O'REILLY
Artificial Intelligence

A Practical Guide to Conducting an AI Snake Oil Sniff Test

Josh Joseph
jj@alphafeatures.com

Chief Science Officer, Alpha Features

Lots of conversations about AI

Analytics? Isn't it just averaging?

Most of the "crowds" I know are idiots

How is this Artificial Intelligence?

Are our internal results not robust?

Rigorous and robust results

Expert insights

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

How are there insights in sentiment?

I don't see how we can mathematically quantify health

Sentiment over social media

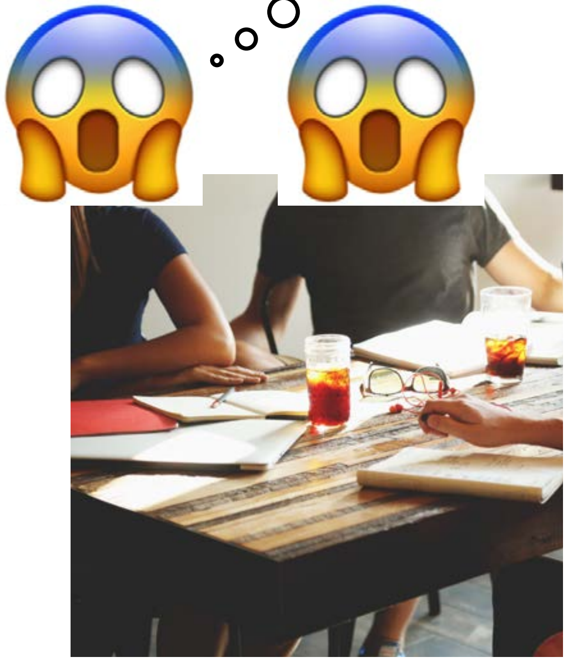
Cutting edge analytics

Mathematical quantification of company health

Lots of conversations about AI

Analytics? Isn’t it just averaging?

Most of the “crowds” I know are idiots



How is this Artificial Intelligence?

Are our internal results not robust?

How are there insights in sentiment?

I don’t see how we can mathematically quantify health

Rigorous and robust results

Expert insights

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Sentiment over social media

Cutting edge analytics

Mathematical quantification of company health

Lots of…vague conversations about AI

Smart team!

Awesome they have machine learning in the cloud!

What a great school they are from!

Joe was just talking about how random forests are super cool

And they're doing that! Of experts! In the cloud!

<Insert super important business decision being made>

Rigorous and robust results

Expert insights

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Sentiment over social media

Cutting edge analytics

Mathematical quantification of company health

Lots of…vague conversations about AI

Awesome they have
machine learning in
the cloud!

Smart team!

What
th

Joe was just talking
about how random
forests are super cool

An
Of

<Insert super important
business decision being made>

**Rigorous and
robust results**

**Expert
insights**

**e Combination of the
and Expert Knowledge**

**Sentiment
over social
media**

**Cutting edge
analytics**

**Mathematical quantification
of company health**

Lots of…vague conversations about AI

Smart team!

Awesome they have machine learning in the cloud!

What th

Joe was just talking about how random fores

Of

<Insert super important business decision being made>

Rigorous and robust results

Expert insights

Sentiment over social media

Cutting edge analytics

Mathematical quantification of company health

This talk is about how to prevent this from happening

Lots of…vague conversations about AI

Awesome they have machine learning in the cloud!

Smart team!

What a great school they are from!

Joe was just talking about how random forests are super cool

And they're doing that! Of experts! In the cloud!

<Insert super important business decision being made>

Rigorous and robust results

Expert insights

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Sentiment over social media

Cutting edge analytics

Mathematical quantification of company health

Maybe that's just the way it is? We know there's a lot of hype…

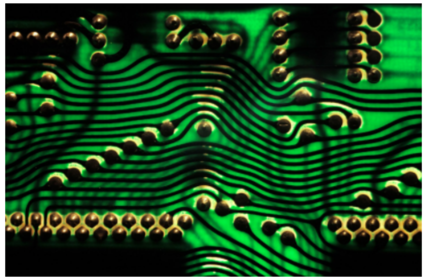
THE NEW YORKER

CURRENCY

THE HYPE—AND HOPE—OF ARTIFICIAL INTELLIGENCE

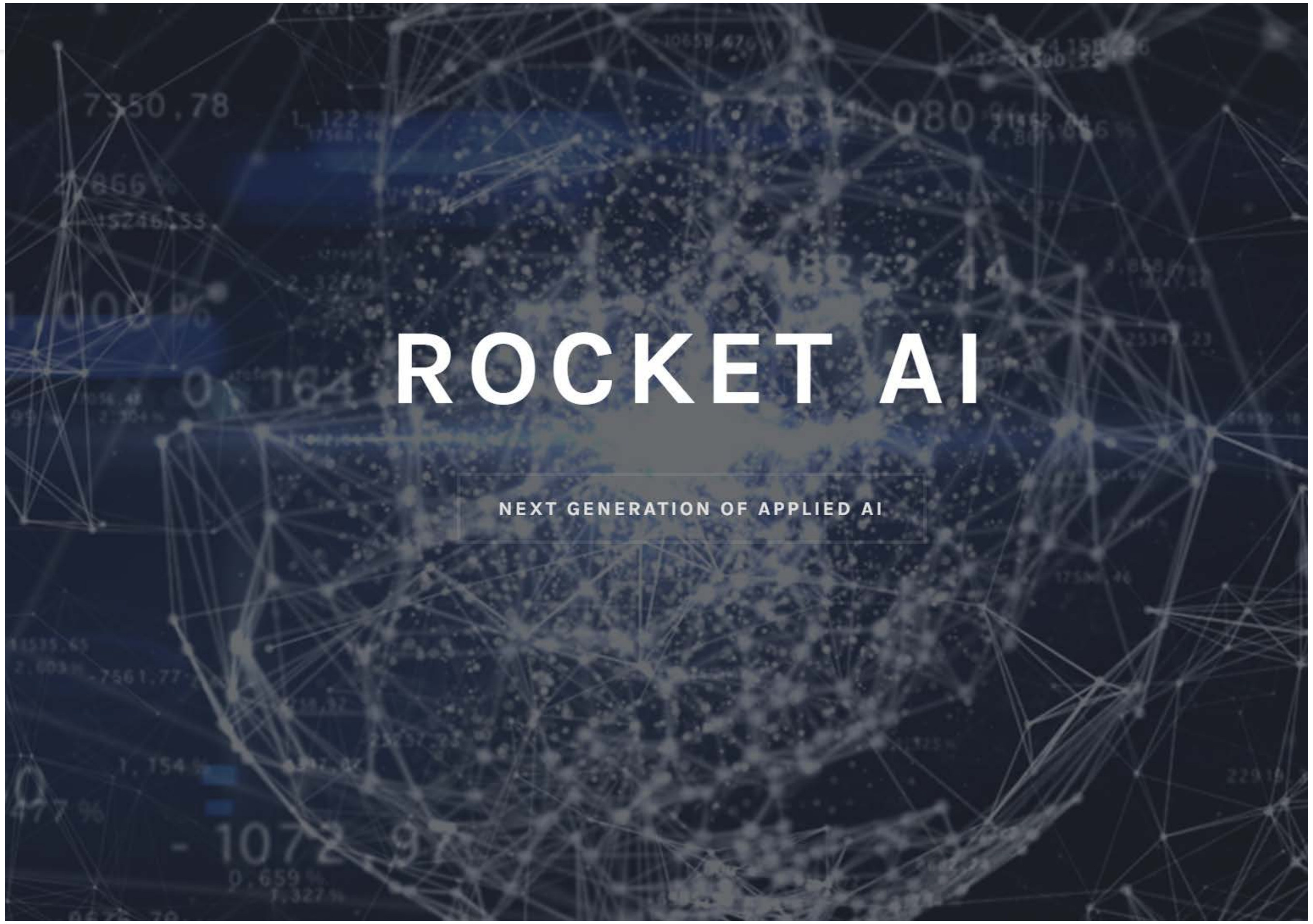
By Om Malik August 26, 2016

Earlier this month, on his HBO show "Last Week Tonight," John Oliver skewered media companies' desperate search for clicks. Like many of his bits, it became a viral phenomenon, clocking in at nearly six million views on YouTube. At around the ten-minute mark, Oliver took his verbal bat to the knees of Tronc, the new name for Tribune Publishing



Much like "the cloud," "big data," and "machine learning" before it, the term "artificial intelligence" has been hijacked by marketers and advertising copywriters.

Photograph by Erich Hartmann / Magnum



Maybe that's just the way it is? We know there's a lot of hype…

Gmail

What can AI do for your business?
1 message

Comcast Business <reply@notice.comcastbusiness.com> Wed, Aug 30, 2017 at 12:36 PM
Reply-To: Comcast Business <CB_Replies@comcastbusiness.com>
To:

Inc.

COMCAST BUSINESS B4B BUILT FOR BUSINESS

PRODUCTIVITY@WORK

AUGUST 2017

NAVIGATING THE SMB TECH REVOLUTION

Artificial intelligence (AI) has been the darling of science fiction writers for decades, but it's fiction no more—and is now becoming accessible to small and

While it's easy to make fun of the hype…there are repercussions

JEFFERIES FRANCHISE NOTE

Forward looking research offering fresh insights

Target Change

USA | Technology | IT Hardware

July 12, 2017

IBM (IBM)

Creating Shareholder Value with AI? Not so Elementary, My Dear Watson

Key Takeaway

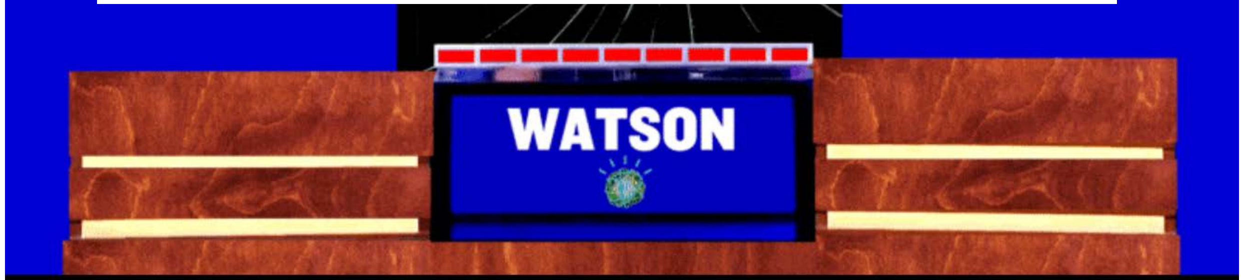
Our checks suggest that while IBM offers one of the more mature cognitive computing platforms today, the hefty services component of many AI deployments will be a hindrance to adoption. We also believe IBM appears outgunned in the war for AI talent and will likely see increasing competition. Finally, our analysis suggests that the returns on IBM's investments aren't likely to be above the cost of capital. Reiterate Underperform.

AI is the New Electricity…. Our checks confirm that a wide range of organizations are exploring incorporating AI in their business, mostly using Machine and Deep Learning for speech and image recognition applications.

…But Competitive Environment Doesn't Favor IBM. Our checks suggest that IBM's Watson platform remains one of the most complete cognitive platforms available in the marketplace today. However, many new engagements require significant consulting work to gather and curate data, making some organizations balk at engaging with IBM. As outlined

GIZMODO

"Their marketing and PR has run amok—to everyone's detriment."



Why Everyone Is Hating on IBM Watson—Including the People Who Helped Make It

Jennings Brown
8/10/17 8:45am • Filed to: WATSON

175.9K 99 6

So what do we do?

FORTUNE SUBSCRIBE

ARTIFICIAL INTELLIGENCE

Beware the Hype of Artificial Intelligence

Jonathan Vanian
Jun 23, 2017

Artificial intelligence has made great strides in the past few years, but it's also generated much hype over its current capabilities.

That's one takeaway from a Friday panel in San Francisco involving leading AI experts hosted by the Association for Computing Machinery for its 50th annual Turing Award for advancements in computer science.

Michael Jordan, a machine learning expert and computer science professor at University of California, Berkeley, said there is "way too much hype" regarding the capabilities of so-called chat bots. Many of these software

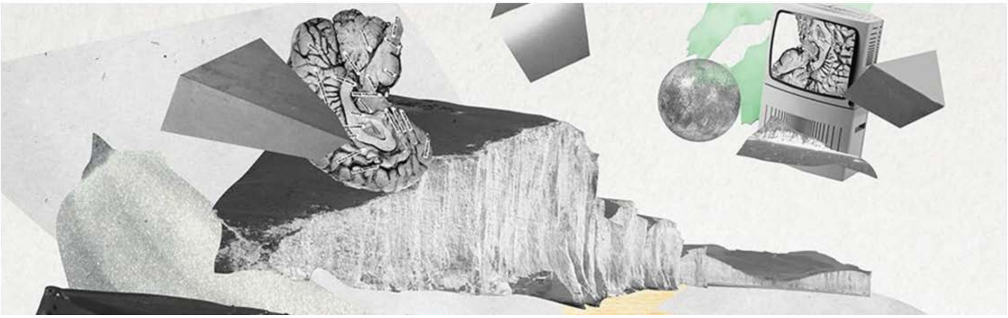
WIRED ARTIFICIAL INTELLIGENCE

Artificial Intelligence

Don't believe the hype when it comes to AI

Artificial intelligence may be subject to more hype than any other field. While this creates funding opportunities, it could also damage AI's long-term potential

By **LUKE DORMEHL**
18 Feb 2017



Frank Chen Follow

Partner at Andreessen Horowitz. Writes about tech, startups, venture investing, science, the future. Likes ex...

Jun 26 · 3 min read

In a few years, no investors are going to be looking for AI startups

But the reason I believe that no investor will be funding startups calling themselves AI-powered startups (and no startup CEO will differentiate themselves as an AI-first company like Google) is because investors will *assume* the startup is using the best available AI techniques to solve the problem they are solving.

Lots of…vague conversations about AI

Smart team!

