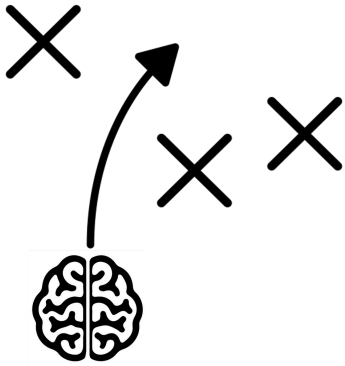
Human-Centered Investigation Playbooks

An Investigation Playbook Standard Designed for Human Analysts

v.1.1

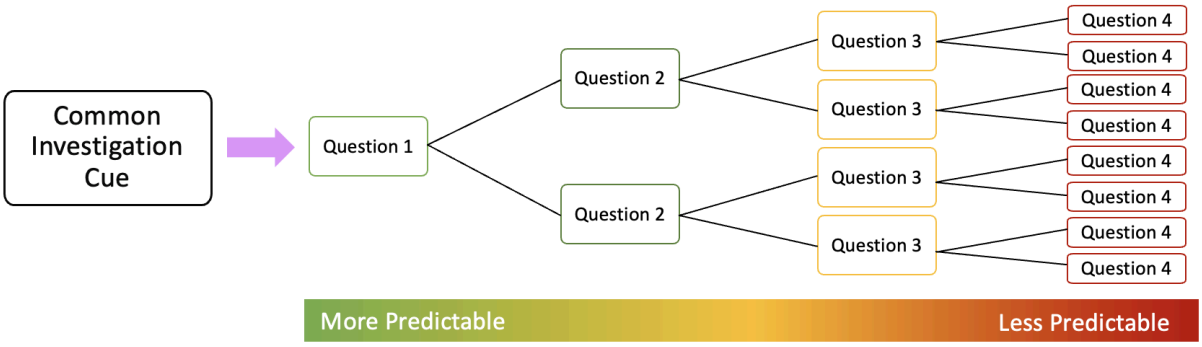
Overview

The purpose of this document is to define a standard for the expression of human-centered investigation playbooks. While other playbook standards exist, none are explicitly focused on interpretation by human analysts and integration into analyst-focused tools. This specification is based on recent research into the cognitive skills leveraged by expert analysts.



This research tells us that:

1. In any given investigation, analysts ask investigative questions that they answer with data (evidence) to determine what happened and if malicious activity occurred.
2. Analysts encounter common scenarios (cues) across diverse investigations based on the evidence they encounter and their forecasting of potentially related events.
3. Analysts and detection engineers can predict many of the initial investigative questions analysts will ask in response to these cues.
4. If you can predict the questions analysts will ask in an investigation, providing the analyst with a list of those questions when they encounter the cue has significant performance benefits.



In many investigation scenarios, analysts' initial investigative questions are predictable.

While organizations should never use playbooks to replace human analysts completely, they can augment analysts by helping them overcome the limitations of their memory, generate new ideas for specific investigation scenarios, minimize the limitations of their intuition, and collectivize individual knowledge. Furthermore, the development of investigation playbooks provides a mechanism for deliberately practicing and developing analysis skills.

Human-Centered Playbooks are flexible and straightforward. This standard enables the easy creation, modification, and sharing of playbooks by various audiences seeking to support analysts.

Goals

The goal of this project is to provide a standardized way to document investigation playbooks in a manner that is meaningful for human analysts performing cybersecurity investigations. Subgoals include:

- Expression of playbooks that are easily interpreted by human analysts
- Providing meaningful investigative steps while allowing cognitive flexibility
- Allowing for the expression of investigation steps as questions for analysts to answer
- Ability to point analysts toward evidence sources that could answer investigative questions.
- Structured in a manner that is parseable by software for integration into investigation tools.
- Linkability to related playbooks
- Linkability to publicly available detection signatures without replicating signature content
- Linkability to privately created detection signatures without exposing the signature content
- Easily created, modified, and shared

Playbook Types

All playbooks must be assigned a type based on their input. This distinction primarily serves as a mechanism to identify when someone should use the playbook. In the case of signature-linked playbooks, the type indicates that additional properties should be specified when creating a playbook of this type.

	Playbook Type	Investigation Input Types
	**Artifact**	**Input**: Encountering a suspicious artifact. **Examples**: IP Address, Domain Name, File Name, File Hash
	**Attack Technique**	**Input**: Suspicion of the use of an attack technique **Examples**: Phishing, Credential Theft, Web Shell, SQL Injection
	**Attack Phase**	**Input**: Suspicion of an attacker at an attack phase **Examples**: Persistence, Recon, Lateral Movement, Exfiltration
	**Malware Family**	**Input**: Encountering an indicator of malware family use **Examples**: Emotet, Rig EK, Cobalt Strike, Qbot, Bazar, Ryuk
	**Detection Signature Linked**	**Input**: An alert from a detection mechanism **Examples**: Suricata SID 4029184,

		YARA Rule ID 4821

Standalone Playbook Format

Analysts reference a standalone playbook when they identify a specific cue within an investigation. Those cues can be related to artifacts, techniques, attack phases, or the presence of a malware family. The playbook provides a series of investigative questions that the analysts can answer to further their identification of events on the attack timeline or their disposition.

