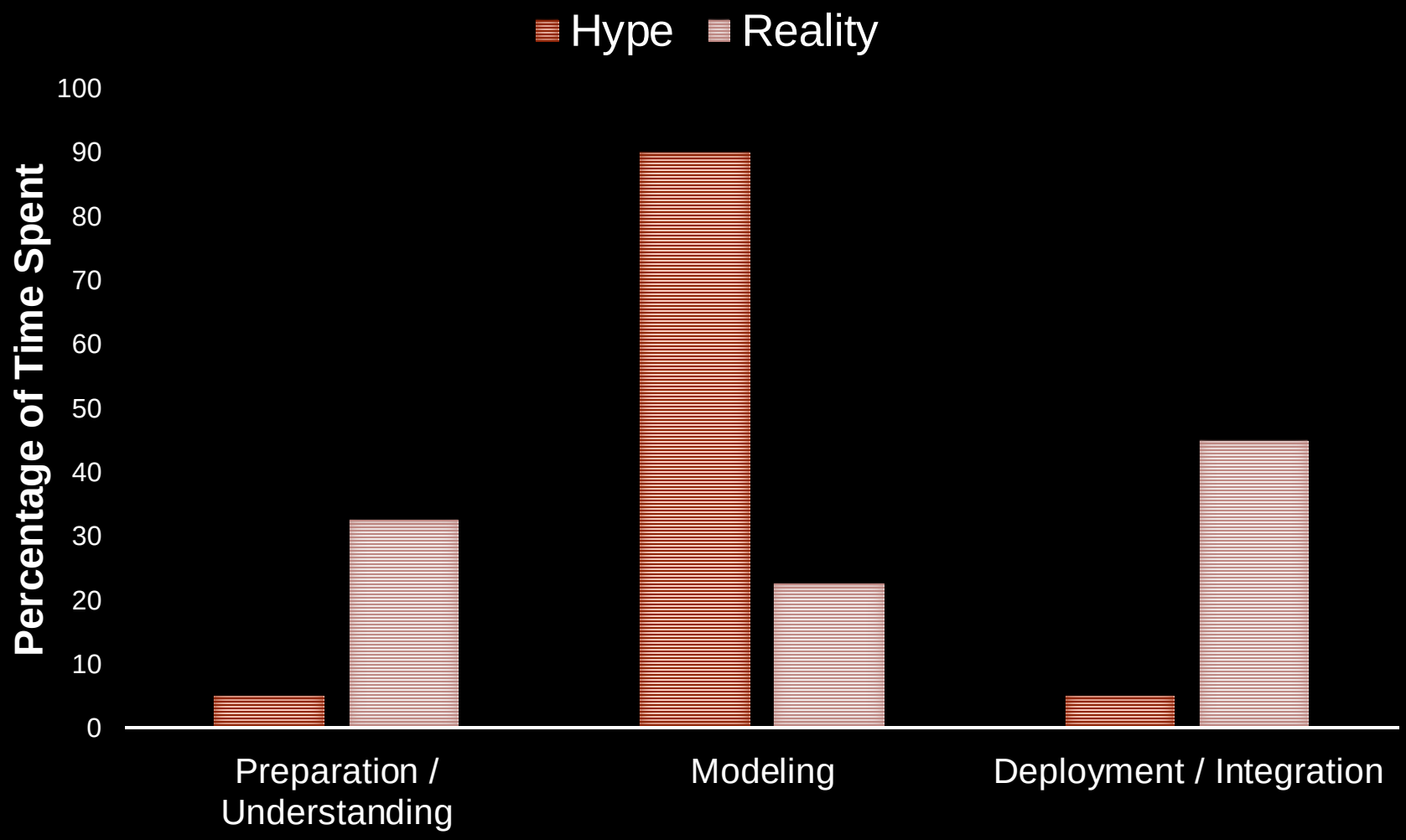
INTRODUCTION TO FRAUD

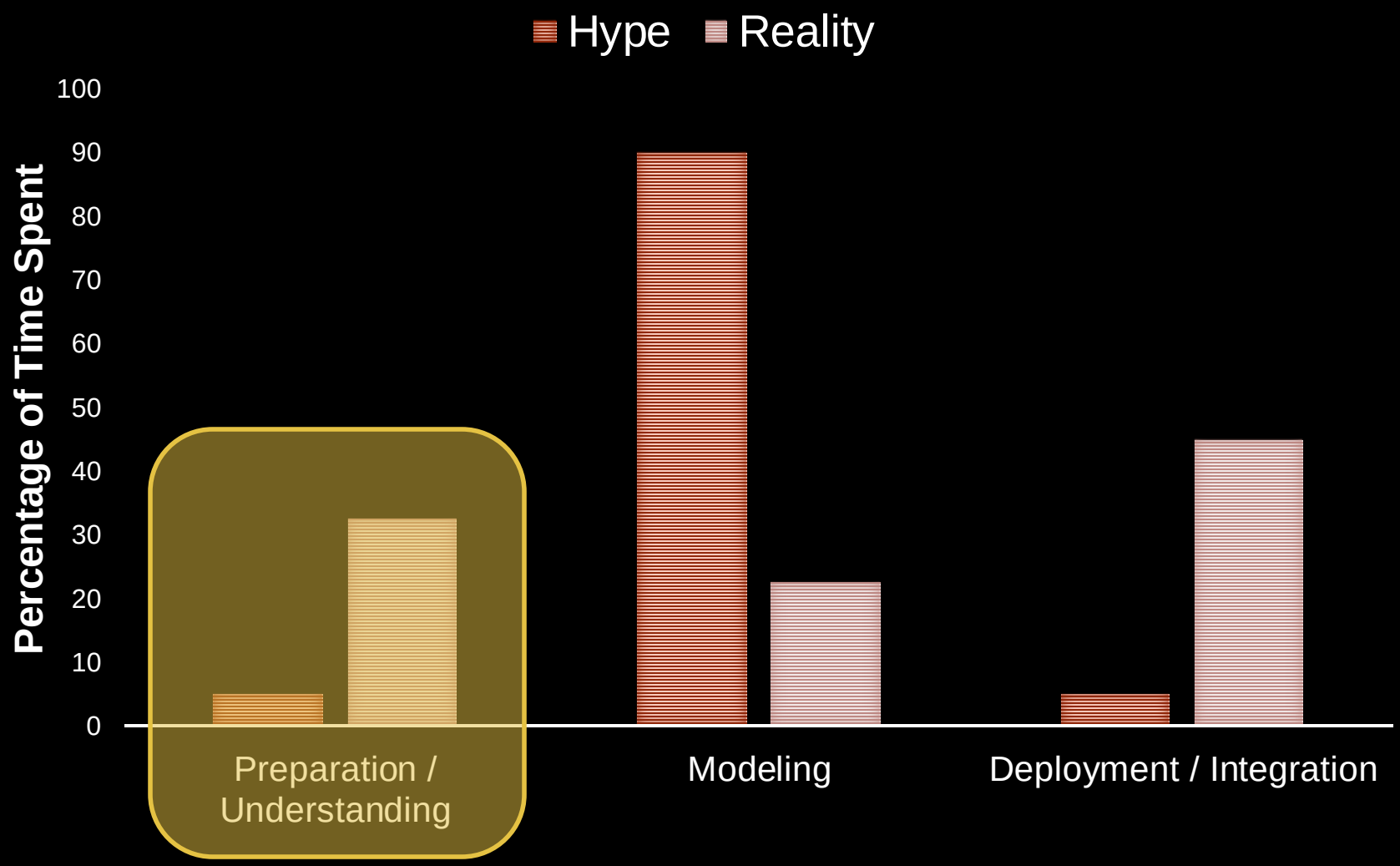
Dr. Aric LaBarr

Institute for Advanced Analytics

Data Science Hype vs. Reality



Data Science Hype vs. Reality



FRAUD PROBLEM

What is Fraud?

- Oxford English Dictionary:

  Fraud – Wrongful or criminal deception intended to result in financial or personal gain.

- Definition is not too helpful when determining how to set-up and solve the problem…

What is Fraud?

- Analytics Papers:

  Fraud – Uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types of forms.

- Definition provides a LOT more insight to things we need to account for in a solution to fraud.

Fraud Characteristics

- 5 Main Characteristics of Fraud:
  1. Uncommon / rare
  2. Well considered & concealed
  3. Evolving over time
  4. Carefully organized
  5. Many forms

1. Uncommon

- In 2023, the ACFE (Association of Fraud Examiners) estimated that organizations lose approximately 5% of their revenues to fraud.
- Based on 2023 world GDP (IMF estimates) this would mean approximately $5.28 trillion is lost each year due to fraud.

1. Uncommon

- Identifying fraud can be extremely difficult because fraud is a rare event.
- Rare event modeling:
  - 5% or less target
  - Limited number of KNOWN fraud
  - Undersampling, oversampling, SMOTE, etc.

2. Well Considered / Concealed

- In non-fraud data sets, observations are indifferent to be analyzed and discovered.
- In fraud data sets, observations are **trying to not be analyzed** or discovered – blending in.
  - Planned ahead of time – otherwise easier to detect in modeling.