Awesome they have machine learning in the cloud!

What a great school they are from!

Joe was just talking about how random forests are super cool

And they're doing that! Of experts! In the cloud!

<Insert super important business decision being made>

Rigorous and robust results

Expert insights

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

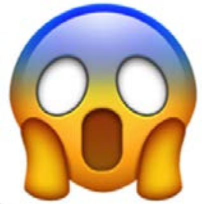
Sentiment over social media

Cutting edge analytics

Mathematical quantification of company health

Many, many experiences working through this

- Wide range of ML methods for my PhD at MIT
- Variety of AI/ML consulting work: biotech, co-working space pricing, robotics
- Alpha Features and previous propriety trading experience evaluating dozens “analytics/signals”
- Due diligence on over a hundred AI/ML hedge funds

…so you can imagine there’s been a ***lot*** of 

and I’d like to share our process for working through it

How does AI and this talk even apply to my industry?

Is this about catching liars using AI?

What's a sniff test?

A Practical Guide to Conducting an AI Snake Oil Sniff Test

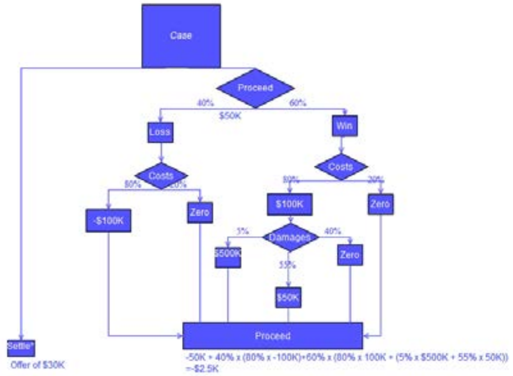
I've been at 10 AI conferences and I still don't get what AI is

Something about "it's no longer AI if we understand it"?

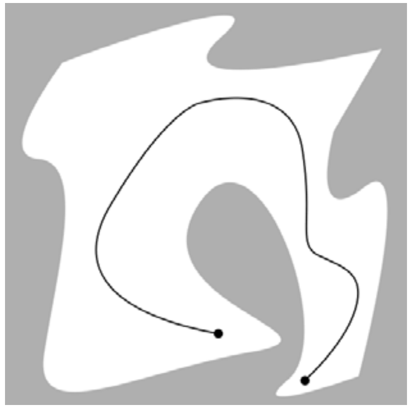
Fingers crossed he's done with the emoji jokes

A Practical Guide to Conducting an AI Snake Oil Sniff Test

A hodgepodge of rules discovered somehow
(that can be implemented in a computer)



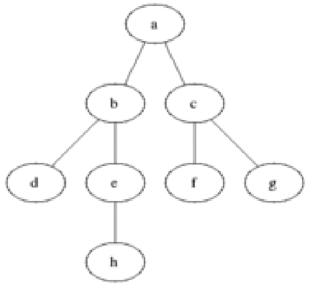
```
If (boolean condition) Then
    (consequent)
Else
    (alternative)
End If
```

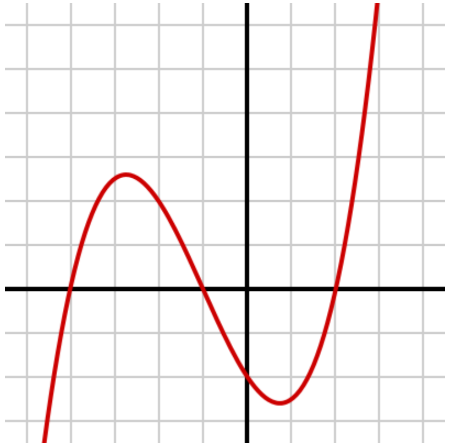


A Practical Guide to Conducting an AI Snake Oil Sniff Test

An amalgamation of rules discovered somehow
(that can be implemented in a computer)

$$\forall x \forall y (P(f(x)) \rightarrow \neg(P(x) \rightarrow Q(f(y), x, z)))$$





*pictures from wikipedia

A Practical Guide to Conducting an $\boxed{\text{AI}}$ Snake Oil Sniff Test

An amalgamation of rules discovered somehow
(that can be implemented in a computer)

I'm going to focus on machine learning for most of the talk

An amalgamation of rules discovered by **making assumptions and following principles** which allow us to believe something about the future performance (that can be implemented in a computer)

The **liar** cares about the truth and attempts to hide it; the **bullshitter** doesn't care if what they say is true or false, but rather only cares whether or not their listener is persuaded.
-Harry Frankfurt

A Practical Guide to Conducting an AI Snake Oil Sniff Test

An amalgamation of rules discovered somehow (that can be implemented in a computer)

I'm going to focus on machine learning for most of the talk

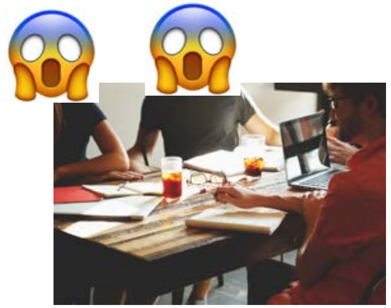
An amalgamation of rules discovered by **making assumptions and following principles** which allow us to believe something about the future performance (that can be implemented in a computer)

The **liar** cares about the truth and attempts to hide it; the **bullshitter** doesn't care if what they say is true or false, but rather only cares whether or not their listener is persuaded.
-Harry Frankfurt

A Practical Guide to Conducting an AI Snake Oil Sniff Test

An amalgamation of rules discovered somehow (that can be implemented in a computer)

30 minute to 2 hour meeting



I'm going to focus on machine learning for most of the talk

An amalgamation of rules discovered by **making assumptions and following principles** which allow us to believe something about the future performance (that can be implemented in a computer)

concrete process, questions, tips

The **liar** cares about the truth and attempts to hide it; the **bullshitter** doesn't care if what they say is true or false, but rather only cares whether or not their listener is persuaded.
-Harry Frankfurt

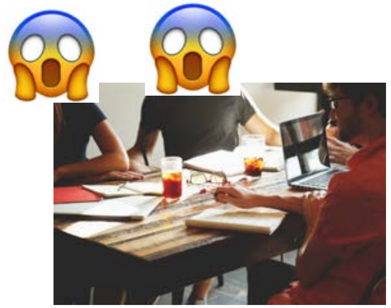
A Practical Guide to Conducting an AI Snake Oil Sniff Test

An amalgamation of rules discovered somehow (that can be implemented in a computer)

30 minute to 2 hour meeting

I'm going to focus on machine learning for most of the talk

An amalgamation of rules discovered by **making assumptions and following principles** which allow us to believe something about the future performance (that can be implemented in a computer)



Components of a Productive ML Sniff Test

- Surface-level understanding of some core ML concepts
- Sniff test procedure
- General tips

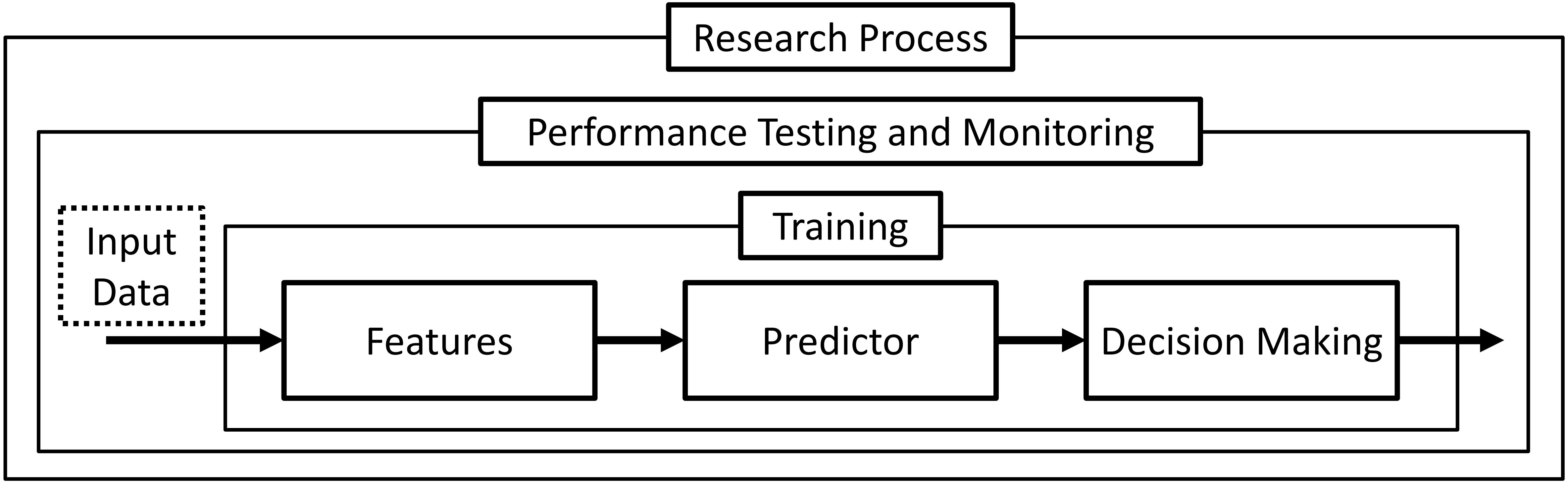
A Quick Brush of Core ML Concepts

- Generalization
  - How will my system perform when I turn it on?
  - To the future, changing environment, new users, additional markets, different geographies, etc.
- The amalgamation of rules inside these systems make assumptions
  - Not knowing the assumptions you're making is gambling
- No such thing as zero human involvement

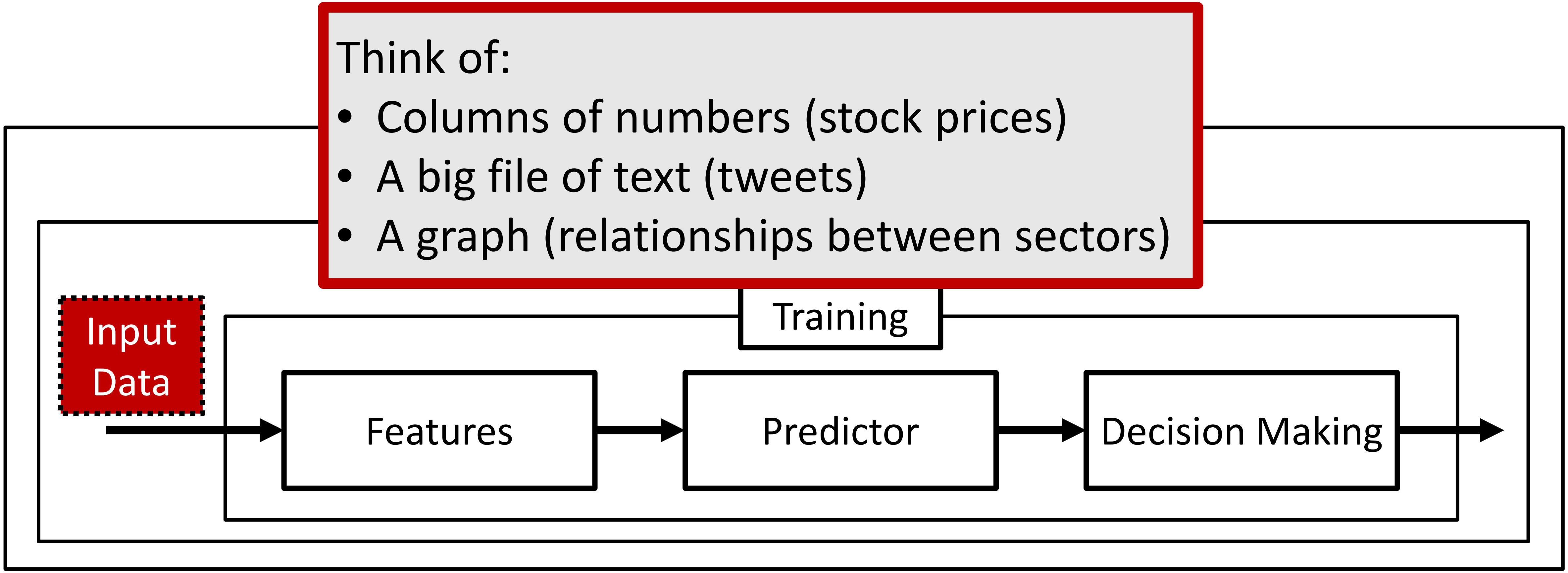
Components of a Productive ML Sniff Test

- Surface-level understanding of some core ML concepts
- **Sniff test procedure**
  - Construct your mental picture of their overall approach
  - Dig deeper and refine the edges
- General tips

High-Level ML System Initial Mental Picture



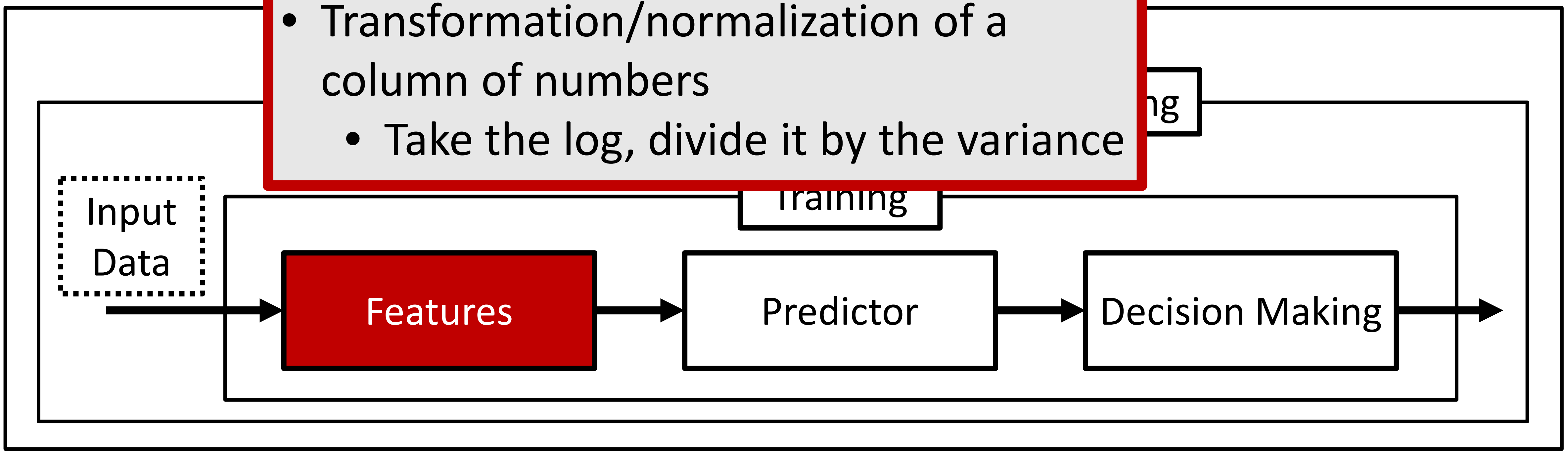
High-Level ML System Initial Mental Picture