- **Playbook Name [*name*]**: A short descriptive name for the playbook
- **Playbook ID [*id*]**: A unique identifier for the playbook. This should be a whole number greater than 0 and not within the reserved range of 1000000-1999999.
- **Playbook Description [*description*] {optional}**: A longer description of the playbook. This description can include useful investigative context for the playbook that is not captured in the other fields.
- **Playbook Type [*type*]:** The category of playbook. For standalone playbooks, this can be artifact, technique, phase, or malware.
- **Related Playbooks [*related*] {optional}**: References to other playbooks that may be useful in investigating observations commonly tied to this playbook.
- **Playbook Contributors [*contributors*] {optional}**: A list of people who contributed to the playbook, beginning with the original author.
- **Created Date [*created]:*** The date the playbook was initially created on.
- **Last Modified Date [*modified*]**: The most recent date when the playbook was added to or modified.
- **Tags [*tags*] {optional}**: Additional categorization properties.
- **References [reference] {optional}**: Links that may be helpful to the analyst while performing investigations with the playbook or that were used to inspire its creation.
- **Investigative Questions [*questions*]**: The investigative question that the play should help answer. A playbook may contain multiple questions. Each question has properties associated with it.
    - **Question [*question*]**: The investigative question written in plain language for human consumption.
        - **Context [*context*] {optional}:** A description of the question's purpose or rationale. Use this field to describe why the question is meaningful or why the analyst should care about its answer.
        - **Answering Data Sources [*answer_sources*] {optional}**: The data sources an analyst can use to answer the question. These sources can reference common values in a published or organization-specific taxonomy.
        - **Relative Time Range [*range*] {optional}**: The time range for which evidence data should be examined to answer the question. The range

should be expressed in terms relative to the observed event time, if applicable.

- **Queries [*queries*] {optional}**: Search queries analysts can use to gather evidence data to answer the question. Specify the search technology and the query.

Standalone Playbook YAML Examples

```
name: Lateral Movement Investigation
id: 91831dcb-ea8a-43b4-a732-67254f48e5d3
description: This playbook includes actions that assist in the
investigation of lateral movement. Analysts can leverage this
playbook when they suspect lateral movement may have occurred but do
not have any specific leads to follow.
type: phase
related:
  - Windows Authentication Playbook
contributors:
  - Chris Sanders
  - Josh Brower
created: 1/23/2025
modified: 1/24/2025
tags:
  - windows
  - auth
  - attack.t0008
questions:
  - question: "Were there any internal authentication attempts from
this host after the compromise occurred?"
    context: “After an attacker compromises a system, they may
attempt to use stolen credentials to authenticate to other systems.
Any authentication to another system during the compromise period
becomes suspicious, particularly if it is to a system where
authentication does not normally occur.”
    answer_sources:
        - windows_security
    range: +1day
    queries:
        - splunk: sourcetype=windows_security eventid=4624
hostname:{hostname}
        - seconion_hunt: winlog.channel:"Security" AND
event.code:"4624" AND host.name:{hostname}
  - question: "Was Psexec executed on the system?"
```

```
    context: “Attackers often use psexec to execute code remotely on
systems to facilitate lateral movement, since it works well and is
often used for legitimate purposes.”
    answer_sources:
        - windows_security
        - windows_registry
        - edr
    range: +1hr
    queries:
        -  splunk: sourcetype=windows_security eventid=4688
           process_name: psexec.exe
        - seconion_hunt: winlog.channel:"Security" AND
event.code:"4688" AND process.name:”psexec.exe”

================================================================
================================================================

name: Phishing Investigation
id: d20bfa8b-e5ae-46b5-9f90-228bdc06e862
description: This playbook includes actions that assist analysts in
determining if a user has been the victim of a phishing-based attack.
This playbook may commonly be used when an analysts has discovered a
compromised host and suspects that the initial attack vector may have
been phishing-related.
type: technique
related:
  - Message Header Analysis
contributors:
  - Chris Sanders
created: 3/18/2025
tags:
  - mail
  - initial.access
  - attack.ta0001
  - attack.t1566
questions:
  - question: "Did the user receive any messages with suspicious
subject lines?"
    context: “A suspicious subject line may be overly generic,
references a request for information, seems irrelevant to the users
job role, or appears to offer any sort of deal or surprising benefit
for the recipient.”
    answer_sources:
        - mail_tx
    range: -3day
```

```
  - question: "Did the user receive any messages with suspicious
links?"
    context: “A suspicious link may be one that references a domain
that you have never heard of, appears algorithmically generated,
appears to be mimicking a legitimate domain, or is tied to an obscure
top level domain. The link may also go directly to an IP address
rather than a domain.”
    answer_sources:
        - mail_tx
    range: +3day
  - question: "Did the user receive any messages with suspicious
attachments?"
    context: “A suspicious attachment may be one that has a name that
is overly generic, appears to request information from the user, or
offers them something valuable. Suspicious attachments may also be of
file types that are commonly used for code execution or redirection
by attackers, like executable files, office documents, or PDFs. They
may also attempt to hide their file type by using file extensions
that don't match their content or multiple file extensions like
.exe.pdf.”
    answer_sources:
        - mail_tx
    range: +3day
  - question: "Did the user receive any messages from accounts they
have never received messages from before?"
    context: “Phishing messages, unless targeted and spoofed, are
more likely to come from accounts that a user has never sent or
received mail to/from.”
    answer_sources:
        - mail_tx
    range: before
  - question: "Did the user visit any links that were from recently
received messages?"
    context: “By reviewing visits to links received from recent
messages, you may identify malicious activity that was not otherwise
obvious from the origin email or link themselves.”
    answer_sources:
        - mail_tx
        - flow
        - pcap
        - http_proxy
    range: -3day+1day
 - question: "Did the user receive any messages from sender IP
addresses that appear on public blocklists?"
```

```
    context: "Messages received from servers listed on public block
lists are more likely to be spam or associated with potential
malicious activity."
    answer_sources:
        - mail_tx
        - mail_message_headers
        - reputation
    range: all
```