3. Evolving Over Time

- In non-fraud data sets, observations are indifferent to be analyzed and discovered.
- In fraud data sets, observations are **trying to not be analyzed** or discovered – blending in.

↓

Models have short shelf lives.

3. Evolving Over Time

- In non-fraud data sets, observations are indifferent to be analyzed and discovered.
- In fraud data sets, observations are **trying to not be analyzed** or discovered – blending in.

↓

Models have short shelf lives.

↓

Models must be adapted often.

3. Evolving Over Time

- In non-fraud data sets, observations are indifferent to be analyzed and discovered.
- In fraud data sets, observations are **trying to not be analyzed** or discovered – blending in.

↓

Models have short shelf lives.

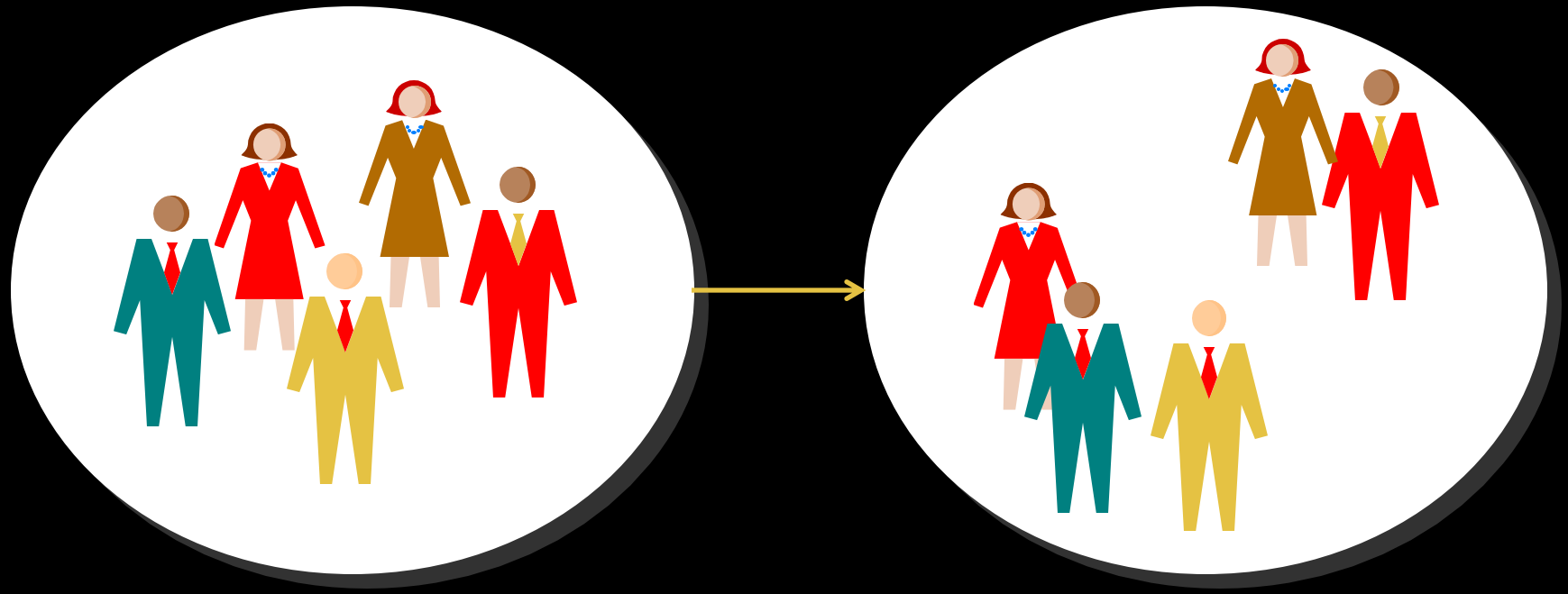
↓

Models must be adapted often.

- Cat-and-mouse game.

4. Carefully Organized

- J L Moreno founded a social science called **sociometry**, where **sociometrists** believe that society is made up of individuals **and** their social, economic, or cultural ties.



4. Carefully Organized

- J L Moreno founded a social science called **sociometry**, where **sociometrists** believe that society is made up of individuals **and** their social, economic, or cultural ties.

- Fraud is often an organized crime.
  - No independence
  - Copycat
  - Homophily: “Birds of a feather flock together.”

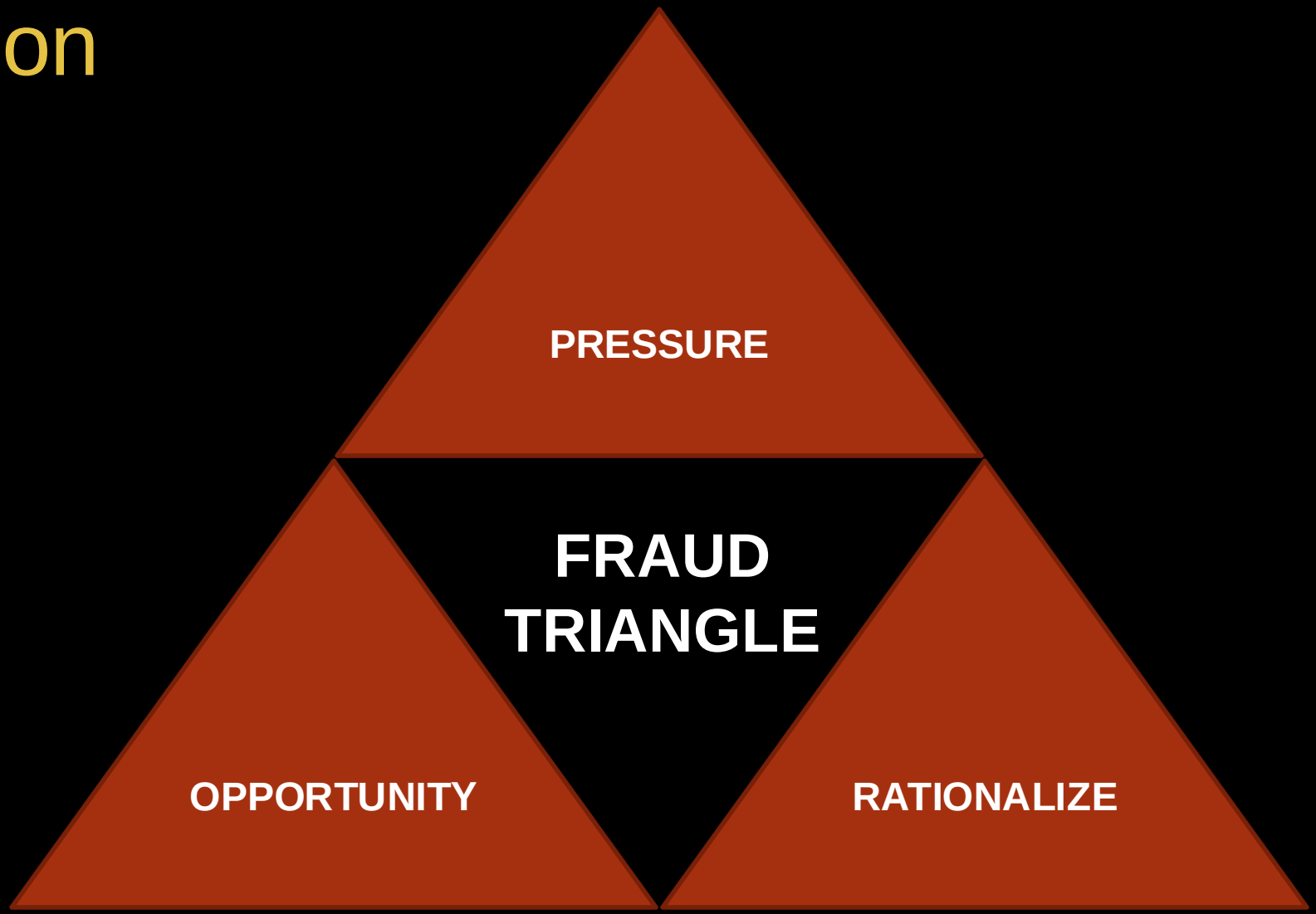
5. Many Forms

- Fraud is ever changing and comes in a variety of forms.
- The technology, economic, and social structures of today provide more and more opportunities for fraudulent activities to occur.

