When Public Data Becomes a Targeting Pipeline: A Guide to Automated OSINT

Maria Cattini | 17/09/2026 | OSINT

The most important change brought by AI to OSINT may not be better search. It may be the ability to turn a sequence of investigative tasks into a continuous, automated pipeline.

A researcher can already search public profiles, identify organisations, retrieve professional roles, compare accounts, map relationships and summarise findings. What changes when those operations are chained together and executed automatically across hundreds of people?

The distinction matters because the underlying information may remain public while the nature of the operation changes substantially. The material examined by ProjectOSINT describes an Iranian-nexus actor that, according to Anthropic, used Claude as part of an automated system to profile hundreds of people in Israel, the United States and the Jewish diaspora from an existing target list. The system collected, correlated and synthesised publicly available information. Anthropic also reports broader cases in which AI systems were used to orchestrate reconnaissance, exploitation and data exfiltration, while humans retained decisions such as target selection and use of the resulting intelligence.

For an OSINT practitioner, the useful question is therefore not simply whether AI can automate research. It is how to recognise what an automated workflow is actually doing.

Start with the workflow, not the AI

A common mistake is to focus on the model: which LLM was used, how capable it was, whether it operated autonomously.

For investigative analysis, start somewhere else. Reconstruct the pipeline.

A useful first model is:

Collection → Enrichment → Correlation → Profiling → Targeting

These stages should not be treated as rigid technical categories. They are an analytical framework for understanding how an operation changes as information moves through it.

The crucial point is that automation can connect stages that, when performed manually, would normally involve separate searches, decisions and checks.

1. Collection: what information is being gathered?

Collection is the familiar starting point of OSINT: retrieving information from publicly accessible sources.

For a person, this might include professional profiles, institutional biographies, public social accounts, company records, conference appearances or other openly available material.

At this stage, ask basic questions:

What is the starting identifier?

A name, username, organisation, email address, domain or another entity may initiate the search.

Which sources are queried?

Do not treat “the web” as one source. Record the platforms, registries, websites, archives or datasets involved.

Is collection targeted or exploratory?

Searching for information about one entity is operationally different from feeding hundreds of predetermined names into a system.

How much human intervention is required?

A researcher performing individual searches is different from a workflow that accepts a list and processes every entry automatically.

This last distinction is increasingly important. Automation changes the economics of collection even when it does not change the underlying accessibility of the information.

2. Enrichment: what does the system add to the starting data?

A name alone may reveal little. An enrichment process attempts to attach additional attributes to it.

Suppose a workflow begins with:

Person A

It may subsequently identify a professional position, employer, public accounts, organisational affiliations, conference appearances and other associated entities.

The methodological question is no longer simply “Did the system find information?” It becomes:

How did the system determine that the information belongs to the same entity?

This is where automated OSINT can produce convincing but incorrect results.

Two people can have the same name. An old professional biography may no longer be current. A social account may belong to someone else. A company record may concern another legal entity with a similar name.

Every enrichment step therefore creates an attribution problem.

A useful rule is:

Discovery is not identity resolution.

If an automated system finds an apparently relevant account, document or organisation, the association should remain provisional until corroborating evidence supports it.

3. Correlation: when separate facts become intelligence

Correlation is where public information begins to acquire substantially greater analytical value.

Imagine that separate sources establish that a person has a particular professional role, works for an organisation, has participated in a public event and appears in material associated with other individuals.

None of those facts necessarily says much in isolation.

Connecting them can produce a map of relationships.

This is why the distinction between public and private information is insufficient for evaluating an OSINT workflow. The relevant capability may lie not in discovering a hidden fact but in connecting numerous visible ones.

For practitioners, this creates a verification requirement.

For every important relationship generated automatically, ask:

What is the evidence for the edge?

In a relationship graph, two nodes connected by a line can look authoritative. But the visual connection is only as strong as the evidence behind it.

Did two people work for the same organisation at the same time?

Did they merely appear on the same website?

Does one follow the other on a social platform?

Were they photographed at the same event?

Are they directors of the same company?

Those relationships have different evidentiary meanings. A system that converts all of them into identical graph edges hides precisely the distinction an analyst needs to preserve.

4. Profiling: recognise the change in analytical purpose

The next transition is more consequential.

Collection asks:

What can I find?

Profiling asks:

What can these combined observations tell me about this person?

A structured profile can combine identity, role, affiliations, relationships and activity into a single analytical object.

This is where scale becomes especially important.

Producing one profile manually imposes costs: researcher time, searches, comparison, note-taking and synthesis. A pipeline capable of processing hundreds of predetermined individuals changes that constraint.

The source material describes exactly this concern: AI-assisted workflows can reduce the human effort required to collect, correlate and synthesise information about large numbers of people.