High-Level ML System Initial Mental Picture

Think of:

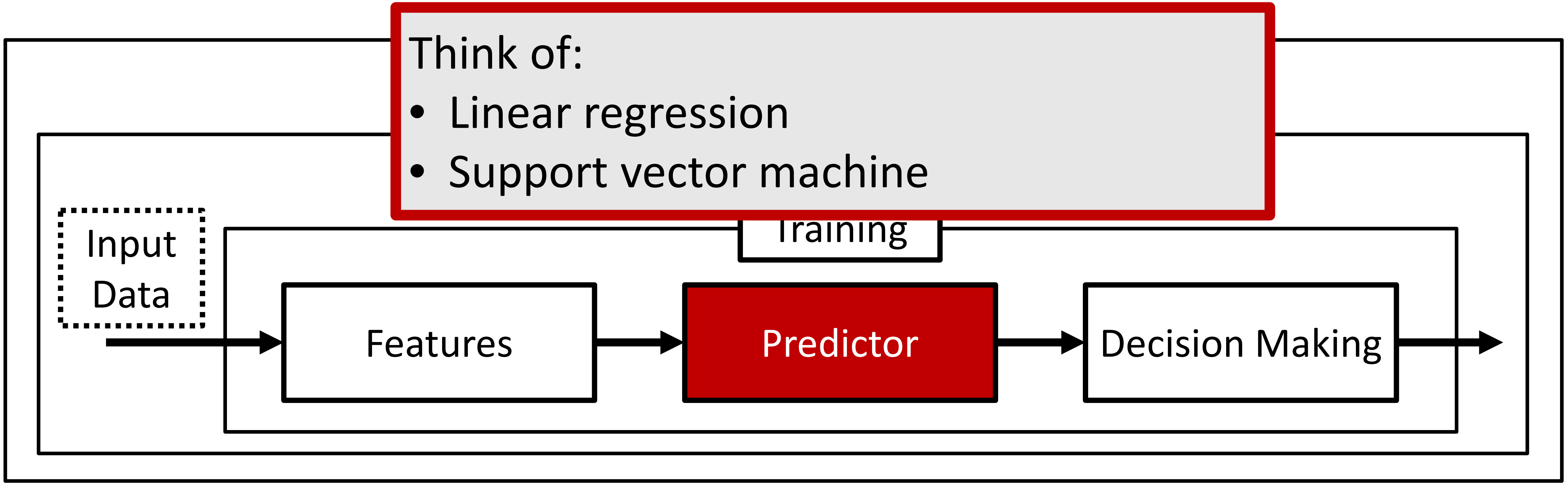
- Transformation/normalization of a column of numbers
  - Take the log, divide it by the variance



High-Level ML System Initial Mental Picture

Think of:

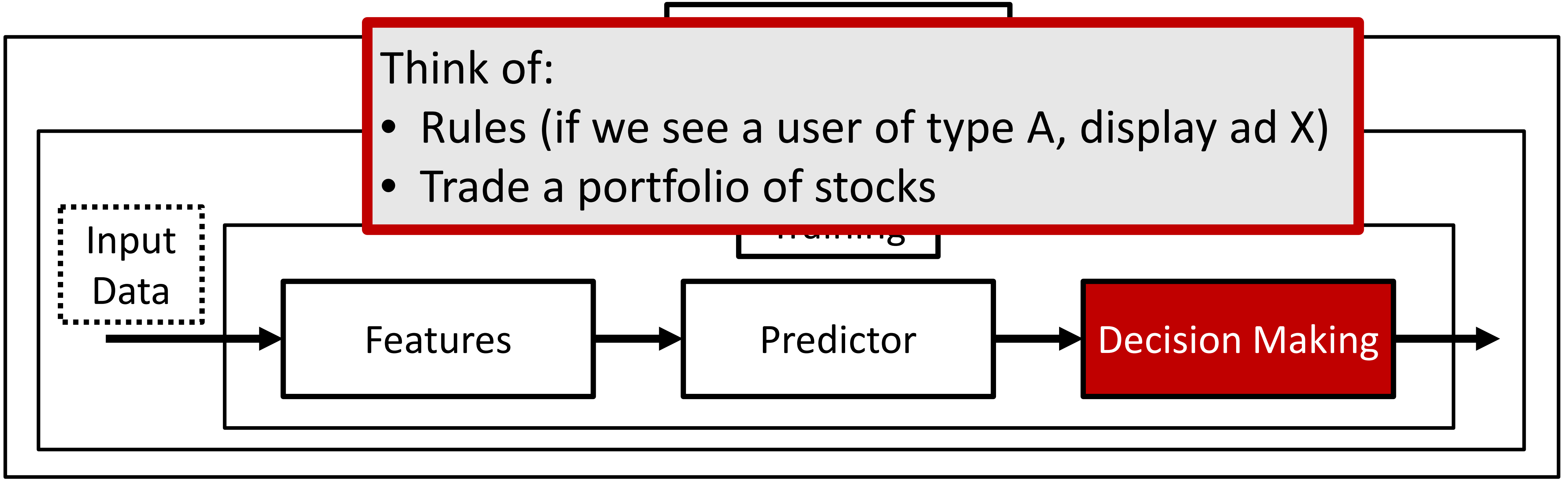
- Linear regression
- Support vector machine



High-Level ML System Initial Mental Picture

Think of:

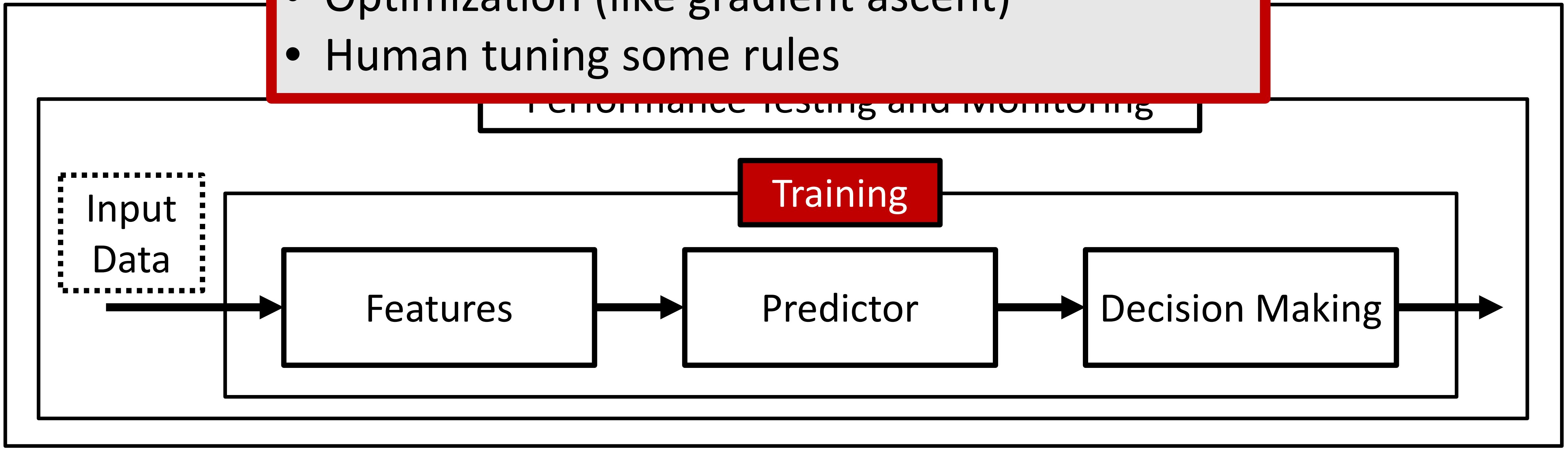
- Rules (if we see a user of type A, display ad X)
- Trade a portfolio of stocks



High-Level ML System Initial Mental Picture

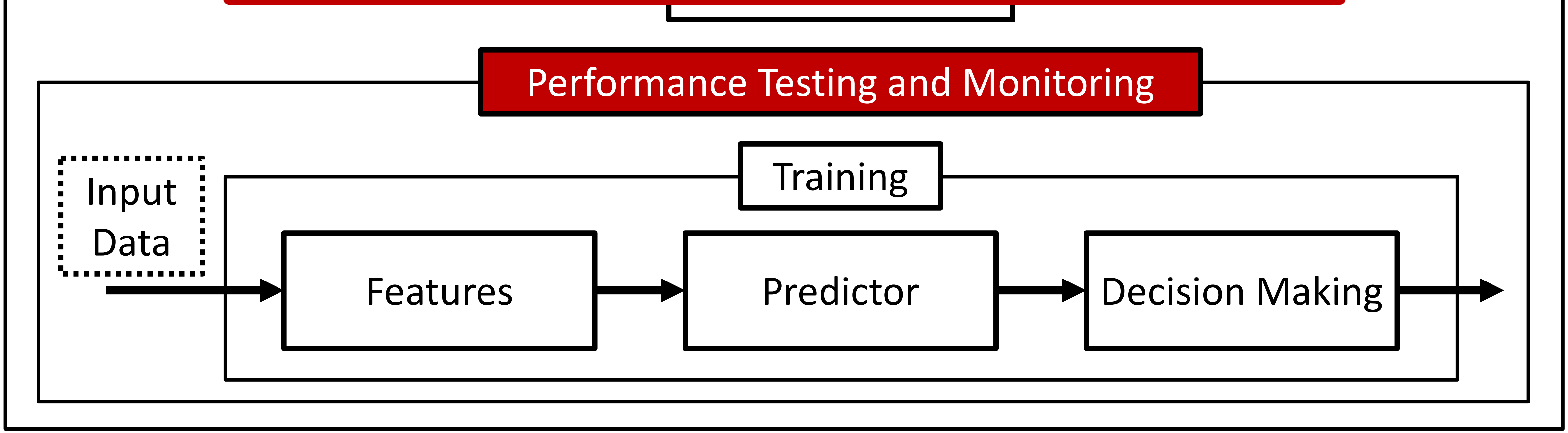
Think of:

- Optimization (like gradient ascent)
- Human tuning some rules



High-Level ML System Initial Mental Picture

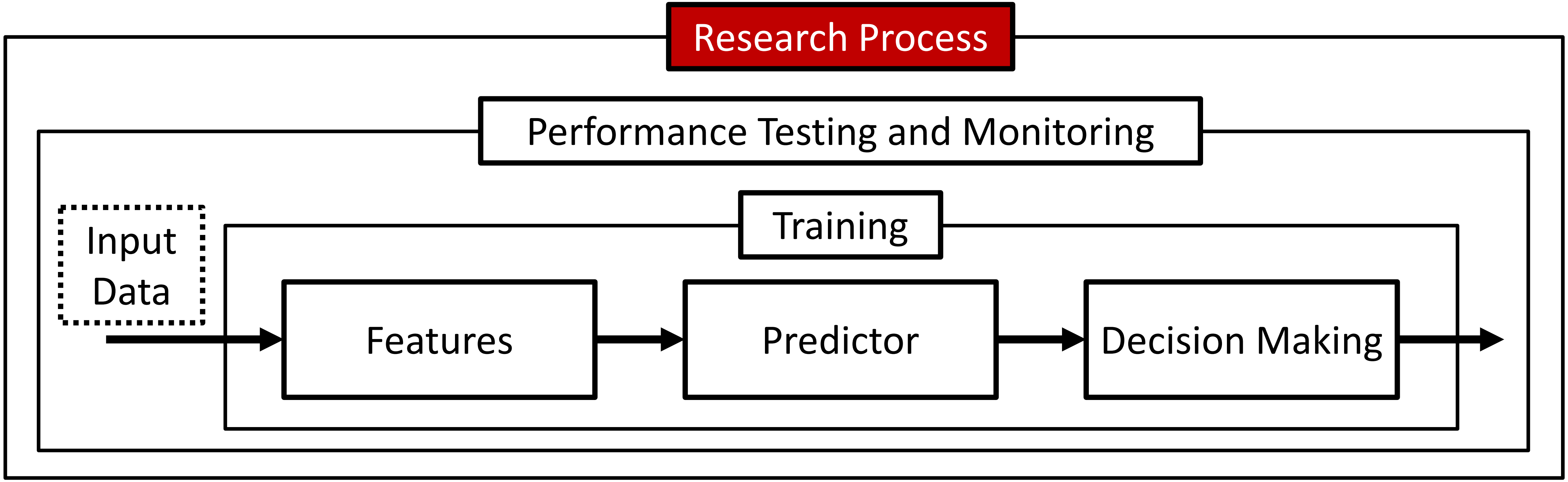
Think of:
- Computing out-of-sample performance
- Comparing live performance to a historical estimate
- Statistical testing