Detection Signature-Linked Playbook Format

A signature-linked playbook is associated with a specific detection mechanism signature or capability. These may be assigned to privately created signatures (whether for specific organizations or proprietary vendor technology) or publicly available signatures (like Suricata, Sigma, or YARA).

- **Playbook Name [*name*]**: A short descriptive name for the playbook. For detection signature-linked playbooks, this can be the name of the alert/signature taken from the source and may include detection platform details.
- **Playbook ID [*id*]**: A unique identifier for the playbook. This should be a whole number greater than 0. The range 1000000-1999999 is reserved.
- **Detection ID [*detection_id*]**: Used only for detection-linked playbooks. Contains the unique identifier of the source detection signature (ex. Suricata SID or Sigma ID).
- **Playbook Description [*description*] {optional}**: A longer description of the playbook. This description can include useful investigative context for the playbook that is not captured in the other components of the playbook. This description may be blank for detection signature-linked playbooks, which can rely on the description included in the detection rule.
- **Playbook Type [*type*]**: The playbook category. This field will always be 'detection' for playbooks linked to detection rules.
- **Detection Type [*detection_type*]**: The tools responsible for generating the alert, such as Suricata, ESET Endpoint Security, or Cisco ESA. You can use a specific detection/alerting technology or a generic tool class like siem, nids, hids, or edr. This field is unique for playbooks of the alert type.
- **Related Playbooks [*related*] {optional}**: References to other playbooks that may be useful in investigating this alert.
- **Playbook Contributors [*contributors*] {optional}**: A list of people who contributed to the playbook, including the original author.
- **Created Date [*created]:*** The date the playbook was initially created on.

- **Last Modified Date [*modified*]**: The most recent date when the playbook was added to or modified.
- **Tags [*tags*] {optional}**: Additional categorization properties.
- **References [reference] {optional}**: Links that may be helpful to the analyst while performing investigations with the playbook or were used during its creation.
- **Investigative Questions [*questions*]**: The investigative question that the play should help answer. A playbook may contain multiple questions. Each question has properties associated with it.
    - **Question [*question*]**: The investigative question written in plain language for human consumption, in the form of a question.
        - **Context [*context*] {optional}:** A description of the question's purpose or rationale. Use this field to describe why the question is meaningful or why the analyst should care about its answer.
        - **Answering Data Sources [*answer_sources*] {optional}**: The data sources an analyst can use to answer the question. These sources can reference common values in a published or organization-specific taxonomy.
        - **Relative Time Range [*range*] {optional}**: The time range for which evidence data should be examined to answer the question. The range should be expressed in terms relative to the alert time.
        - **Queries [*queries*] {optional}**: Search queries analysts can use to gather evidence data to answer the question. Specify the search technology and the query.
            - **Aggregation [*boolean*]**: Specifies whether the query is an aggregation (grouping) of results.

