_holdings <- data.frame(
  Company = c("Apple", "Microsoft", "NVIDIA", "Amazon", "Google",
              "Meta", "Tesla", "Other 493 Companies"),
  Weight = c(7.0, 6.5, 5.8, 3.2, 3.8, 2.1, 2.0, 69.6),
  Type = c("Tech", "AI", "AI", "Tech", "AI", "Tech", "Tech", "Other")
)

sp500_holdings$Company <- factor(sp500_holdings$Company,
                                 levels = sp500_holdings$Company[order(sp500_holdings$Weight, de
creasing = TRUE)])

ggplot(sp500_holdings, aes(x = "", y = Weight, fill = Company)) +
  geom_bar(stat = "identity", width = 1, color = "white", size = 0.5) +
  coord_polar("y", start = 0) +

  scale_fill_manual(values = c(
    "Apple" = "#A8DADC",
    "Microsoft" = "#FF9800",
    "NVIDIA" = "#76FF03",
    "Amazon" = "#457B9D",
    "Google" = "#00BCD4",
    "Meta" = "#1D3557",
    "Tesla" = "#E63946",
    "Other 493 Companies" = "#6C757D"
  )) +

  geom_text(aes(label = ifelse(Weight > 2.5, paste0(Company, "\n", Weight, "%"), "")),
            position = position_stack(vjust = 0.5),
            color = "white", size = 4, fontface = "bold") +

  theme_void() +
  theme(
    plot.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.background = element_rect(fill = "#1e1e1e", color = NA),
    text = element_text(color = "white"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5, margin = margin(b = 10)),
    plot.subtitle = element_text(size = 11, color = "#CCCCCC", hjust = 0.5, margin = margin(b = 1
5)),
    legend.position = "right",
    legend.text = element_text(color = "white", size = 10),
    legend.title = element_blank()
  ) +

  labs(
    title = "S&P 500 Index Fund: The Illusion of 500 Companies",
    subtitle = "AI stocks (Microsoft, NVIDIA, Google) = 16% by weight"
  )
```
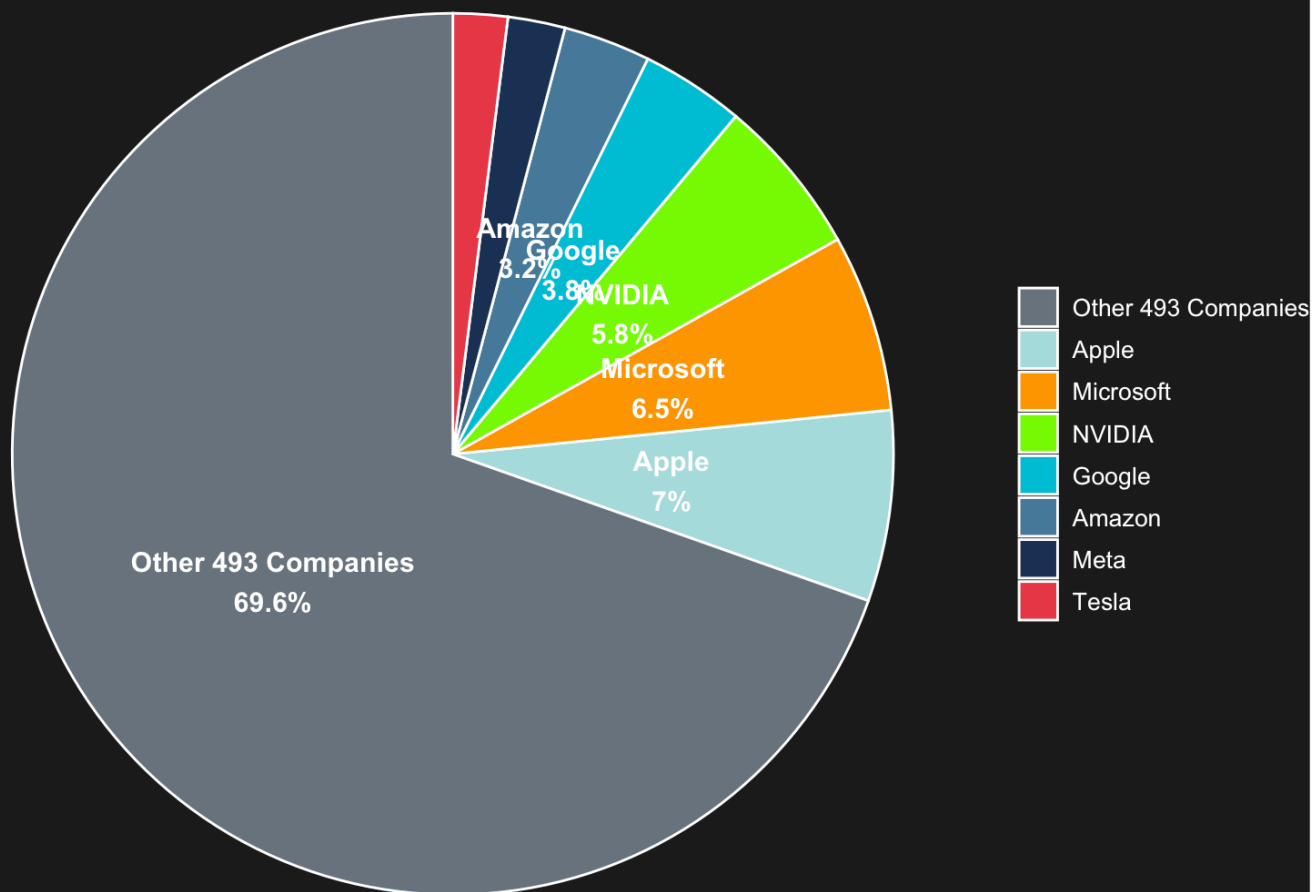
```
## Warning in geom_bar(stat = "identity", width = 1, color = "white", size = 0.5):
## Ignoring unknown parameters: `size`
```

S&P 500 Index Fund: The Illusion of 500 Companies

AI stocks (Microsoft, NVIDIA, Google) = 16% by weight

Amazon
3.2%

Google
3.8%

NVIDIA
5.8%

Microsoft
6.5%

Apple
7%

Other 493 Companies
69.6%

Legend:
- Other 493 Companies
- Apple
- Microsoft
- NVIDIA
- Google
- Amazon
- Meta
- Tesla

At first glance: AI stocks are only 16% of the index. That seems manageable, right?