For an OSINT practitioner assessing such a system, four questions are useful:

Is the subject already predetermined?
There is a meaningful difference between investigating an event and receiving a list of people to profile.

Is the output structured around individuals?
A collection of search results is not the same thing as a dossier.

Does the system infer relationships or characteristics?
Inference introduces another layer of uncertainty beyond retrieved facts.

Can the process operate at scale without equivalent growth in human review?
If 500 profiles can be generated with roughly the same human oversight previously required for a handful, automation has altered the nature of the capability.

5. Targeting: look at what happens after the profile

The final stage cannot be understood by examining data collection alone.

Ask what the output is designed to enable.

A profile created for historical research, journalistic verification or defensive threat analysis may use some of the same public sources as a profile created to support surveillance or an operation against a person.

The data alone does not tell you the purpose.

This means an OSINT assessment should document not only inputs and processing but, where evidence permits, the intended use of outputs.

A practical investigation can therefore be represented as:

Input → Sources → Enrichment → Correlation → Profile → Decision or action

The final arrow deserves as much scrutiny as the first.

If evidence for that final stage is absent, say so. Do not infer targeting merely because profiling exists.

How to audit an automated OSINT pipeline

When encountering an AI-based investigative system, do not begin by asking whether it is “autonomous”. That label compresses too many different behaviours into one word.

Instead, reconstruct what humans and machines actually do.

Create a simple workflow table:

Stage	Key question	Main verification risk
Collection	What sources are queried?	Incomplete or unreliable sources
Enrichment	How are attributes attached to entities?	False identity matches
Correlation	How are relationships established?	Unsupported associations
Profiling	What conclusions are generated?	Facts and inference merged
Targeting	How are profiles used?	Purpose inferred without evidence

Then identify the human checkpoints.

Who chooses the subjects?

Who approves sources?

Who resolves ambiguous identities?

Who validates relationships?

Who reviews generated profiles?

Who decides what happens next?

This produces a much more useful picture than a percentage claiming that an operation is “90% autonomous” or “mostly AI-driven”. The meaningful issue is **where human judgement remains in the pipeline and where it has been delegated**.

Three checks before trusting automated profiling

The first is **provenance**.

Every significant assertion should remain traceable to its source. If a generated dossier states that a person works for an organisation, the analyst should be able to identify the evidence supporting that statement.

The second is **entity resolution**.

Automated systems must not silently collapse similar names, accounts or organisations into a single identity. High-confidence matching requires corroborating attributes rather than superficial similarity.

The third is **fact–inference separation**.

“Person A is listed as an employee of Organisation B” and “Person A is influential within Organisation B” are different types of statements. One may be directly evidenced; the other may be an analytical judgement.

An effective pipeline should preserve that distinction instead of presenting both as equivalent facts.

Scale is itself an analytical variable

Traditional OSINT discussions often concentrate on access: can this information legally and technically be obtained from an open source?

AI makes another variable harder to ignore: **scale**.

An operation that would require a researcher to spend hours on each subject may become capable of processing hundreds of subjects through the same sequence of searches and correlations.

The source material frames this as an increase in scale available to each human operator rather than simply an increase in machine “intelligence”.

That distinction is useful beyond the Anthropic cases.

When evaluating any automated OSINT platform, measure at least three forms of scale:

Subject scale: how many entities can be investigated?

Source scale: how many sources can be queried and correlated?

Operational scale: how many investigations can run concurrently?

A fourth metric should then be placed beside them:

Verification capacity: how much meaningful human checking can actually be performed?

If the first three grow dramatically while the fourth remains almost unchanged, the system may generate more intelligence products while simultaneously increasing the opportunity for attribution errors, false correlations and unsupported inference.

Public does not mean context-free

There is a broader lesson here for OSINT.

The public availability of an individual piece of information tells us something about access. It does not, by itself, resolve questions about aggregation, profiling, purpose or risk.

A professional biography may have been published so that someone can understand a person's role. A conference programme exists to identify speakers. A corporate registry serves a transparency function.

An automated system can extract those separate fragments from their original contexts and combine them into something none of the individual sources contained: a structured representation of a person and their relationships.

For investigators, this should change the way automated OSINT systems are evaluated.

Do not ask only:

Can the system find this information?

Ask:

What does it collect, how does it connect the evidence, what does it infer, how many people can it process, where does human verification occur, and what is the resulting profile used for?

That is the practical threshold worth watching.

AI does not need to discover secret information to change intelligence work. Automating the path from public data to structured profiles can be consequential enough.