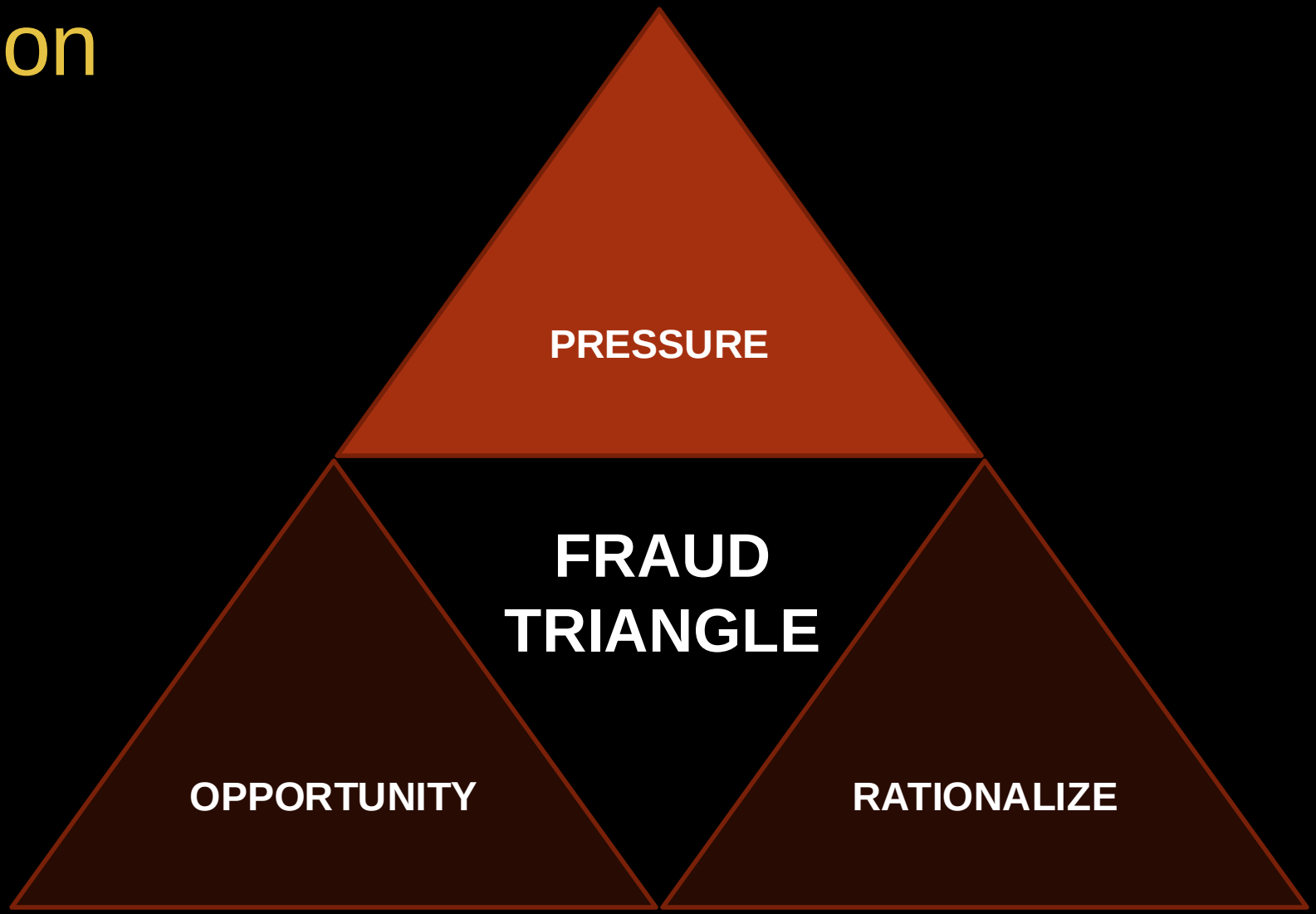
5. Many Forms

- Fraud can occur in many industries and across many aspects of industries:
  - Credit card fraud
  - Insurance fraud
  - Counterfeit
  - Healthcare fraud
  - Money laundering
  - Identify theft
  - Tax evasion

