VIKRAM JHA - DEFENSIBLE AI
REFERENCE TAXONOMY

AGENT ACTION-CLASS TAXONOMY

Version 1.0 - 24 July 2026
https://vikramjha.work/artifacts/agent-action-class-taxonomy

Ten classes of thing an agent can do, and the obligation each one drags in the moment it does it.

WHY IT EXISTS

Almost every governance argument about agents is conducted at the wrong altitude. Teams debate the model - its accuracy, its provider, whether it hallucinates - while the obligations attach somewhere else entirely: to the action. A regulator does not ask which model declined the claim. It asks who declined the claim, on whose authority, and whether the person affected can find out why. Sorting an estate by action class rather than by model is usually the single change that makes an agent inventory governable, because it produces a list whose rows map one-to-one onto duties that already exist.

WHAT IT CONTAINS

1. Ten action classes, ordered by how much of the organization's authority each one spends
2. For each: what actually changes at the moment an agent takes it rather than a person
3. The obligation family it triggers, with the regime that usually carries it
4. The evidence a reviewer asks for first - and the artifact that closes it
5. The boundary test that tells you which class a given capability really belongs to

THE TEN ACTION CLASSES

Ordered by escalating authority. An agent almost never sits in one class: classify per capability, not per agent, and inventory the highest class it can reach without a human in the path.

01 RETRIEVE

What it is: Reads and summarizes. No writes, no assertions to a third party.
What changes when an agent does it: Agents retrieve broadly to answer narrowly. The access pattern inverts: where a person opened three records, an agent opens three hundred and uses four.
Obligation family: Access control, minimum necessary, purpose limitation. HIPAA Security Rule 164.312; India DPDP purpose limitation.
First evidence asked for: Access logs at the volume an agent actually generates, and the entitlement that scoped the query.

02 ASSERT

What it is: States something a counterparty can act on - a policy, an entitlement,

a fact about their account.
What changes when an agent does it: The statement becomes the organization's statement. Moffatt v. Air Canada (2024 BCCRT 149) settled that the argument 'the chatbot said it, not us' does not survive.
Obligation family: Misrepresentation, conduct, consumer protection.
First evidence asked for: What the agent said, to whom, and what the authoritative source said at that moment.

03 RECOMMEND
What it is: Produces an output a human is expected to act on, without acting itself.
What changes when an agent does it: The human-in-the-loop defense only holds if the human had the material and the time. An approver clicking through at nine seconds per case is a rubber stamp with an audit trail.
Obligation family: Model risk, effective challenge, clinical-decision-support boundaries. US interagency model risk guidance (2026).
First evidence asked for: What the reviewer was shown, what they decided, and how long they had.

04 COMMIT
What it is: Binds the organization - a quote, a promise, a service commitment, a term.
What changes when an agent does it: Authority to bind is a well-understood legal concept that almost nobody has mapped onto a non-human actor. Delegated underwriting authority is the closest existing analog.
Obligation family: Contract formation, delegated authority, producer conduct.
First evidence asked for: The grant that authorized a commitment of that size, in that class, at that moment.

05 TRANSACT
What it is: Moves value: payments, limits, pricing, reserves, positions.
What changes when an agent does it: The blast radius stops being reputational. Controls designed around a human's rate of work - four eyes, daily limits, batch review - do not survive an actor operating at machine rate.
Obligation family: Payment controls, segregation of duties, model risk, fraud and AML.
First evidence asked for: The value and rate limits actually enforced at the point of action, not the ones in the policy.

06 MUTATE RECORD
What it is: Writes to a system of record: claims, patient, customer, ledger.
What changes when an agent does it: The record becomes evidence in someone else's proceeding. Integrity and attribution now matter more than accuracy.
Obligation family: Records integrity and attribution. 21 CFR Part 11 for regulated records; books-and-records duties.
First evidence asked for: An immutable event tying the change to the agent, the authority and the input that justified it.

07 CONFIGURE
What it is: Changes a control: a threshold, a rule, an entitlement, a routing decision.
What changes when an agent does it: The agent is now acting on the control plane rather than inside it. A system that can widen its own limits has no limits.
Obligation family: Change management, segregation of duties, control self-assessment.
First evidence asked for: Change records with approval, and proof the agent cannot alter the controls that bind it.

08 DELEGATE
What it is: Invokes another agent or tool, or grants authority onward.
What changes when an agent does it: Authority becomes transitive. Most stacks cannot answer which human's grant a third-hop tool call was ultimately spending.
Obligation family: Delegation chain, sub-processing, third-party risk. OCC 2023-17; NYDFS Part 500.11 (scoped to access).
First evidence asked for: A machine-readable chain from the accountable human to the executing call.

09 COMMUNICATE EXTERNALLY
What it is: Sends to a customer, counterparty, or a regulator.
What changes when an agent does it: Everything sent is disclosable and retained. Language, channel and timing all acquire obligations.
Obligation family: Conduct, disclosure, retention. CBUAE AI/ML Guidance Note 4b expects Arabic and English disclosure for UAE licensed institutions.
First evidence asked for: The full outbound record, retained for the regime's period, not the application's default.

10 DENY OR TERMINATE
What it is: Refuses service, declines a claim, closes an account, rejects an application.
What changes when an agent does it: This is the class that creates a person with a grievance and a right. It is also the class most often built first, because refusal feels safer than approval.
Obligation family: Adverse action, fairness, appeal and human-review rights. CBUAE 7c; EU AI Act Annex III where applicable.
First evidence asked for: A reconstructable reason in the terms the affected person is entitled to receive.

BOUNDARY TESTS - WHICH CLASS DOES THIS CAPABILITY REALLY BELONG TO?
Applied in order. The first test that returns yes sets the class; capabilities are routinely classified a tier too low.

CAN ITS OUTPUT REACH A PERSON OUTSIDE THE ORGANIZATION WITHOUT A HUMAN READING IT FIRST?
If yes: At least Assert. Not Retrieve.

WOULD A REASONABLE RECIPIENT RELY ON IT?
If yes: At least Assert, and probably Commit.

DOES ANYTHING CHANGE IN A SYSTEM OTHER PEOPLE READ AS TRUE?
If yes: At least Mutate record.

DOES MONEY, A LIMIT, A PRICE OR A RESERVE MOVE?
If yes: Transact, regardless of size.

CAN IT CHANGE A THRESHOLD, RULE OR PERMISSION - INCLUDING ITS OWN?
If yes: Configure. Treat as the highest tier present.

CAN IT CALL SOMETHING THAT CAN DO ANY OF THE ABOVE?
If yes: It inherits that class. Authority is transitive; classification has to be too.

IS THERE A HUMAN APPROVAL STEP?
If yes: It only lowers the class if you can evidence what the approver saw and how long they had.

HOW TO USE IT

1. List capabilities, not agents. One agent commonly spans four classes.
2. For each capability, run the boundary tests in order and record the first yes.
3. Take the highest class reachable without a human in the path - that is the agent's governing class.
4. Map that class to the obligation family, then to the specific regime that binds you.
5. Where the evidence column names something you cannot produce today, that is the gap. Record it as such rather than restating the policy.
6. Re-run when a tool is added. Adding one integration can move an agent two classes.

Built from public standards and general practice. No client information appears anywhere in this document. Regulatory instruments move - check anything cited here against its primary source on the day you rely on it.