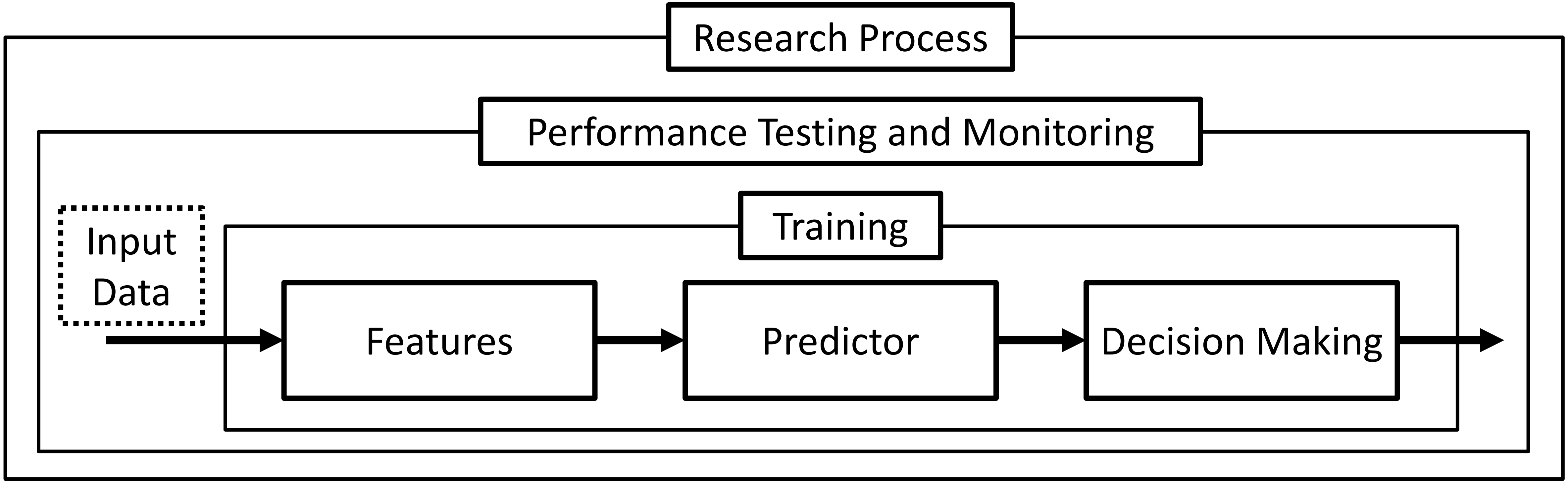
High-Level ML System Initial Mental Picture

Think of:

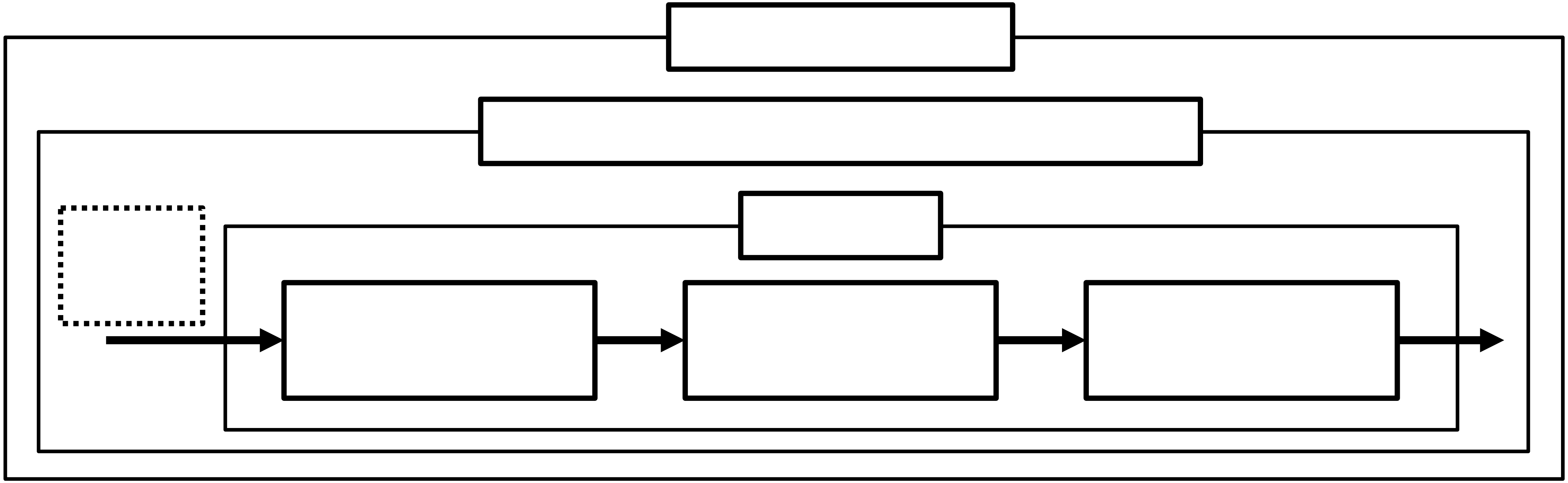
- How does the system get improved?
- What happens when something stop working well?



High-Level ML System Initial Mental Picture



High-Level ML System Initial Mental Picture



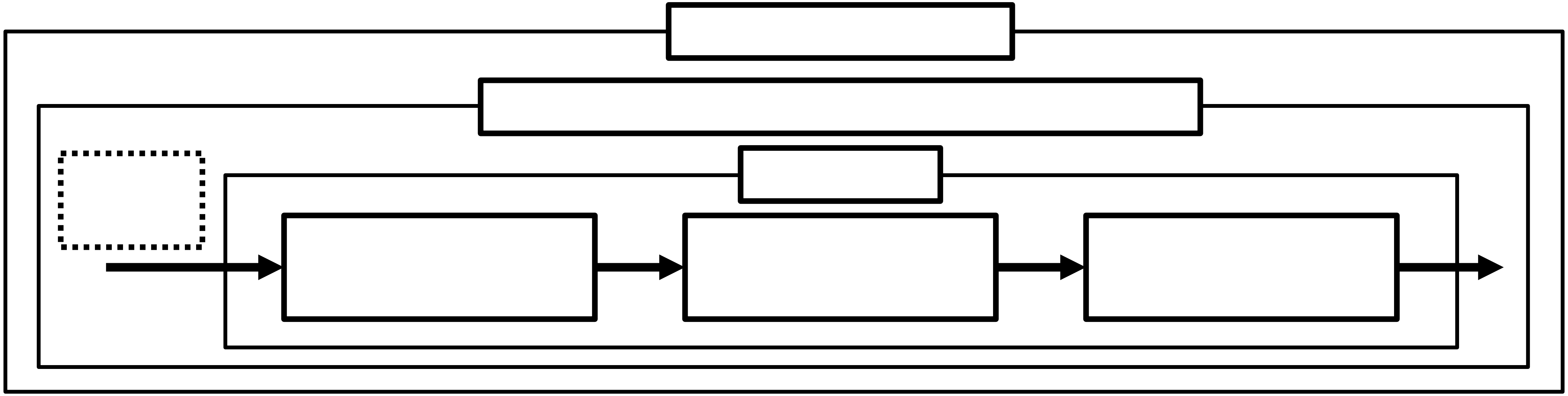
Components of a Productive ML Sniff Test

- Surface-level understanding of some core ML concepts
- Sniff test procedure
  - Construct your mental picture of their overall approach
    - High-level ML system initial mental picture
    - **Find the fuzzy-power words game to connect the story to the system**
  - Dig deeper and refine the edges
- General tips

Find the Fuzzy-Power Words Game

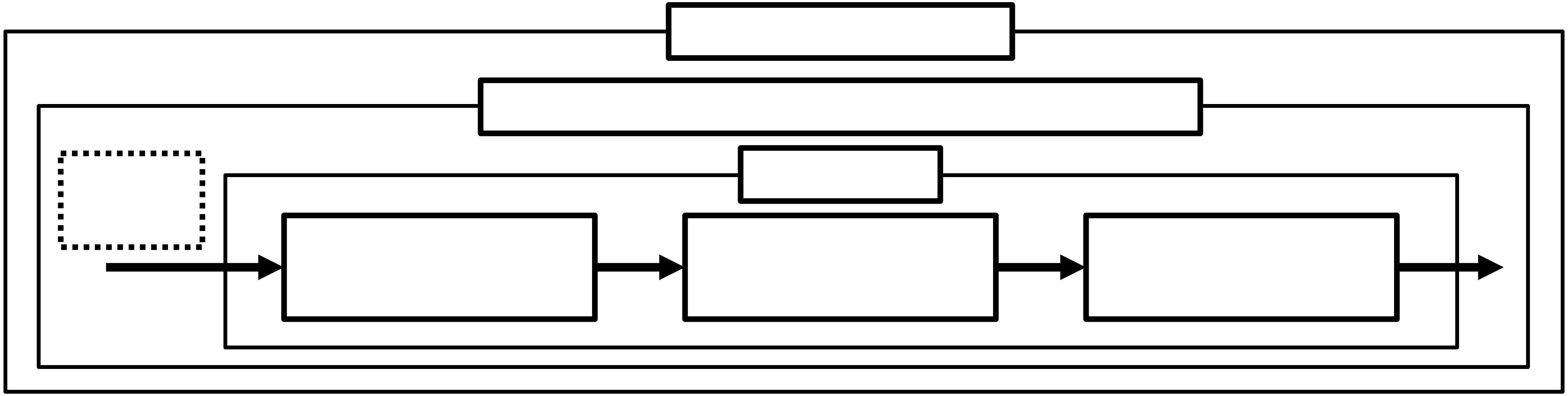
- Words/phrases that are said in a way say to convey of sense of
  - “The bullshitter doesn't care if what they say is true or false, but rather only cares whether or not their listener is persuaded”
- (For tech people) you’ve hit one if it’s not clear how it is implementable in a computer
- Best way to articulate this is by example
- (Names and details have been changed to protect the innocent)

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund



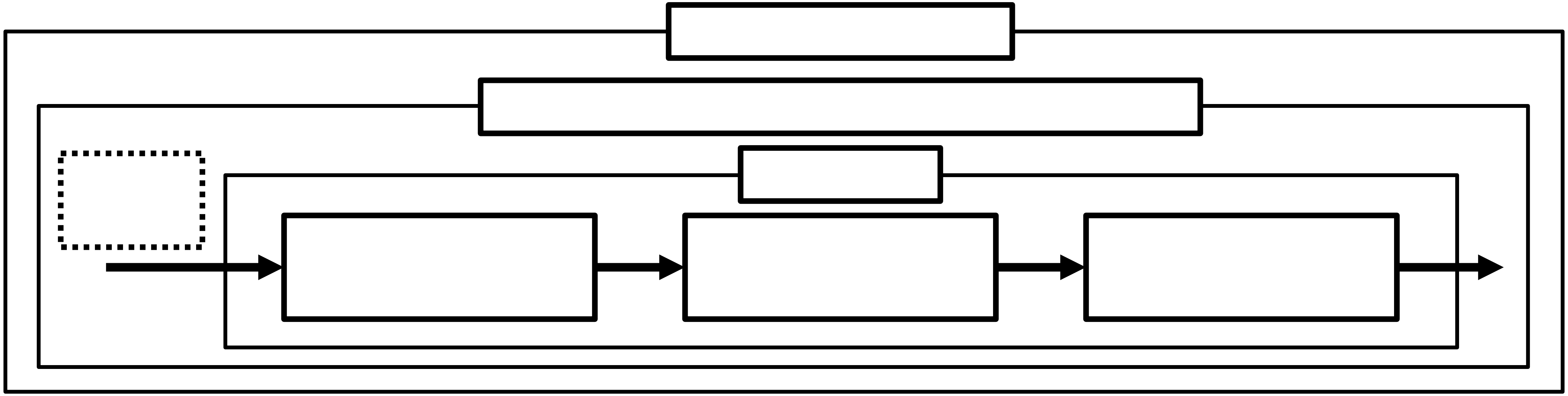
Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge



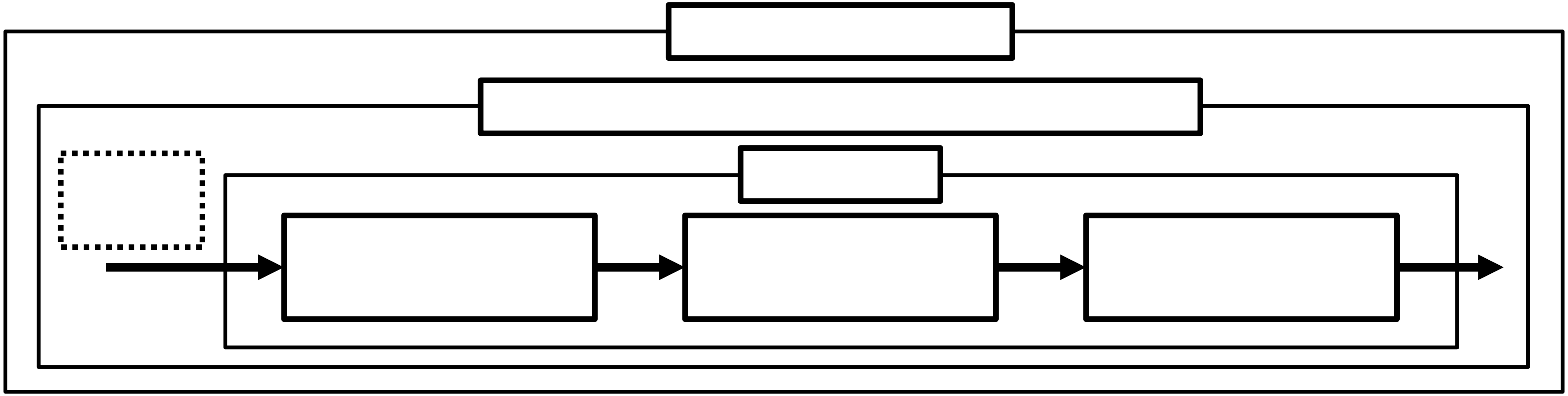
Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Sentiment of social media