Detection Signature-Linked YAML Examples

```
================================================================
================================================================

name: "Whoami Execution"
id: e28a5a99-da44-436d-b7a0-2afc20a5f413
description: Detects the execution of whoami, which is often used by
attackers after exploitation / privilege escalation but rarely used
by administrators.
type: detection
mechanism: edr
contributors:
  - Josh Brower
created: 1/23/2025
modified: 1/24/2025
questions:
  - question: "What user ran it, on what system and what is the
parent process? Are all of these expected for your environment?"
    context: “Since whoami is not an application typically run by
normal users, the information collected here can help determine if
this is typical behavior for the user and system.”
    answer_sources:
        - windows_security
        - interview
    queries:
        seconion_hunt: query

  - question: "What other processes did the parent process create?"
    context: “Identifying other related processes can help you
contextualize the disposition of the action and understand its role
in potentially broader activity.”
    answer_sources:
         - process_auditing
         - windows_security
    range: -1hr+1hr
    queries:
        seconion_hunt: "event.dataset: "process_creation" AND
process.ppid: "{PPID}" AND host.hostname: "{Hostname}" | groupby
"process.executable" "process.command_line"

================================================================
```

```
================================================================

name: AWS STS AssumeRole Misuse
id: 905d389b-b853-46d0-9d3d-dea0d3a3cd49
description: Identifies the suspicious use of the AWS AssumeRole
action. This is a common activity performed by developers, admins,
and CI/CD systems whose authorization should be documented, but
attackers could use AssumeRole to move laterally to roles with
elevated privileges using stolen credentials.
type: detection
mechanism: siem
related:
  - AWS IAM Backdoor Users Keys
  - AWS STS GetSessionToken Misuse
  - AWS Suspicious SAML Activity
contributors:
  - Alek Rollyson
created: 1/23/2025
modified: 1/24/2025
questions:
  - question: "Who normally assumes this role, if anyone?"
    context: “This helps to set a baseline for normal use of this
action.”
    answer_sources:
        - cloudtrail
    range: -1mo
    queries:
        splunk: sourcetype=aws:cloudtrail
userIdentity.type=AssumedRole
userIdentity.sessionContext.sessionIssuer.type=Role
responseElements.assumedRoleUser.arn=<{ARN}> | stats count by
userIdentity.username
  - question: "Does this user normally perform role assumptions and,
if so, what roles do they normally assume?"
    context: “This question facilitates a baseline comparison for the
user who conducted this action. Deviations from normal behavior may
indicate a malicious action.”
    answer_sources:
        - cloudtrail
    range: -1mo
    queries:
        splunk: sourcetype=aws:cloudtrail
userIdentity.type=AssumedRole
userIdentity.sessionContext.sessionIssuer.type=Role
```

```
userIdentity.username=<{username}> | stats count by
responseElements.assumedRoleUser.arn
  - question: "Is this a location and useragent this user normally
makes API calls from?"
    context: “If the location and useragent are different from
baseline, it may indicate a malicious disposition.”
    answer_sources:
        - cloudtrail
    range: -1mo
    queries:
        splunk: sourcetype=aws:cloudtrail eventType=AwsApiCall
userIdentity.username=<{username}> | stats count by sourceIPAddress,
useragent
  - question: "Is this user part of authorized development groups?"
    context: “This type of activity is normal for development group
users. If the user is not part of this group, the action may indicate
a malicious action.”
    answer_sources:
        - active_directory
        - internal_docs
    queries:
        ldapsearch: ldapsearch -x sAMAccountName={username} memberOf
  - question: "What API calls did this role make after it was
assumed?"
    context: “Analyzing additional API calls made from this user
after this action should provide more information about the user’s
intentions and whether they were malicious in disposition.”
    answer_sources:
        - cloudtrail
    range: +6hr
    queries:
        splunk: sourcetype=aws:cloudtrail userIdentity.arn=<{ARN}> |
stats count by eventName
```

Additional Field Specifications

Relative Time Ranges [*range*]

Analysts typically narrow searches with time ranges relative to events they are interested in. Human-centered playbooks allow for specifying relative time ranges recommended for answering investigative questions. These time ranges are anchored to whatever event led the analyst to reference the playbook (which may include an alert for signature-linked playbooks).

Relative time ranges can be expressed as time before (-) OR after (+) the anchoring event. You should include a numeric value and time unit.

- +1hr: Within an hour after the event
- -1day: Within a day before the event
- -100ms: Within 100 milliseconds before the event

Relative time ranges can be expressed as time before AND after an event by combining keywords. The before (-) time unit should appear first.

- -10min+10min: Within 10 minutes before or after the event
- -1yr+1day: Within 1 year before the event and 1 day after the event

You may also designate all time, all time before an event, or all time after an event by using the all, before, and after time units, respectively.

The following time units are supported:

Time Unit	Description
all	All Time - Before or After
before	All Time - Before
after	All Time - After
yr	Years
mo	Months
wk	Weeks
day	Days
hr	Hours
min	Minutes
sec	Seconds

ms	Milliseconds

License