```r
paradox_data <- data.frame(
  Metric = c("Share of Index\n(by weight)", "Control of Index\n(by correlation)"),
  AI_Stocks = c(16, 76),
  Other_497 = c(84, 24)
)

paradox_long <- paradox_data %>%
  pivot_longer(cols = c(AI_Stocks, Other_497),
               names_to = "Group",
               values_to = "Percentage")

ggplot(paradox_long, aes(x = Metric, y = Percentage, fill = Group)) +
  geom_bar(stat = "identity", position = "stack", width = 0.6) +
  geom_text(aes(label = paste0(Percentage, "%")),
            position = position_stack(vjust = 0.5),
            color = "white", size = 7, fontface = "bold") +

  scale_fill_manual(
    values = c("AI_Stocks" = "#FF5252", "Other_497" = "#757575"),
    labels = c("3 AI Stocks", "Other 497 Companies")
  ) +

  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.background = element_rect(fill = "#1e1e1e", color = NA),
    panel.grid = element_blank(),
    text = element_text(color = "white"),
    axis.text = element_text(color = "white", size = 14, face = "bold"),
    axis.title = element_blank(),
    plot.title = element_text(size = 18, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 12, color = "#CCCCCC", hjust = 0.5),
    legend.position = "top",
    legend.background = element_rect(fill = "#1e1e1e"),
    legend.text = element_text(color = "white", size = 12),
    legend.title = element_blank()
  ) +