Expert insights



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

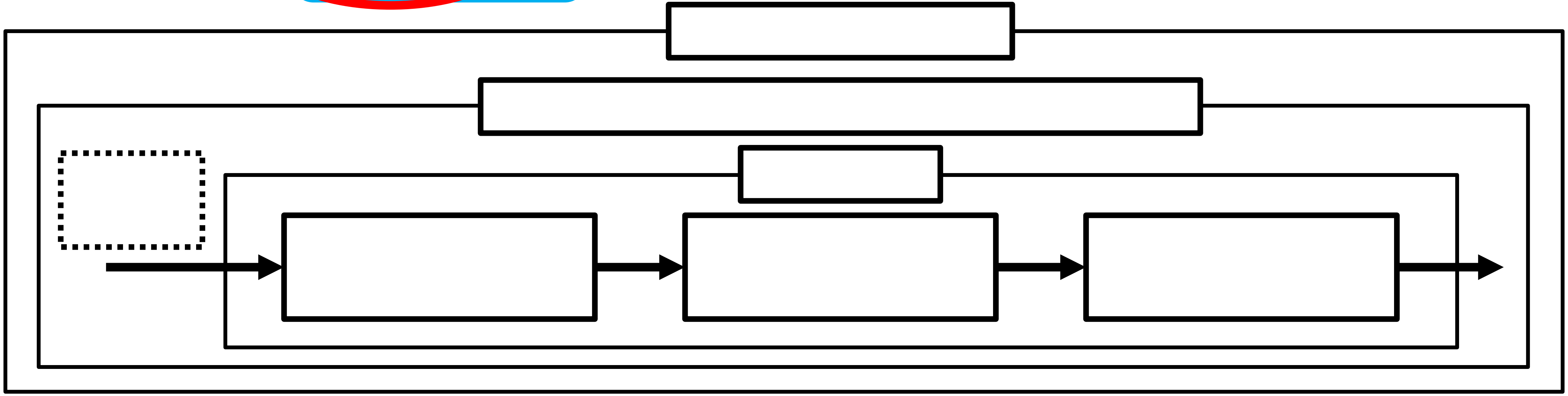
Sentiment of social media

Expert insights

Mathematical quantification of company health

Proprietary and curated tweets

Public expectations



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

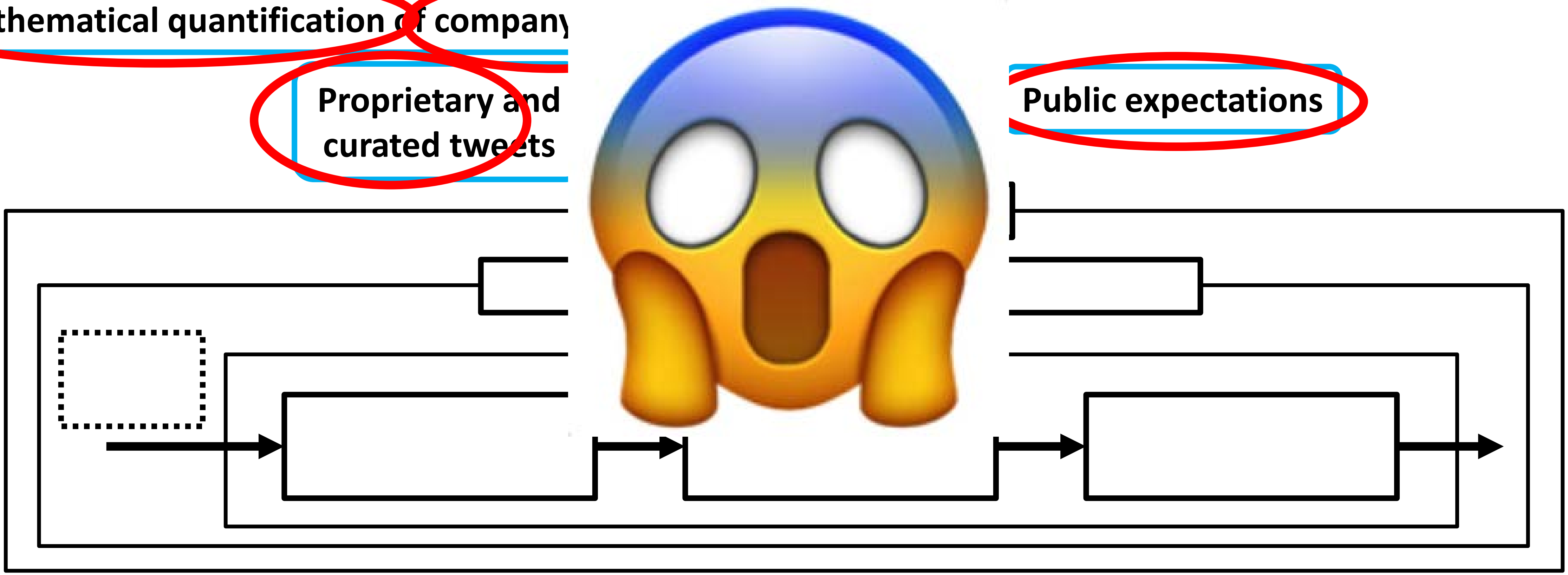
Sentiment of social media

Expert insights

Mathematical quantification of company

Proprietary and curated tweets

Public expectations



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

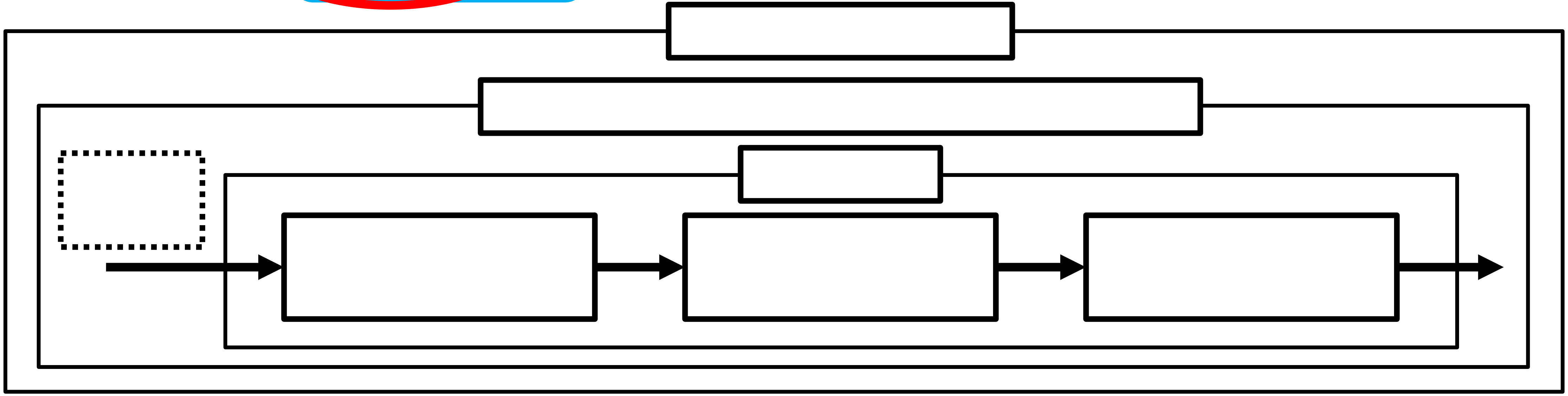
Sentiment of social media

Expert insights

Mathematical quantification of company health

Proprietary and curated tweets

Public expectations



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

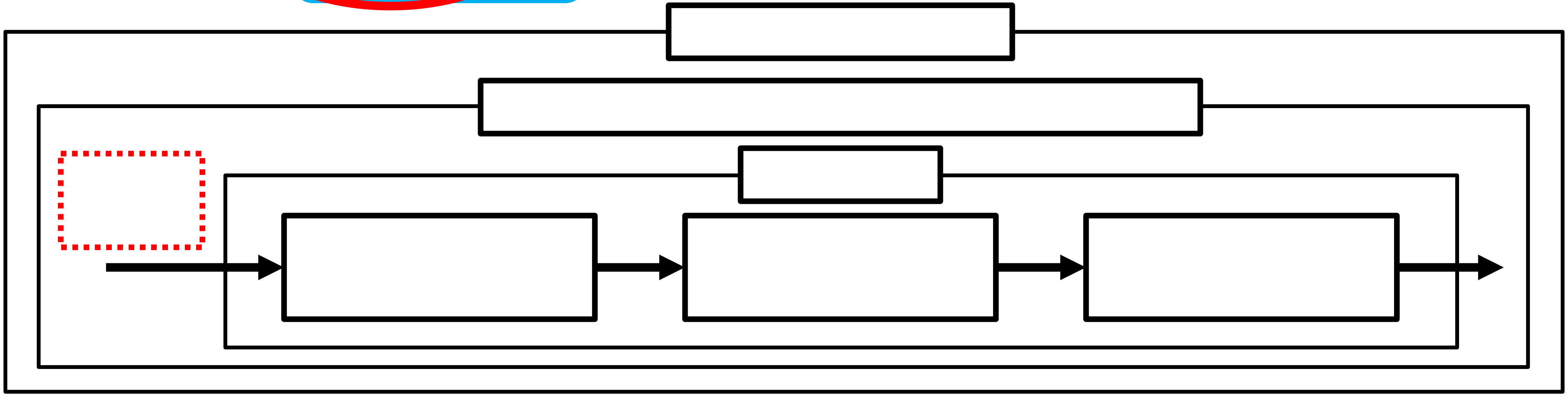
Sentiment of social media

Expert insights

Mathematical quantification of company health

Proprietary and curated tweets

Public expectations



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Sentiment of social media

Expert insights

Mathematical quantification of company health

Proprietary and curated tweets

Public expectations

Tweets from human-selected accounts →

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Sentiment of social media

Expert insights

Mathematical quantification of company health

Positive/negative

Proprietary and curated tweets

Public expectations

company fundamental data

Tweets from human-selected accounts

List of important words mapping to -1/+1 scores

Linear weighting, tuned by-human

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

company fundamental data

Tweets from human-selected accounts → List of important words mapping to -1/+1 scores

→ Linear weighting, tuned by-human → Threshold-based buy/sell decisions →

Performance Testing: Red flags

- They only show in-sample performance
  - (or don't clearly answer what is in-sample / out-of-sample)
- They don't have clear answers to expected future system performance with some understanding of the future performance's confidence
- They cite intuitive agreement as sufficient:
  - evidence of good future performance
  - justification for their methodology

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

company fundamental data → Linear weighting, tuned by-human

Tweets from human-selected accounts → List of important words mapping to -1/+1 scores → Linear weighting, tuned by-human

Linear weighting, tuned by-human → Threshold-based buy/sell decisions →

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

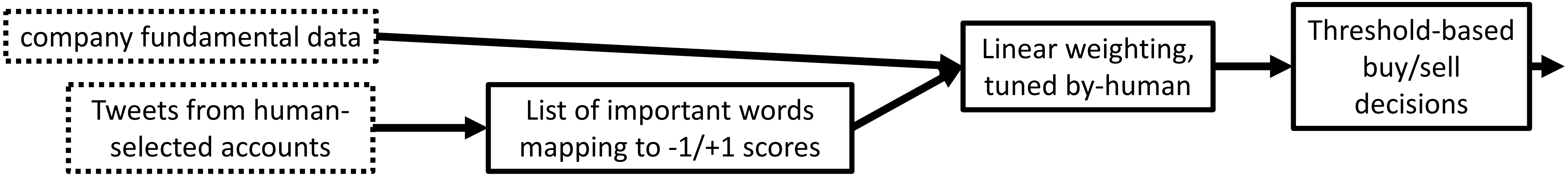
Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

Run backtest



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

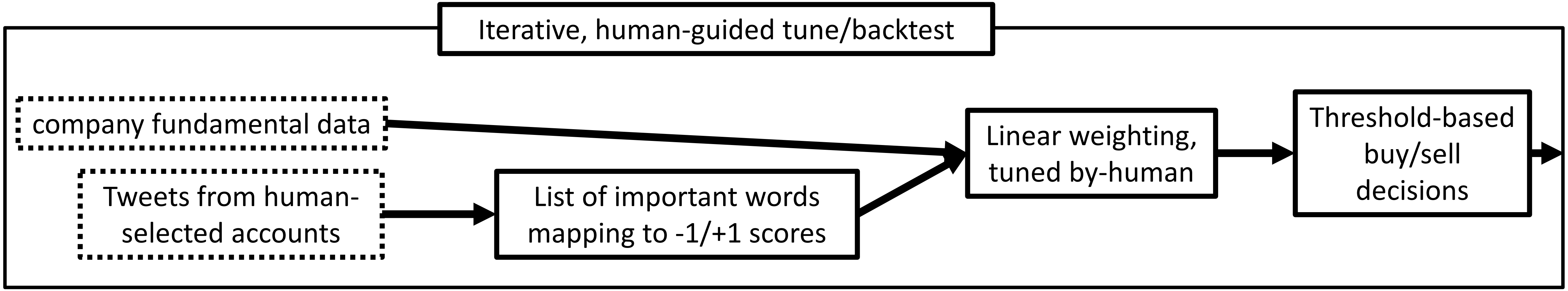
Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