The challenge for OSINT is therefore not simply learning how to automate more research. It is learning how to preserve provenance, identity resolution, evidentiary distinctions and human judgement when automation makes research possible at a scale that manual workflows rarely allowed.
The most important change brought by AI to OSINT may not be better search. It may be the ability to turn a sequence of investigative tasks into a continuous, automated pipeline.

A researcher can already search public profiles, identify organisations, retrieve professional roles, compare accounts, map relationships and summarise findings. What changes when those operations are chained together and executed automatically across hundreds of people?

The distinction matters because the underlying information may remain public while the nature of the operation changes substantially. The material examined by ProjectOSINT describes an Iranian-nexus actor that, according to Anthropic, used Claude as part of an automated system to profile hundreds of people in Israel, the United States and the Jewish diaspora from an existing target list. The system collected, correlated and synthesised publicly available information. Anthropic also reports broader cases in which AI systems were used to orchestrate reconnaissance, exploitation and

data exfiltration, while humans retained decisions such as target selection and use of the resulting intelligence.

For an OSINT practitioner, the useful question is therefore not simply whether AI can automate research. It is how to recognise what an automated workflow is actually doing.

Start with the workflow, not the AI

A common mistake is to focus on the model: which LLM was used, how capable it was, whether it operated autonomously.

For investigative analysis, start somewhere else. Reconstruct the pipeline.

A useful first model is:

Collection → Enrichment → Correlation → Profiling → Targeting

These stages should not be treated as rigid technical categories. They are an analytical framework for understanding how an operation changes as information moves through it.

The crucial point is that automation can connect stages that, when performed manually, would normally involve separate searches, decisions and checks.

1. Collection: what information is being gathered?

Collection is the familiar starting point of OSINT: retrieving information from publicly accessible sources.

For a person, this might include professional profiles, institutional biographies, public social accounts, company records, conference appearances or other openly available material.

At this stage, ask basic questions:

What is the starting identifier?

A name, username, organisation, email address, domain or another entity may initiate the search.

Which sources are queried?

Do not treat “the web” as one source. Record the platforms, registries, websites, archives or datasets involved.

Is collection targeted or exploratory?

Searching for information about one entity is operationally different from feeding hundreds of predetermined names into a system.

How much human intervention is required?

A researcher performing individual searches is different from a workflow that accepts a list and processes every entry automatically.

This last distinction is increasingly important. Automation changes the economics of collection even when it does not change the underlying accessibility of the information.

2. Enrichment: what does the system add to the starting data?

A name alone may reveal little. An enrichment process attempts to attach additional attributes to it.

Suppose a workflow begins with:

Person A

It may subsequently identify a professional position, employer, public accounts, organisational affiliations, conference appearances and other associated entities.

The methodological question is no longer simply “Did the system find information?” It becomes:

How did the system determine that the information belongs to the same entity?

This is where automated OSINT can produce convincing but incorrect results.

Two people can have the same name. An old professional biography may no longer be current. A social account may belong to someone else. A company record may concern another legal entity with a similar name.

Every enrichment step therefore creates an attribution problem.

A useful rule is:

Discovery is not identity resolution.

If an automated system finds an apparently relevant account, document or organisation, the association should remain provisional until corroborating evidence supports it.

3. Correlation: when separate facts become intelligence

Correlation is where public information begins to acquire substantially greater analytical value.

Imagine that separate sources establish that a person has a particular professional role, works for an organisation, has participated in a public event and appears in material associated with other individuals.

None of those facts necessarily says much in isolation.

Connecting them can produce a map of relationships.

This is why the distinction between public and private information is insufficient for evaluating an OSINT workflow. The relevant capability may lie not in discovering a hidden fact but in connecting numerous visible ones.

For practitioners, this creates a verification requirement.

For every important relationship generated automatically, ask:

What is the evidence for the edge?

In a relationship graph, two nodes connected by a line can look authoritative. But the visual connection is only as strong as the evidence behind it.

Did two people work for the same organisation at the same time?

Did they merely appear on the same website?

Does one follow the other on a social platform?

Were they photographed at the same event?

Are they directors of the same company?

Those relationships have different evidentiary meanings. A system that converts all of them into identical graph edges hides precisely the distinction an analyst needs to preserve.

4. Profiling: recognise the change in analytical purpose

The next transition is more consequential.

Collection asks:

What can I find?

Profiling asks:

What can these combined observations tell me about this person?

A structured profile can combine identity, role, affiliations, relationships and activity into a single analytical object.

This is where scale becomes especially important.

Producing one profile manually imposes costs: researcher time, searches, comparison, note-taking and synthesis. A pipeline capable of processing hundreds of predetermined individuals changes that constraint.

The source material describes exactly this concern: AI-assisted workflows can reduce the human effort required to collect, correlate and synthesise information about large numbers of people.