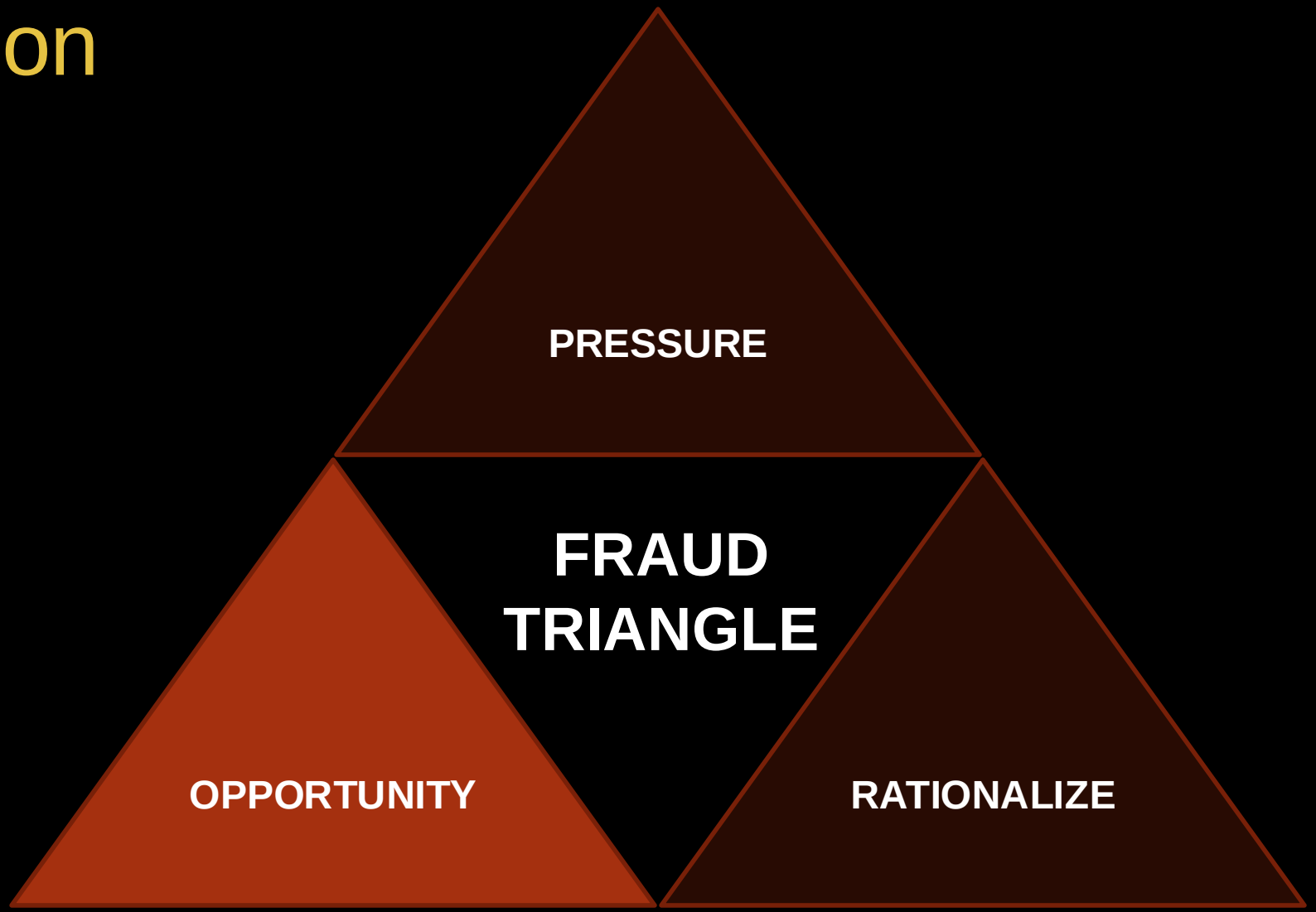
Motivation



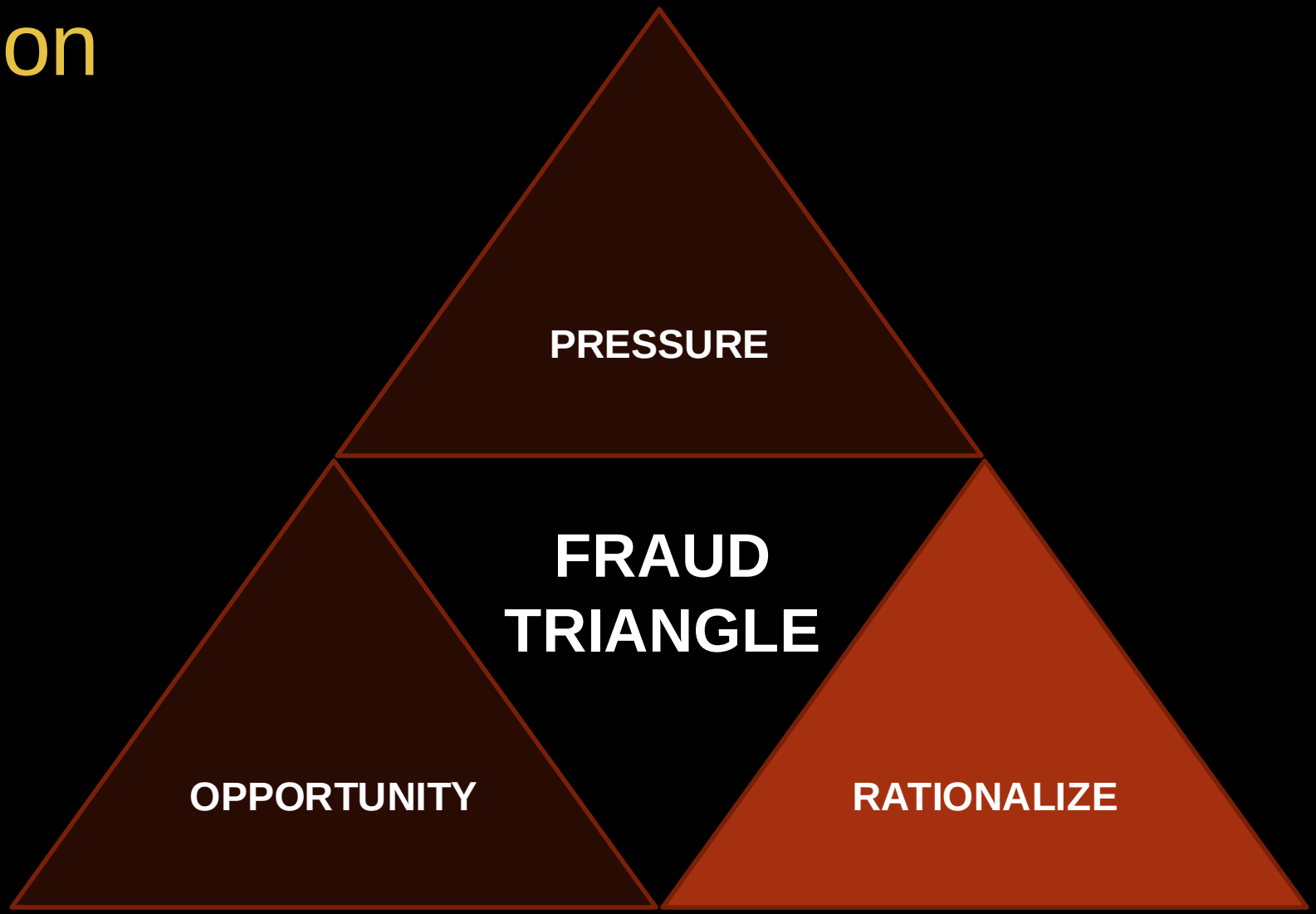
Motivation



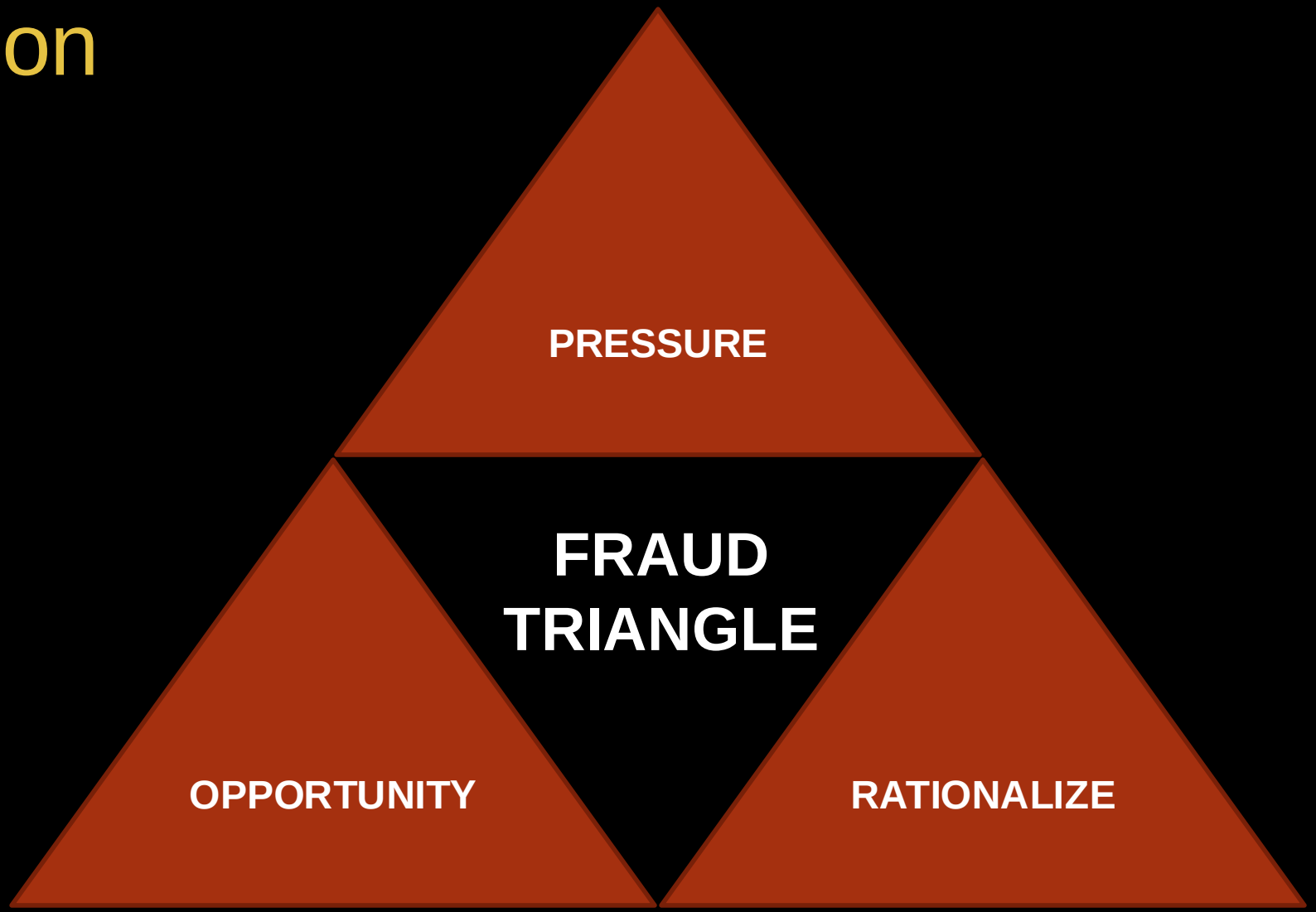
Motivation



Motivation



Motivation



FRAUD DETECTION & PREVENTION

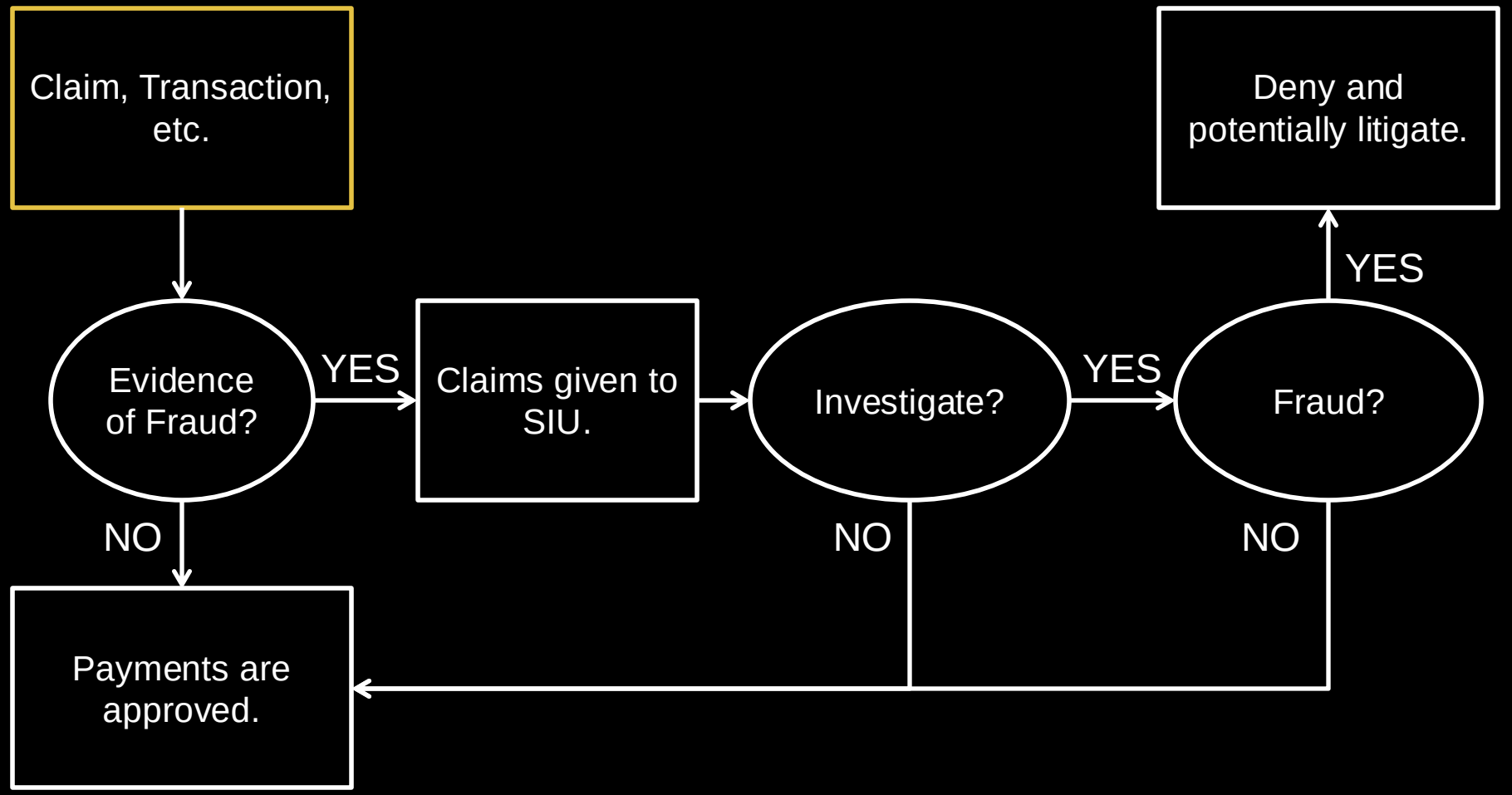
The Fraud Solution

- Regardless of the industry, two things are important for any fraud detection solution:
  1. **DETECTION** – Observing **known** fraudulent observations to determine patterns that may assist in finding other fraudulent observations.