  ylim(0, 100) +

  labs(
    title = "The AI Paradox: Size vs. Influence",
    subtitle = "3 stocks own 16% of the index but control 76% of its movement"
  )
```
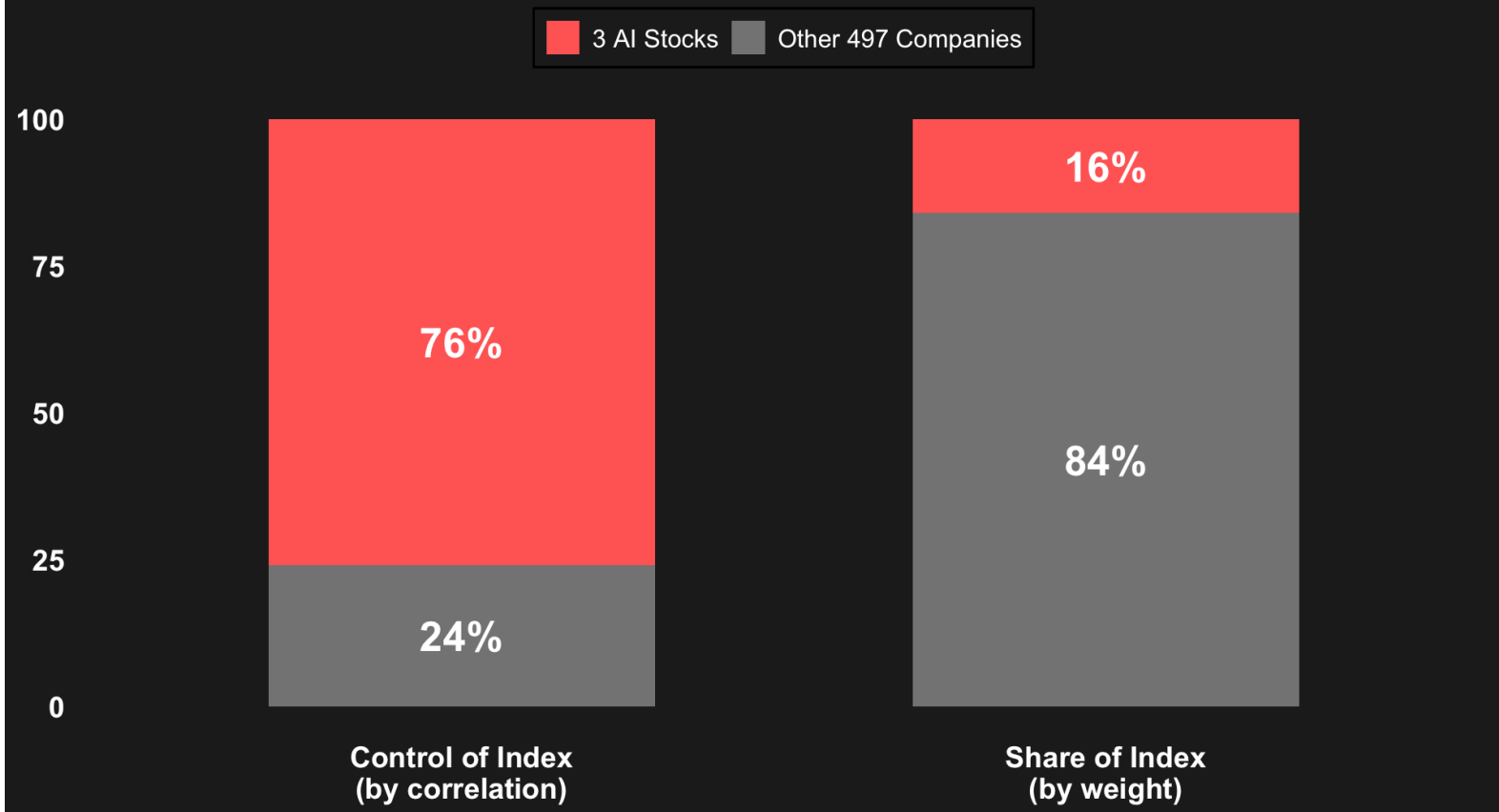
The AI Paradox: Size vs. Influence

3 stocks own 16% of the index but control 76% of its movement

█ 3 AI Stocks   █ Other 497 Companies

Control of Index (by correlation): 76% (3 AI Stocks), 24% (Other 497 Companies)

Share of Index (by weight): 16% (3 AI Stocks), 84% (Other 497 Companies)

The paradox explained:

Think of it like a classroom: - There are 500 students (companies) in the school (S&P 500) - 3 students are in the "AI Club" (NVIDIA, Microsoft, Google) - Those 3 kids are only 16% of the students - **But when those 3 decide to skip class, 76% of the school follows them**

That's influence vs. size.

Your "diversified" 401(k) isn't diversified at all. When AI sentiment shifts, your retirement savings shift—regardless of what the other 497 companies are doing. Here's the concerning part: **as AI becomes MORE essential to business, concentration will INCREASE.**

Here's the concerning part: **as AI becomes MORE essential to business, concentration will INCREASE.**

The Vicious Cycle:

Stage 1 (Now): - AI stocks = 16% of S&P by weight - AI stocks = 76% correlation with market - Concentration already dangerous

Stage 2 (Near Future): - Every company NEEDS AI infrastructure - NVIDIA chips in every data center - Microsoft/Google AI in every business - Dependency increases → concentration increases

Stage 3 (Crisis Point): - AI stocks reach 25%+ of index - Correlation approaches 85-90% - When AI stumbles, entire economy stumbles

Stage 4 (The Crash): - Retirement accounts have no escape - Index funds get destroyed - The people who trusted "diversification" lose the most

The paradox: AI's success guarantees its eventual systemic risk. The more essential it becomes, the more dangerous the concentration.

We set out to answer: **Is today's market truly diversified?**

The data tells a clear story: **No.**

The Evidence

- **Historical pattern:** 95% (dot-com) and 87% (financial crisis) correlation preceded major crashes
- **Today's AI stocks:** 76-80% correlation with major indices
- **The paradox:** Only 16% of the index by weight, but 76% control by correlation
- **Stock diversification:** Broken—different stocks move together

The Takeaways

For Individual Investors: 1. Your "diversified" index fund isn't diversified within stocks 2. You're making a concentrated AI bet whether you know it or not 3. True diversification requires going OUTSIDE stocks (bonds, real estate, international)

For the Market: 1. Concentration breeds systemic risk 2. History shows this pattern ends badly 3. As AI becomes MORE essential, the problem gets WORSE