Run backtest



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

Run backtest

No test/holdout

Iterative, human-guided tune/backtest

company fundamental data

Tweets from human-selected accounts

List of important words mapping to -1/+1 scores

Linear weighting, tuned by-human

Threshold-based buy/sell decisions

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

Run backtest

No test/holdout

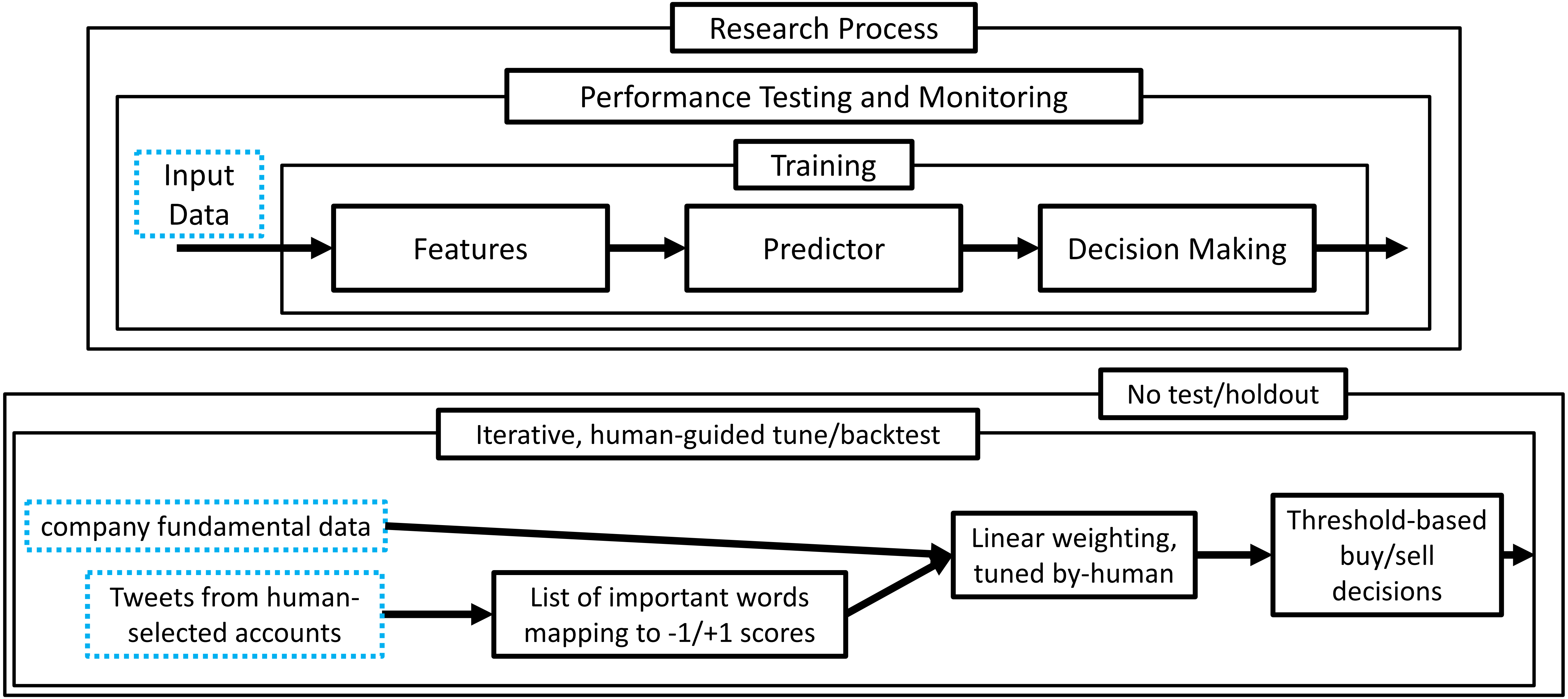
Iterative, human-guided tune/backtest

company fundamental data

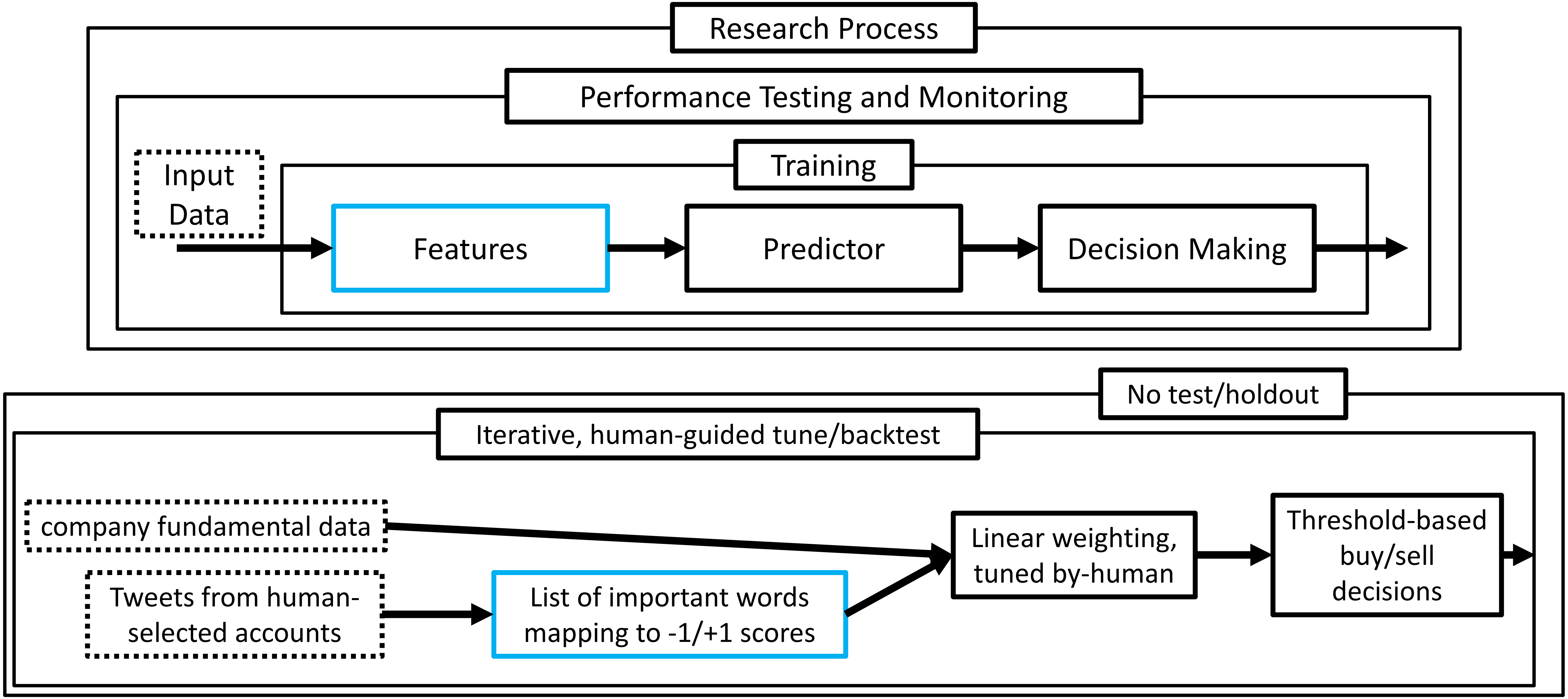
Tweets from human-selected accounts → List of important words mapping to -1/+1 scores

→ Linear weighting, tuned by-human → Threshold-based buy/sell decisions →

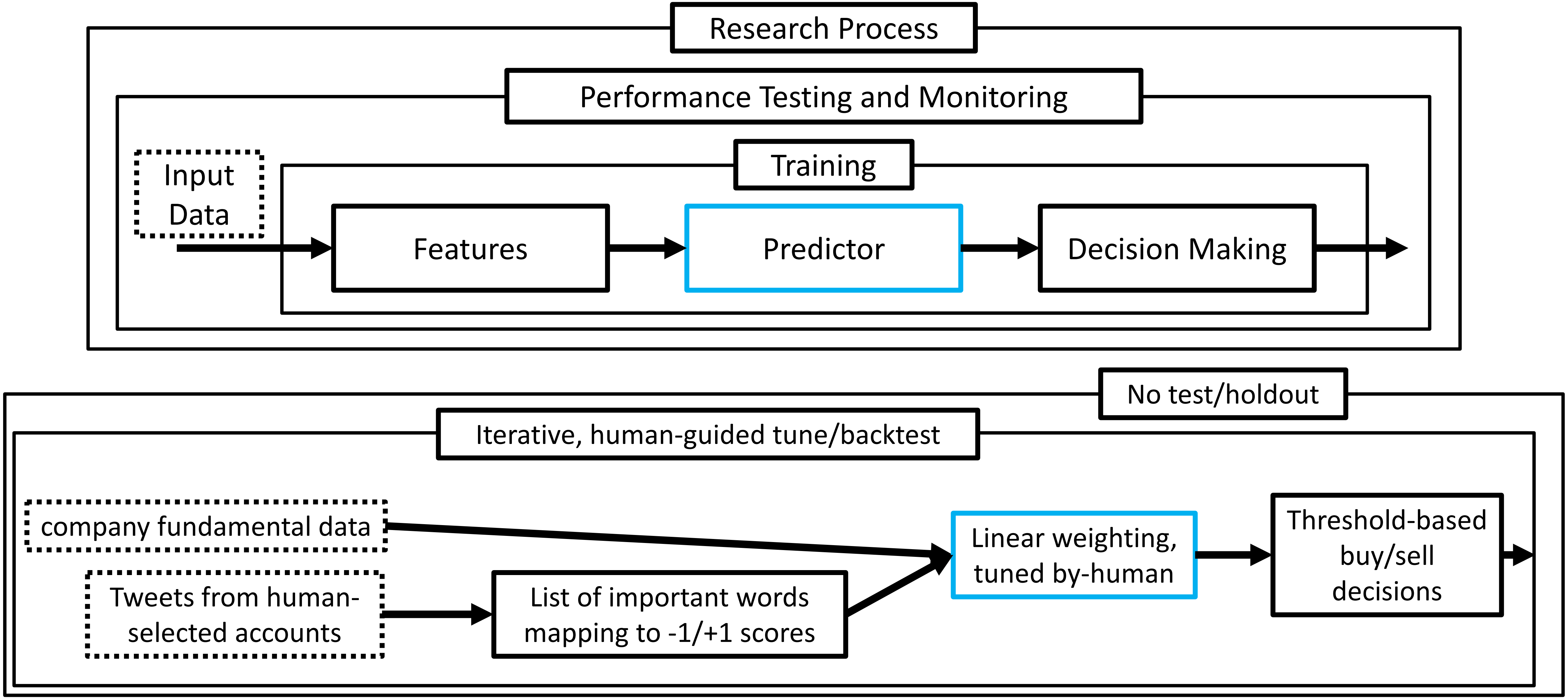
Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund



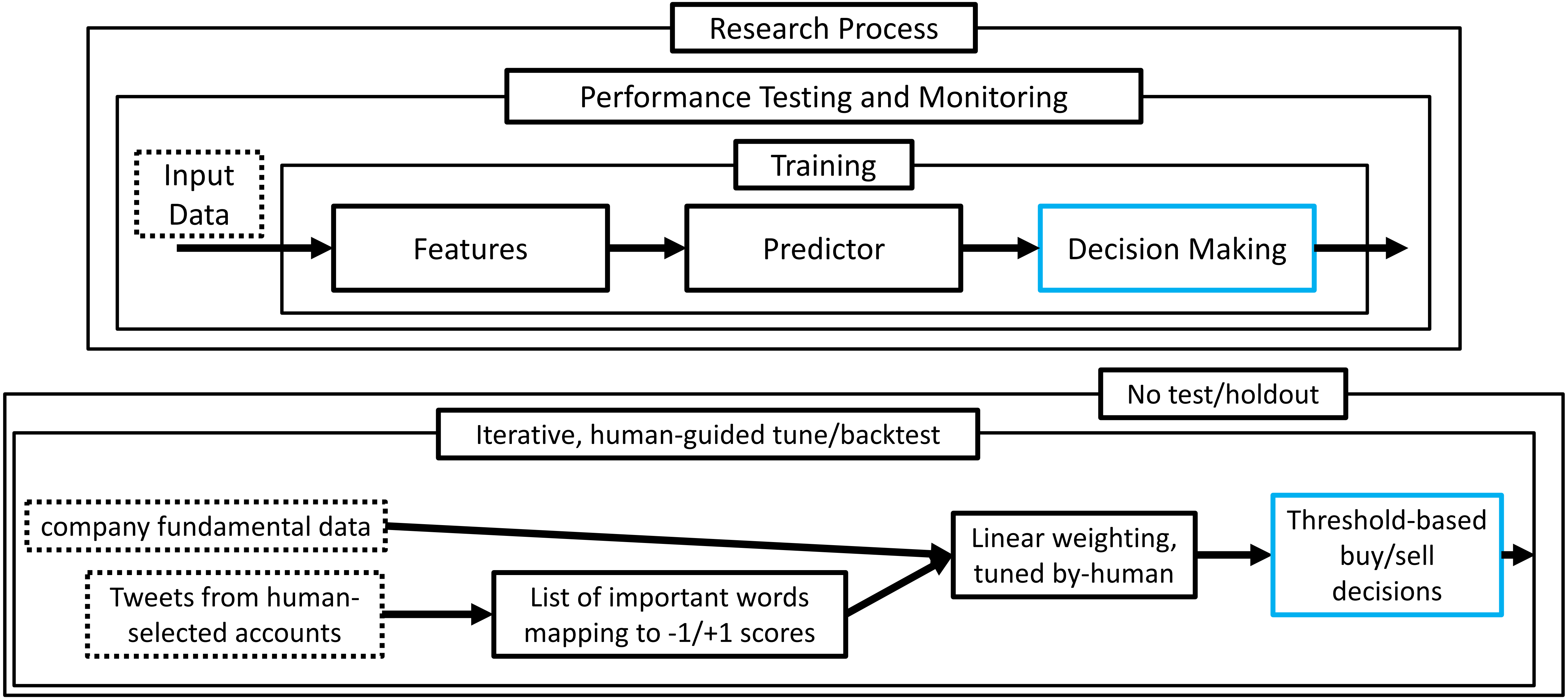
Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund