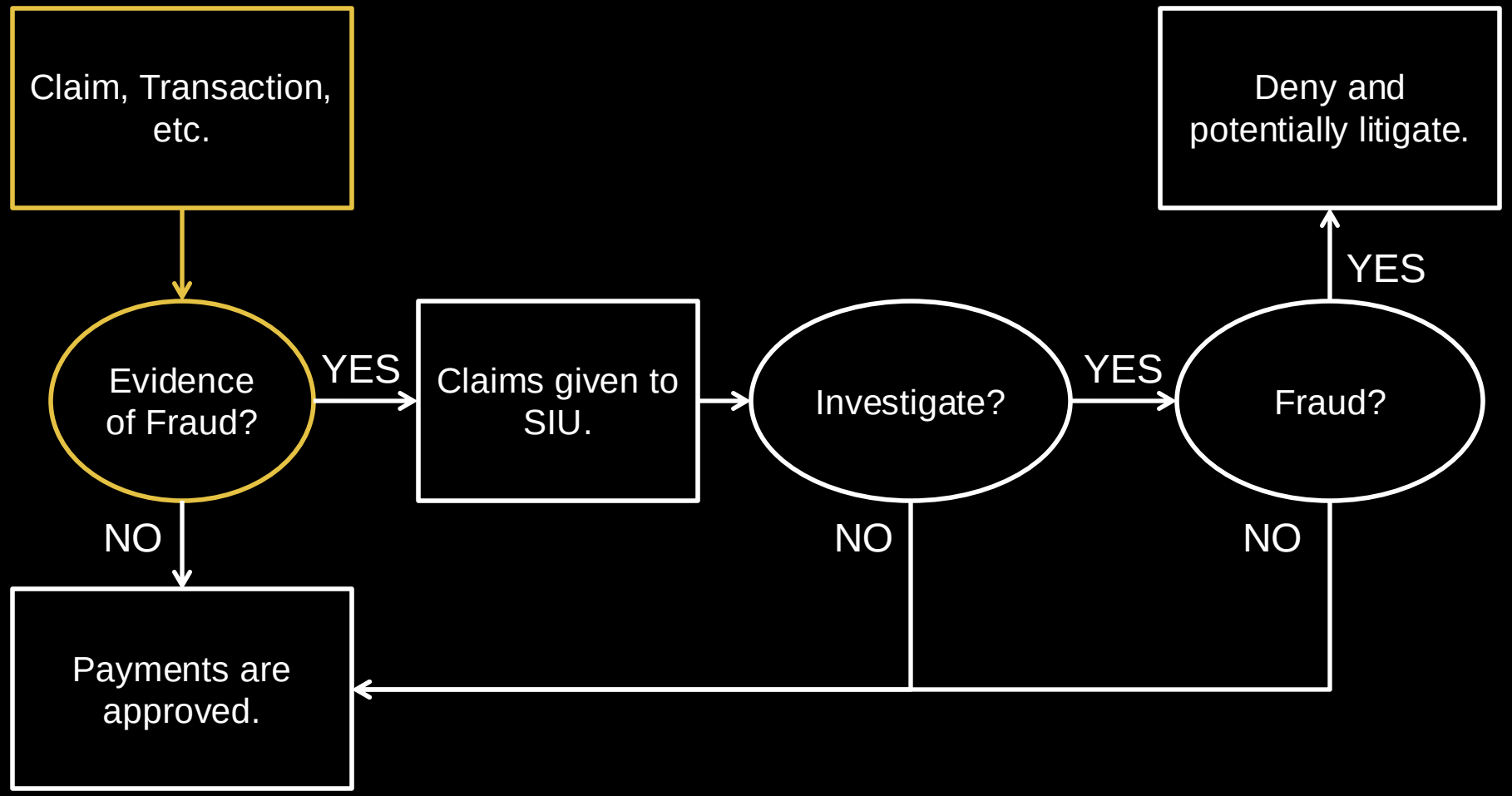
The Fraud Solution

- Regardless of the industry, two things are important for any fraud detection solution:
  1. **DETECTION** – Observing **known** fraudulent observations to determine patterns that may assist in finding other fraudulent observations.
  2. **PREVENTION** – Observing behavior and identifying suspicious actions that might be fraudulent – lead to further investigation and identification of **new** fraudulent observations.