For an OSINT practitioner assessing such a system, four questions are useful:

Is the subject already predetermined?
There is a meaningful difference between investigating an event and receiving a list of people to profile.

Is the output structured around individuals?
A collection of search results is not the same thing as a dossier.

Does the system infer relationships or characteristics?
Inference introduces another layer of uncertainty beyond retrieved facts.

Can the process operate at scale without equivalent growth in human review?
If 500 profiles can be generated with roughly the same human oversight previously required for a handful, automation has altered the nature of the capability.

5. Targeting: look at what happens after the profile

The final stage cannot be understood by examining data collection alone.

Ask what the output is designed to enable.

A profile created for historical research, journalistic verification or defensive threat analysis may use some of the same public sources as a profile created to support surveillance or an operation against a person.

The data alone does not tell you the purpose.

This means an OSINT assessment should document not only inputs and processing but, where evidence permits, the intended use of outputs.

A practical investigation can therefore be represented as:

Input → Sources → Enrichment → Correlation → Profile → Decision or action

The final arrow deserves as much scrutiny as the first.

If evidence for that final stage is absent, say so. Do not infer targeting merely because profiling exists.

How to audit an automated OSINT pipeline

When encountering an AI-based investigative system, do not begin by asking whether it is “autonomous”. That label compresses too many different behaviours into one word.

Instead, reconstruct what humans and machines actually do.

Create a simple workflow table:

Stage	Key question	Main verification risk
Collection	What sources are queried?	Incomplete or unreliable sources
Enrichment	How are attributes attached to entities?	False identity matches
Correlation	How are relationships established?	Unsupported associations
Profiling	What conclusions are generated?	Facts and inference merged
Targeting	How are profiles used?	Purpose inferred without evidence

Then identify the human checkpoints.

Who chooses the subjects?

Who approves sources?

Who resolves ambiguous identities?

Who validates relationships?

Who reviews generated profiles?

Who decides what happens next?

This produces a much more useful picture than a percentage claiming that an operation is “90% autonomous” or “mostly AI-driven”. The meaningful issue is **where human judgement remains in the pipeline and where it has been delegated**.

Three checks before trusting automated profiling

The first is **provenance**.

Every significant assertion should remain traceable to its source. If a generated dossier states that a person works for an organisation, the analyst should be able to identify the evidence supporting that statement.

The second is **entity resolution**.

Automated systems must not silently collapse similar names, accounts or organisations into a single identity. High-confidence matching requires corroborating attributes rather than superficial

similarity.

The third is **fact–inference separation**.

“Person A is listed as an employee of Organisation B” and “Person A is influential within Organisation B” are different types of statements. One may be directly evidenced; the other may be an analytical judgement.

An effective pipeline should preserve that distinction instead of presenting both as equivalent facts.

Scale is itself an analytical variable

Traditional OSINT discussions often concentrate on access: can this information legally and technically be obtained from an open source?

AI makes another variable harder to ignore: **scale**.

An operation that would require a researcher to spend hours on each subject may become capable of processing hundreds of subjects through the same sequence of searches and correlations.

The source material frames this as an increase in scale available to each human operator rather than simply an increase in machine “intelligence”.

That distinction is useful beyond the Anthropic cases.

When evaluating any automated OSINT platform, measure at least three forms of scale:

Subject scale: how many entities can be investigated?

Source scale: how many sources can be queried and correlated?

Operational scale: how many investigations can run concurrently?

A fourth metric should then be placed beside them:

Verification capacity: how much meaningful human checking can actually be performed?

If the first three grow dramatically while the fourth remains almost unchanged, the system may generate more intelligence products while simultaneously increasing the opportunity for attribution errors, false correlations and unsupported inference.

Public does not mean context-free

There is a broader lesson here for OSINT.

The public availability of an individual piece of information tells us something about access. It does not, by itself, resolve questions about aggregation, profiling, purpose or risk.

A professional biography may have been published so that someone can understand a person's role. A conference programme exists to identify speakers. A corporate registry serves a transparency function.

An automated system can extract those separate fragments from their original contexts and combine them into something none of the individual sources contained: a structured representation of a person and their relationships.

For investigators, this should change the way automated OSINT systems are evaluated.

Do not ask only:

Can the system find this information?

Ask:

What does it collect, how does it connect the evidence, what does it infer, how many people can it process, where does human verification occur, and what is the resulting profile used for?

That is the practical threshold worth watching.

AI does not need to discover secret information to change intelligence work. Automating the path from public data to structured profiles can be consequential enough.

The challenge for OSINT is therefore not simply learning how to automate more research. It is learning how to preserve provenance, identity resolution, evidentiary distinctions and human judgement when automation makes research possible at a scale that manual workflows rarely allowed.