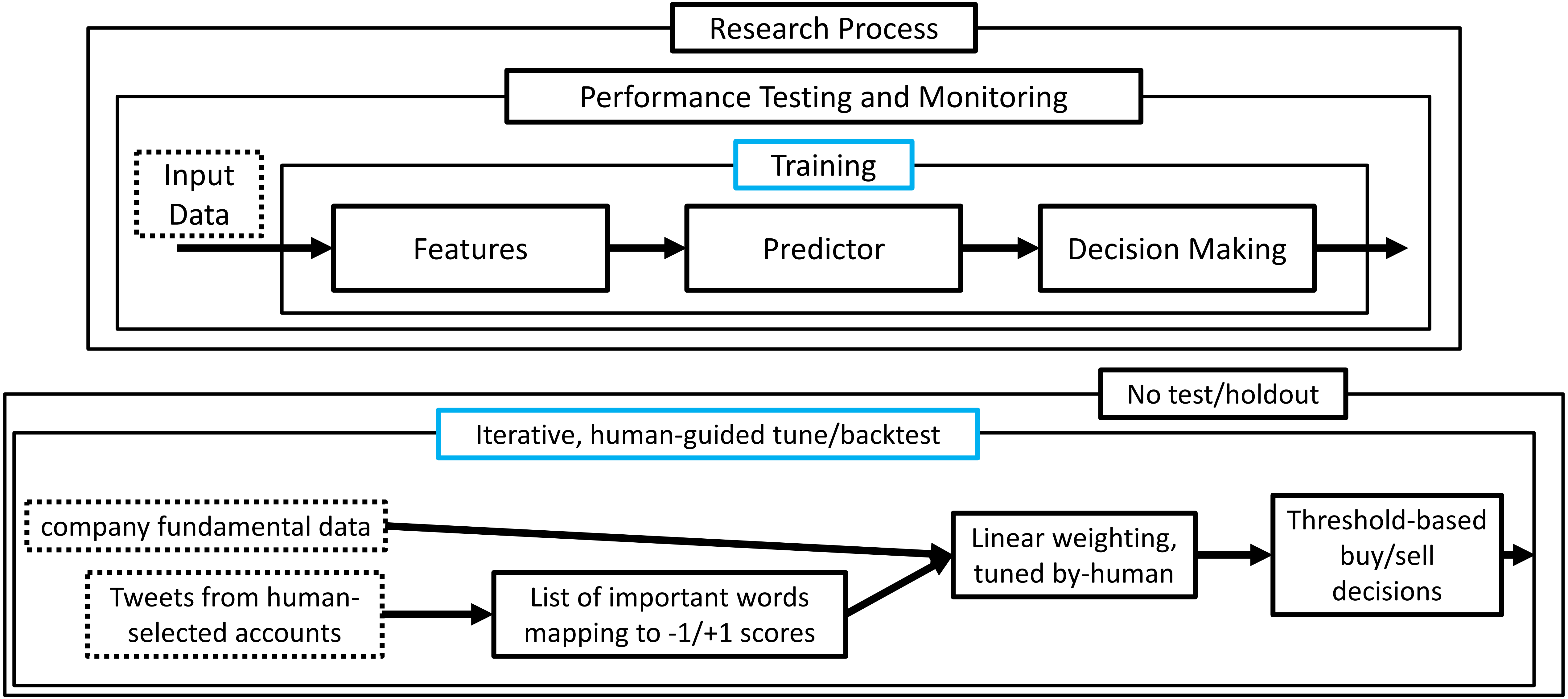
Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund



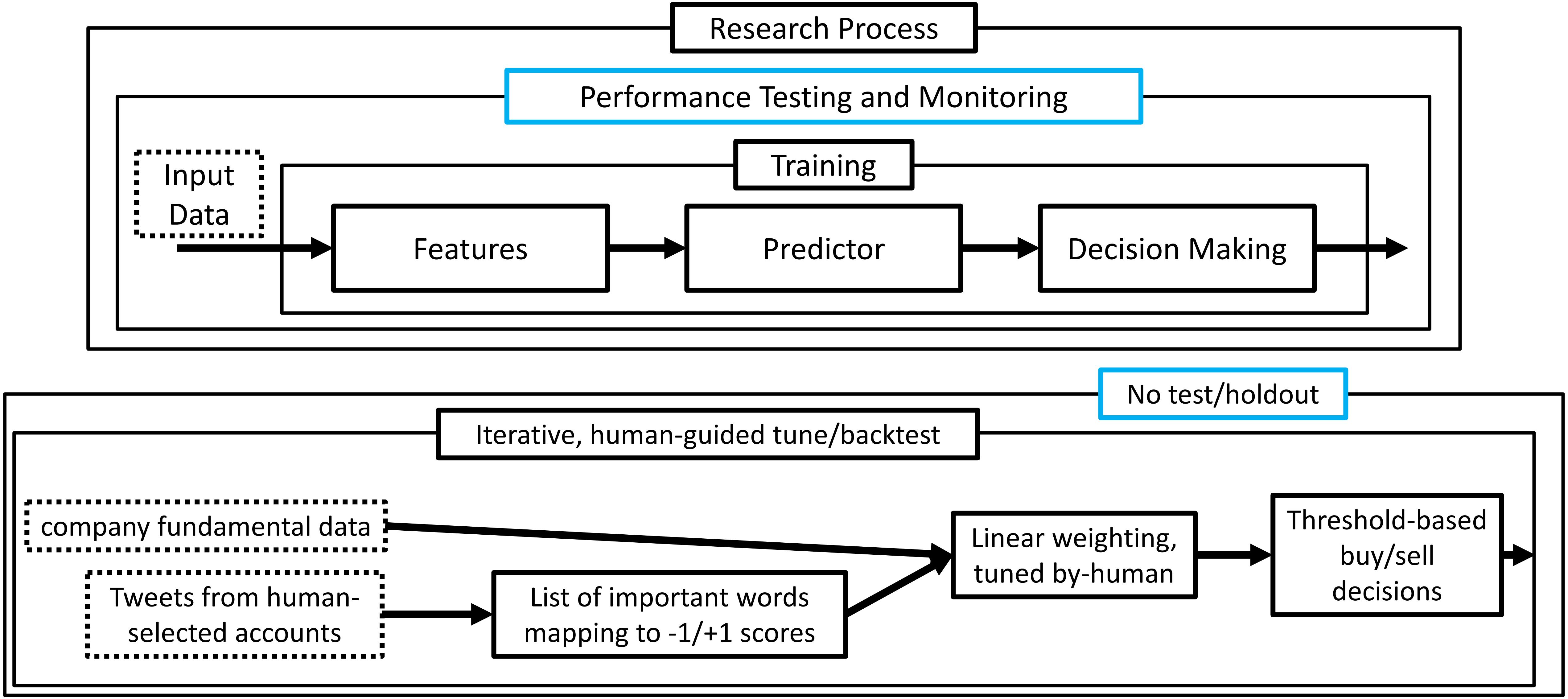
Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund



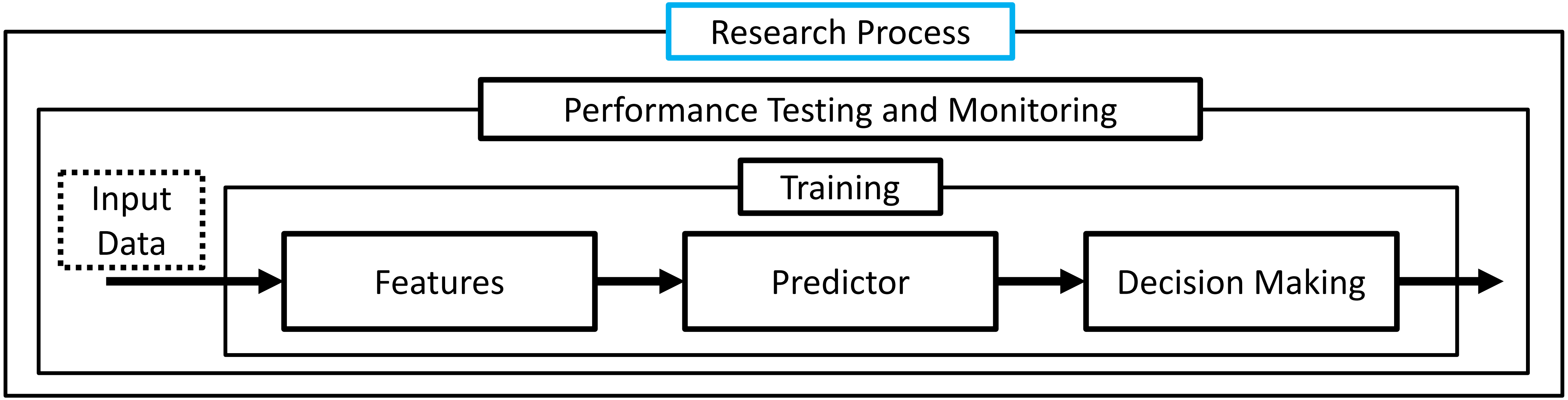
Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

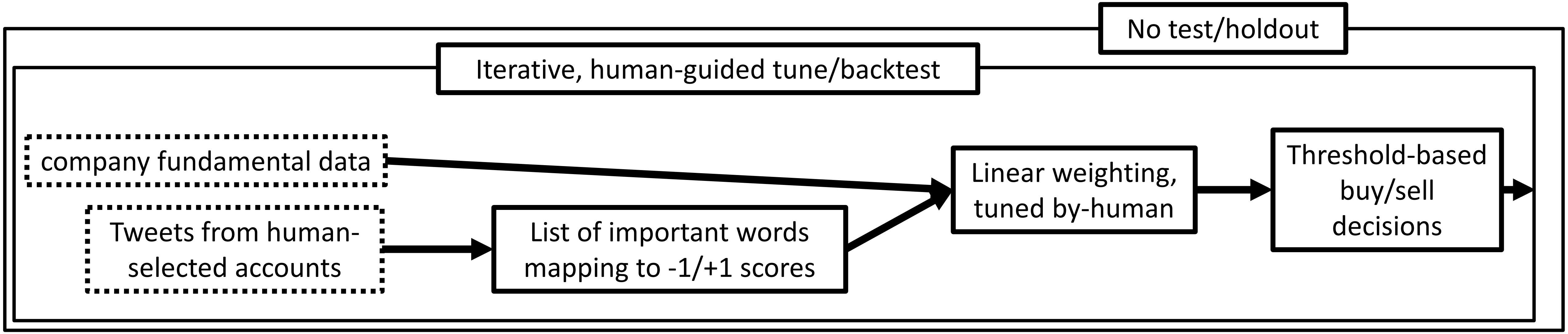


Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund



Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund





Find the Fuzzy-Power Words Game: Company DCC – Deep Crypto Capital

Deep Learning of the Blockchain

Blockchains store the full transaction history

No human involvement – deep learning discovers subtle patterns humans cannot see

Cryptocurrency market AI predictions

The same technology that revolutionized the IMAGENET competition

Its insight automatically selects what it believes has high conviction

Bitcoin and Ethereum price movements

Trained on specialized hardware

We trade on the confidence of price movements

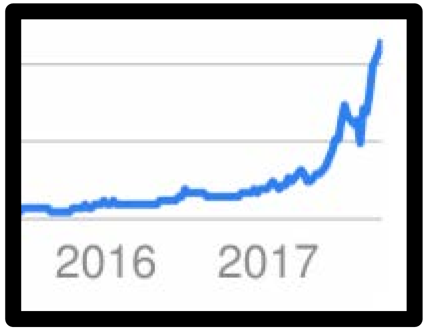
Compute performance on all the historical data

GDAX historical Bitcoin and Ethereum price and volume data

Training on all historical data

AlexNet

Buy if predict it will go up with >90%, sell to fiat if predict it will go down



Find the Fuzzy-Power Words Game: Company DCC – Deep Crypto Capital

Deep Learning of the Blockchain

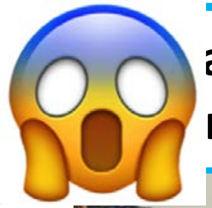
Blockchains store the full transaction history

No human involvement – deep learning discovers subtle patterns humans cannot see

Cryptocurrency market AI predictions

The same technology that revolutionized the IMAGENET competition

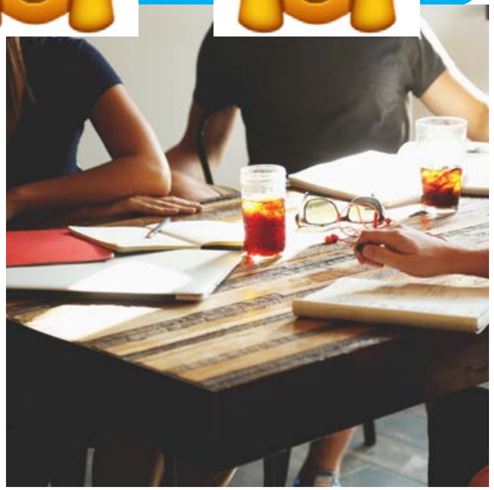
Its insight automatically selects what it believes has high conviction

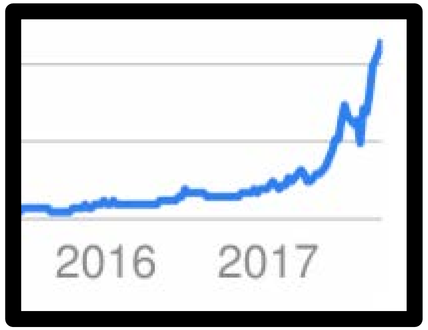


and n
mo

Trained on specialized hardware

We trade on the confidence of price movements





High probability fuzzy-power words

- Algorithm
- Analytics
- Artificial intelligence
- Automatic
- Autonomous
- Big data
- Classification
- Cognitive *
- Curated
- Data science
- Deep *
- Descriptor
- Detect
- Enrich
- Expert
- Indicator
- Insight
- Machine learning
- Method
- Model
- Novel
- Platform
- Prediction
- Reasoning
- Robust
- Signal
- Statistical *
- System
- Technique
- Technology
- Any human-like word
  - Thinks, knows, believes, understands, tries, etc.