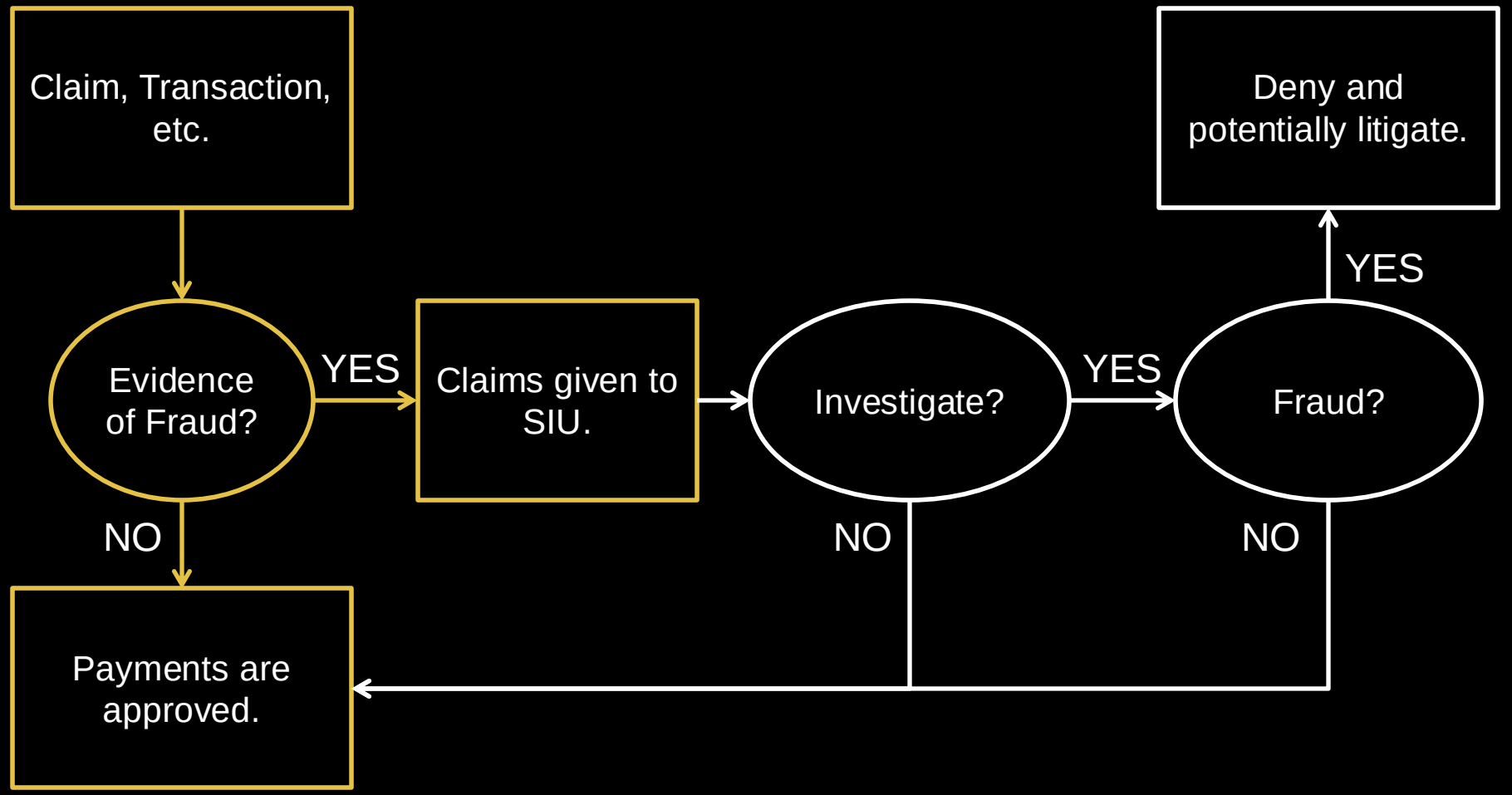
Typical Fraud Framework



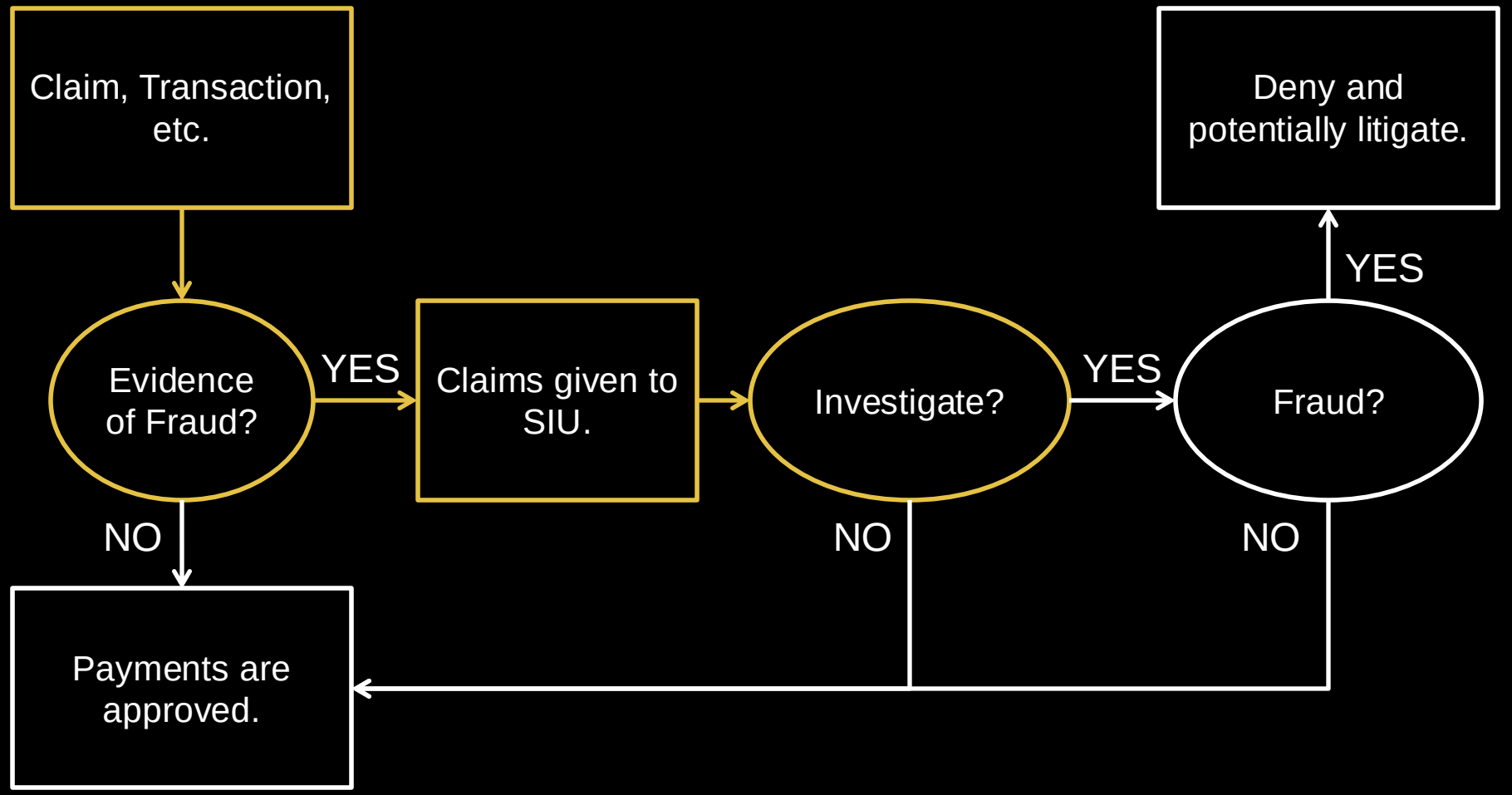
Typical Fraud Framework



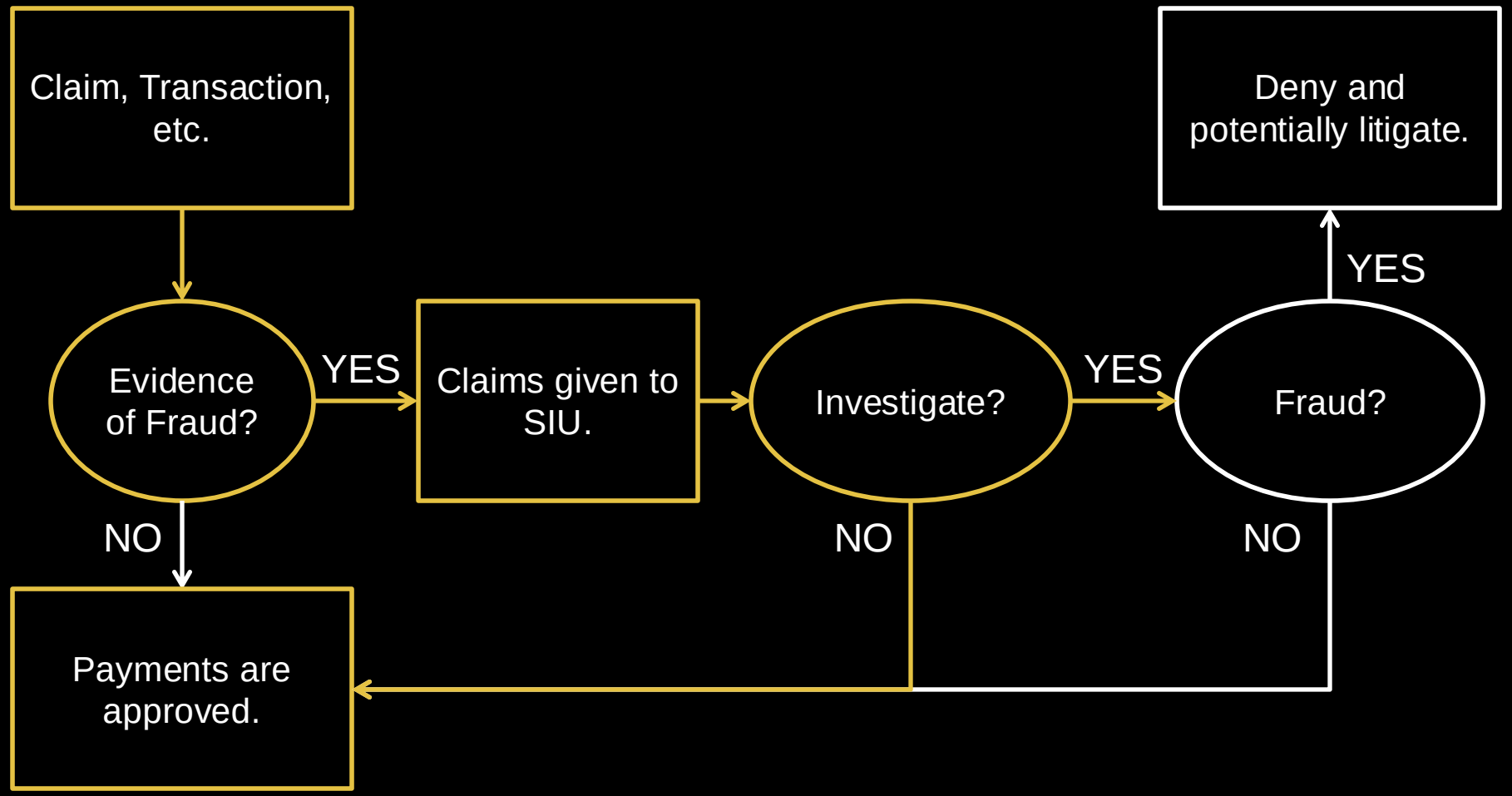
Typical Fraud Framework



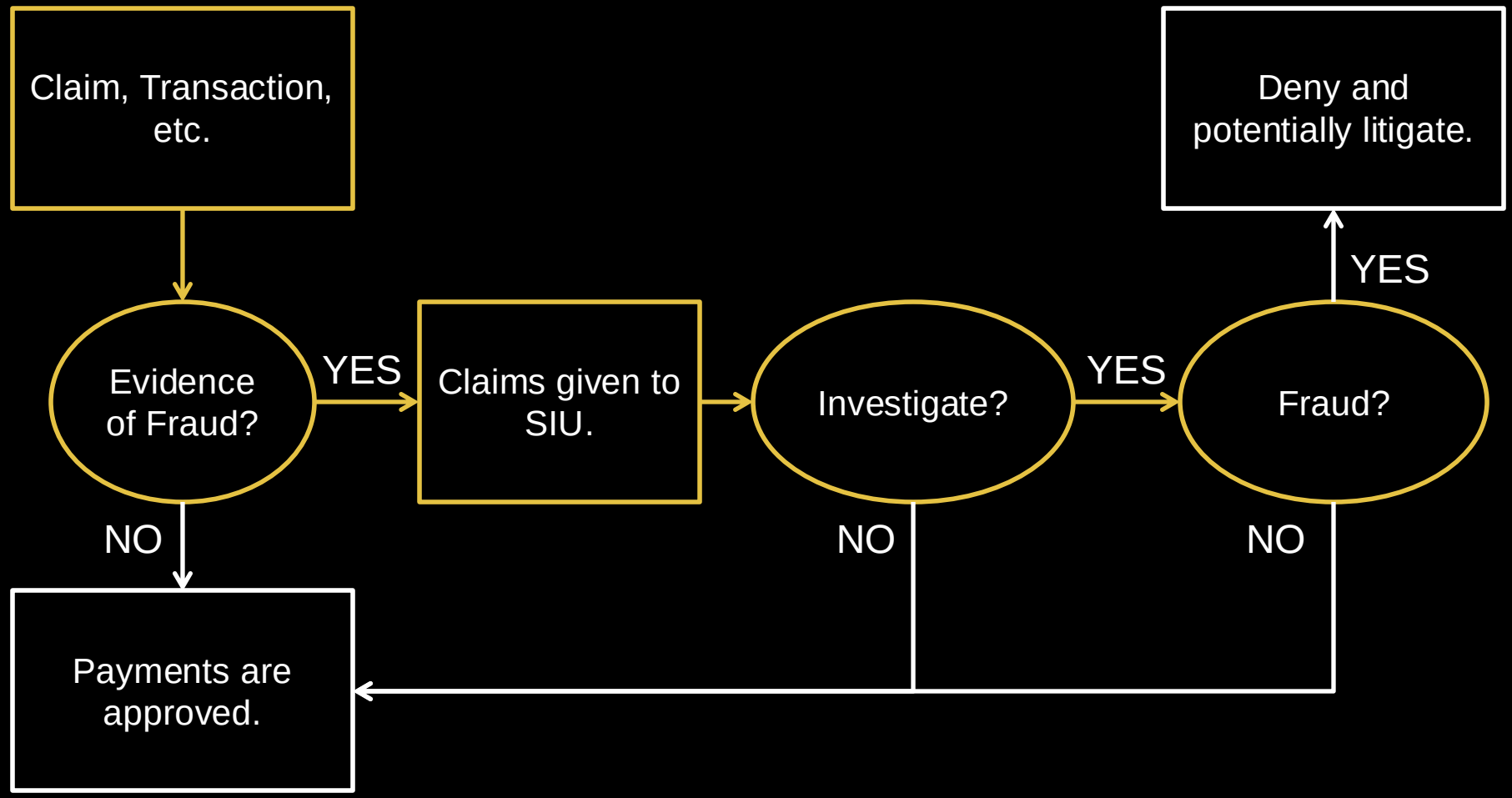
Typical Fraud Framework



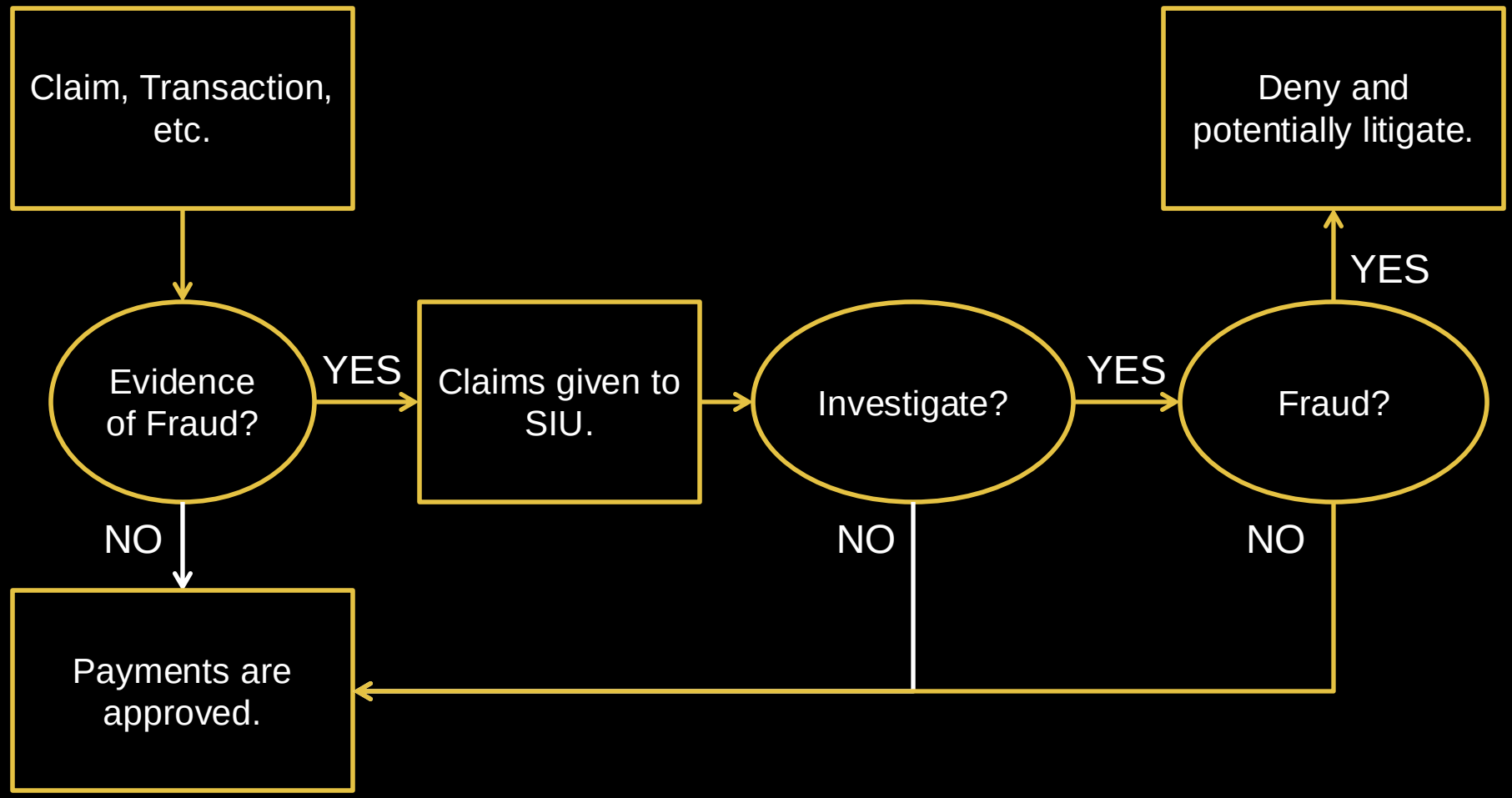
Typical Fraud Framework



Typical Fraud Framework



Typical Fraud Framework



Fraud Maturity

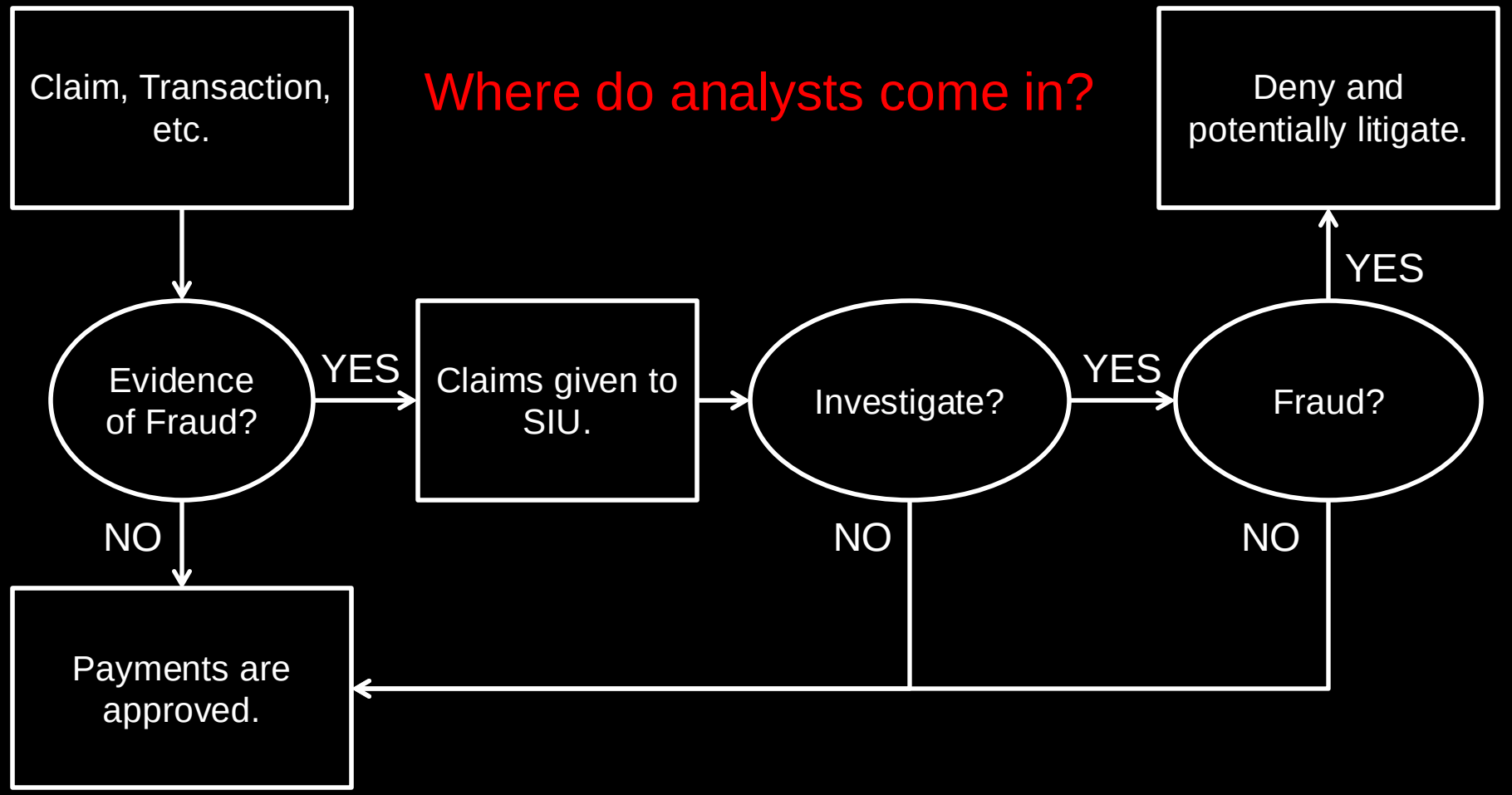
- New / young fraud analytics solutions are based on **business rules**.
- Example:
  - IF:
    - Amount of claim above threshold OR
    - Severe accident, but no police report OR
    - Severe injury, but no doctor report
  - THEN:
    - Flag as suspicious AND
    - Alert SIU

Fraud Maturity

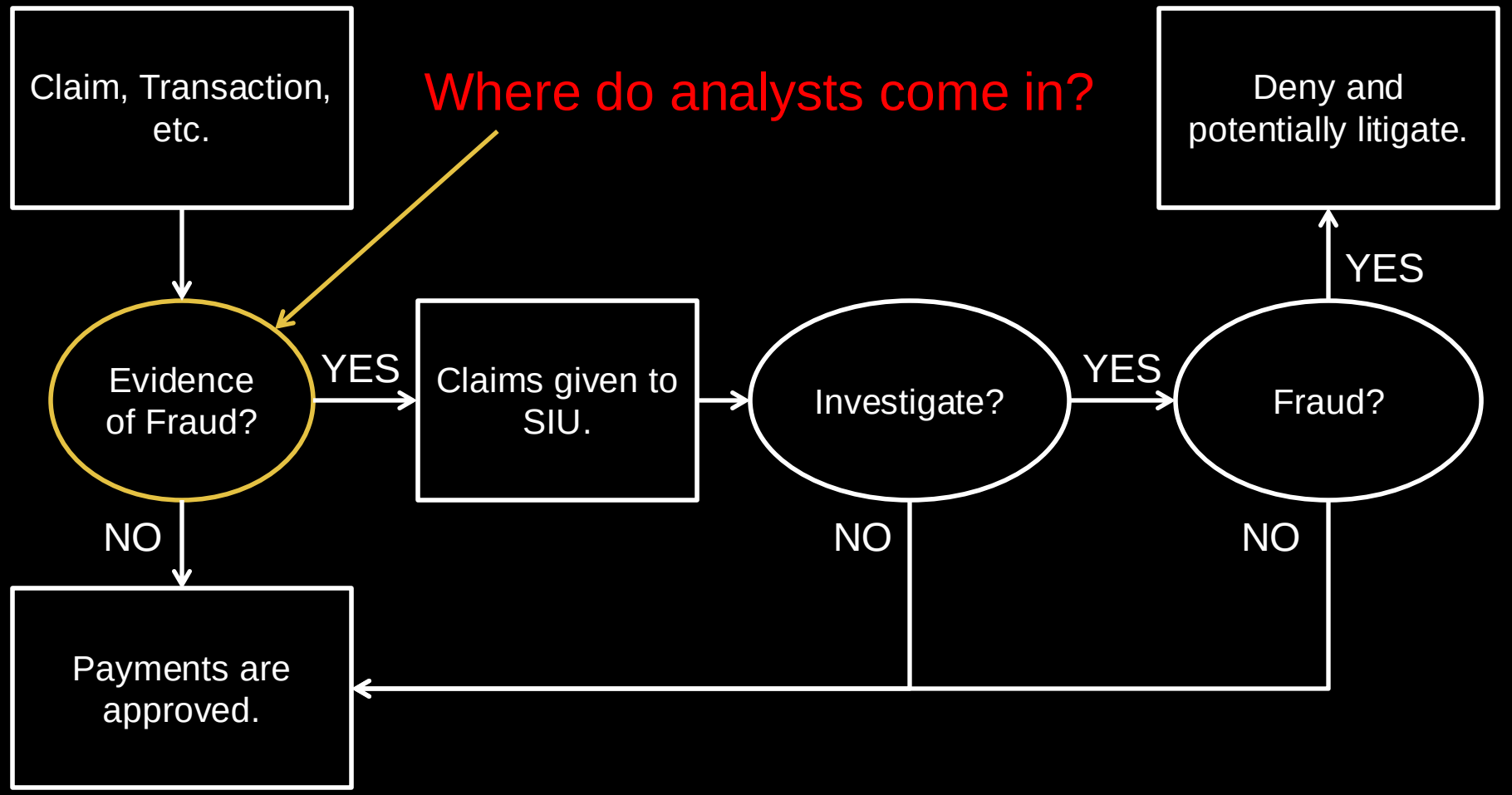
- New / young fraud analytics solutions are based on **business rules**.
- Advantages:
  - Simple
  - Easy to implement
- Disadvantages:
  - Expensive
  - Difficult to maintain and manage
  - Fraudsters discover rules
  - Completely historical

ANALYTICAL FRAUD SOLUTION

Typical Fraud Framework



Typical Fraud Framework



Advantages

1. **Precision**
   - Increased detection power