Components of a Productive ML Sniff Test

- Surface-level understanding of some core ML concepts
- Sniff test procedure
  - Construct your mental picture of their overall approach
    - High-level ML system initial mental picture
    - Find the fuzzy-power words game to build the full tree of how the story and system connects
  - Dig deeper
    - **Probe into integrity gaps across hops**
- General tips

Integrity Gaps Across Hops

- Now look back at the diagram, how much integrity is there in each hop?

consistency

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

Run backtest

No test/holdout

Iterative, human-guided tune/backtest

company fundamental data

Tweets from human-selected accounts

List of important words mapping to -1/+1 scores

Linear weighting, tuned by-human

Threshold-based buy/sell decisions

Find the Fuzzy-Power Words Game: Company SWF - Social Wisdom Fund

AI Trading through the Combination of the Wisdom of the Crowd and Expert Knowledge

Cutting edge analytics

Autonomous trading

Sentiment of social media

Expert insights

Mathematical quantification of company health

Company fundamentals

Rigorous and robust results

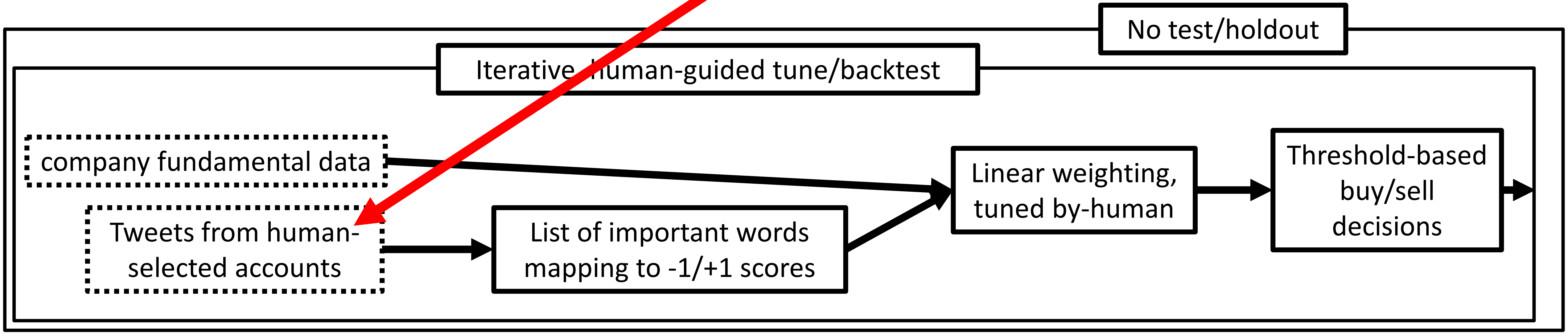
Positive/negative

Proprietary and curated tweets

Expert trading experience

Public expectations

Run backtest



Find the Fuzzy-Power Words Game: Company DCC – Deep Crypto Capital

Deep Learning of the Blockchain

Blockchains store the full transaction history

Cryptocurrency market AI predictions

Bitcoin and Ethereum price movements

No human involvement – deep learning discovers subtle patterns humans cannot see

The same technology that revolutionized the IMAGENET competition

Trained on specialized hardware

Its insight automatically selects what it believes has high conviction

We trade on the confidence of price movements

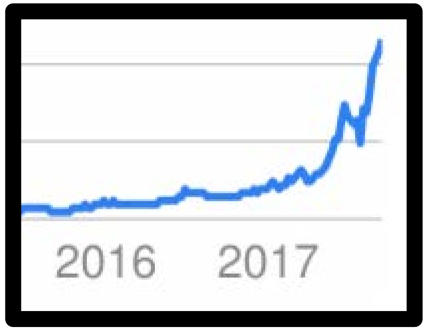
Compute performance on all the historical data

GDAX historical Bitcoin and Ethereum price and volume data

Training on all historical data

AlexNet

Buy if predict it will go up with >90%, sell to fiat if predict it will go down



Find the Fuzzy-Power Words Game: Company DCC – Deep Crypto Capital

Deep Learning of the Blockchain

Blockchains store the full transaction history

No human involvement – deep learning discovers subtle patterns humans cannot see

Cryptocurrency market AI predictions

The same technology that revolutionized the IMAGENET competition

Its insight automatically selects what it believes has high conviction

Bitcoin and Ethereum price movements

Trained on specialized hardware

We trade on the confidence of price movements

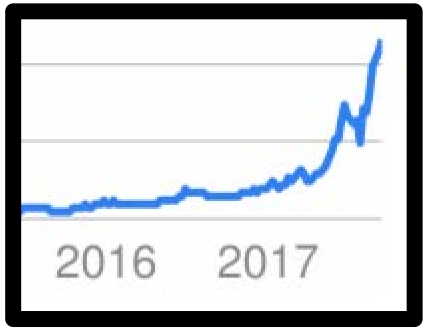
Compute performance on all the historical data

GDAX historical Bitcoin and Ethereum price and volume data

Training on all historical data

AlexNet

Buy if predict it will go up with >90%, sell to fiat if predict it will go down



Integrity Gaps Across Hops

- The purpose here is to discover if these “integrity gaps” are that:
  - They simplified it for our benefit
  - They don’t view the gap as a gap and it’s our misunderstanding
  - They are trying to deceive us

Components of a Productive ML Sniff Test

- Surface-level understanding of some core ML concepts
- Sniff test procedure
  - Construct your mental picture of their overall approach
    - High-level ML system initial mental picture
    - Find the fuzzy-power words game to build the full tree of how the story and system connects
  - Dig deeper and refine the edges
    - Probe into integrity gaps across hops
    - **Understand how well thought through each concrete box is with lists of questions**
- General tips

Lists of questions on each component of the system

Input data:
* Where does the data come from?
* How much history do you have?
* What is the resolution of your data?
* Are there any large gaps or outages in your data?
* What kind of sanity checking, cleaning, and outlier detection/removal do you do?
* How do you check for changes in the data format? How many times has that happened?
* What is the biggest source of noise in the data?
* Who else has access to the same data?
* Has anyone used data like yours to solve a problem like yours?
* How much "human curation" is in your data?
* What filtering is in place?
* How is the data stored?
* Are there sporadic performance-effecting latencies in your data arrival?
* Can you give me specific examples of what your data actually is?
* Do you believe there is systematic noise in your data somehow? How do you correct for it?

Features:
* How do you represent features in your system?
* How often do you add/change/remove features?
* Can you give me three concrete examples of features that are currently live in your system?
* How much human domain knowledge is imbedded in your features?
* Do you normalize/transform your features somehow?
* How do you handle heterogenous data sources?
* Roughly, how many features do you use?
* (If feature selection, dimensionality reduction, etc. methods are used)
  * Why were these methods chosen?
  * What effect do they have on the system's overall performance?
  * Why do these features make sense for the decisions you want to make?

Predictor:
* What is the meaning of the prediction output?
* What type of representation do you use to go from features to a prediction?
* On what timescales is your system outputting predictions?
* How does your system quantify its uncertainty in a prediction?
* Why did you chose that representation from the many, many of other methods?
* What alternate methods did you try? Why were those passed on?
* How much human domain knowledge is imbedded in your prediction representation?
* Do you explicitly regularize or measure/control the complexity your representation somehow?
* What are other successful applications of your chosen representation?
* Why does your predictor representation make sense for the decisions you want to make?
* If more than one predictor (for example, using ensembles):
  * Roughly how many models do you use?
  * Do you find most of your performance comes from a few of the models?
  * How do you validate that you increase performance by adding additional models?

Decision making:
* What is the meaning of the decisions you are making?
* On what timescales is your system outputting decisions?
* What type of representation do you use to go from a prediction to a decision?
* What alternate methods did you try? Why were those passed on?
* How much human domain knowledge is imbedded in your decision making representation?
* How does your system quantify its uncertainty in a prediction?
* Why does your decision making representation make sense for the decisions you want to make?

Training:
* What parts of your system "learn" from data?
* Where do your labels come from? How accurate are they?
* How much human domain knowledge is in the fitting process?
* What is the rough ratio of (# data points)/(# features)?
* What objective function do you use? Why? How did it come about? How much tuning was involved?
* Roughly what order of magnitude of parameters are you fitting?
* What are alterative objective functions you have tried?
* How do you actually perform the fitting/learning/search/optimization?
* What alternative optimization techniques have you previously tried? Why were those insufficient?
* How do you understand if your system is overfitting/underfitting?
* How often do you re-fit your representations? How much does that increase your performance?

Performance testing:
* How do you test the performance of your overall system?
* How do you test the performance of single components of your system (e.g., the decision making)?
* Exactly how large is your train/test/holdout sets? How are they kept separate?
* What assumptions is your overall approach making? How have you validated these assumptions hold?
* What are the metrics you use to measure performance across your system?
* What is your overall performance? With confidences.
* How do common baseline methods perform on your problem? With performance and confidences.
* Can you explain a couple situations of both unexpectedly high and low performance?
* What do you believe is the maximum achievable performance? Why?
* How have the dynamics of your data/problem changed over time?
* How much do your estimated and live performances differ? What are the sources of these differences? How do you represent and compensate for them?
* Do you validate against simulated data where you know your assumptions hold? How is the simulated data generated?
* Does your performance make intuitive sense to you? Why or why not?

Monitoring for performance changes:
* How are you monitoring for performance changes?
* How often do you expect to detect changes in performance?
* Do you have methods for early-detection of performance changes for each individual component before it shows up in overall performance?
* When has your monitoring caught a change in performance?
* How do you think about the difference between an anomalous environmental condition and a change in performance?

Research process:
* Describe your overall research philosophy
* For any change that is made to the overall system (method, implementation, etc), what is the process that validate and approves the change?
* How often is the methodology or a component of the system changed?
* Who has access to the entire data set? What policies are in place to prevent data snooping?
* What change made it through your validation process that turned out to decrease performance? Why did that happen? How was your process changed to prevent this from happening again in the future?
* What is the likely next research piece to make it into production?
* What roles comprise your research team? Why?
* How do you allocate time across your team?
* What part of your system keeps you up at night? Why?
* What has worked much better/worse than you had expected?
* How is your live system influencing/corrupting your future data? How are you correcting for that?
* What other data sources have you investigated?
* How closely does your in-sample data represent your out-of-sample?

github.com/alpha-features/oreilly-sf-ai-conference-2017/

(Don't worry, you don't have to read it all. It's there so you have lists of questions for later.)

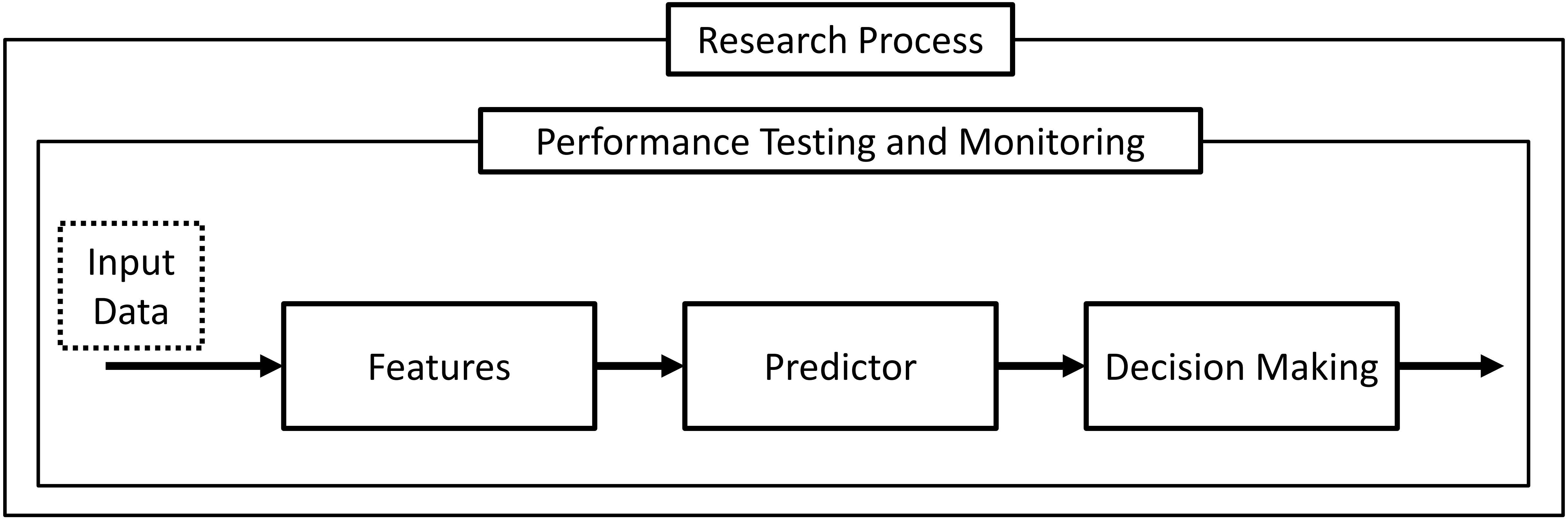
Components of a Productive ML Sniff Test

- Surface-level understanding of some core ML concepts
- Sniff test procedure
  - Construct your mental picture of their overall approach
    - High-level ML system initial mental picture
    - Find the fuzzy-power words game to build the full tree of how the story and system connects
  - Dig deeper and refine the edges
    - Probe into integrity gaps across hops
    - Understand how well thought through each concrete box is with lists of questions
- **General tips**

General Tips

- Tricks to pull out in the meeting
  - Purposefully ask a wrong question 1 out of 5 times (super-power)
  - Jump on strangely out of place, overly precise jargon
- Helpful mindsets
  - You are on the easy side of the table (burden of proof is on the seller)
  - We are uninterested in inferring whether anyone is smart or stupid
    - At worst, their story and methodology gives us no information about their future performance

What about an AI system?



Josh, I feel like I'm still missing half the story here…

A Practical Guide to Conducting an AI Snake Oil Sniff Test

Understanding what they're doing

- **Understanding of ML core concepts**
- **Sniff test procedure**
- **General tips**
- **Initial ML system mental picture**
- **Find the fuzzy-power words**
- **Integrity gaps**
- **Lists of questions**

Evaluation of what they're doing

- **Some niche benchmarks**
- **Run a trial**
- **It's the "best" one based on ad-hoc qualities**
- **Benchmarks that industries care about**
- **Detailed success/failure case studies**
- **AI/ML best practices (by industry?)**
- **Why do they think this works and do we agree with their thinking?**

github.com/alpha-features/oreilly-sf-ai-conference-2017/