   - More information used in decision
   - More fraudsters investigated

Advantages

1. **Precision**
2. **Efficiency in Operations**
   - Automated processing of claims
   - Ranked cases for investigators

Advantages

1. **Precision**
2. **Efficiency in Operations**
3. **Efficiency in Costs**
   - Cheaper in long-run to maintain
   - Quicker identification
   - Higher investigative returns

FRAUD DATA

Fraud Data

- There are 3 common scenarios when it comes to fraud detection data sets:
  1. No previous data on fraudulent cases.

Fraud Data

- There are 3 common scenarios when it comes to fraud detection data sets:
  1. No previous data on fraudulent cases.
  2. Previous data on fraudulent cases, but can not use it.
     - Organizational structure prohibits exchange of information.
     - Retrieving data is too time consuming or expensive.
     - Fraudulent transactions can not be mapped to master database of important information.

Fraud Data

- There are 3 common scenarios when it comes to fraud detection data sets:
  1. No previous data on fraudulent cases.
  2. Previous data on fraudulent cases, but not in electronic form.
  3. Previous data on fraudulent cases that is fully integrated into company databases and structure.

Universe of Potential Fraud Cases

- Even if fraud data exists, a majority of the fraud data has a typical value of “Unknown.”
- While a claim that has never been investigated is most likely not fraud compared to fraud, it is still impossible to correctly label.

ANALYTICAL FRAUD TECHNIQUES

Fraud Maturity

Components	New / Young
Simple Rules	Yes
Unlabeled Data	Yes / No

Fraud Maturity

Components	New / Young	Emerging SIU
Simple Rules	Yes	Yes
Unlabeled Data	Yes / No	Yes / No
Labeled Fraud Cases	No	Yes
Anomaly Models	No	Yes / No

Fraud Maturity

Components	New / Young	Emerging SIU	Fraud Scoring
Simple Rules	Yes	Yes	Yes
Unlabeled Data	Yes / No	Yes / No	Yes
Labeled Fraud Cases	No	Yes	Yes
Anomaly Models	No	Yes / No	Yes
Supervised Models	No	No	Yes

Fraud Maturity

Components	New / Young	Emerging SIU	Fraud Scoring	Holistic Solution
Simple Rules	Yes	Yes	Yes	Yes
Unlabeled Data	Yes / No	Yes / No	Yes	Yes
Labeled Fraud Cases	No	Yes	Yes	Yes
Anomaly Models	No	Yes / No	Yes	Yes
Supervised Models	No	No	Yes	Yes
Non-Fraud Models	No	No	No	Yes
Clusters of not Good	No	No	No	Yes

Course Layout