Govern the changes that alter AI behaviour.

Treat models, prompts, retrieval and tools as versioned parts of a business system with accountable release decisions.

AI change governance

Change	Review before release
Model or prompt	Task quality, refusal behaviour and regression results
Retrieval source	Permission, freshness and source ownership
Tool access	Input validation, authority and recovery
Operating policy	Review coverage, escalation and accepted limitations

Record the candidate configuration

A release record should identify the model, prompt, source set and tool contracts that produced the evaluated behaviour. Otherwise the results cannot be reliably reproduced.

Make stopping possible

Define conditions that disable a tool, route work to manual review or restore a previous release. A recovery plan needs an operator with the authority to use it.

Govern the specific use of AI

Begin with the purpose of the application and the people affected by its outputs. Record what it may recommend, what it may change and which decisions remain with a person. A single organisation-wide statement about responsible AI cannot replace these application-level boundaries.

Maintain a use-case register with a business owner, source systems, external providers and approved operating scope. Identify consequential decisions and the review process appropriate to them. Record unresolved questions about accuracy, privacy, access or recovery before release. Governance should help a team make and explain a decision, rather than require a large document whose conclusions cannot be traced to the implementation.

Use release gates that can be checked

Agree the evidence required for the intended level of autonomy. A read-only assistant, a draft-producing workflow and an agent that changes business records need different acceptance criteria. Evaluate representative tasks, permission boundaries, refusal behaviour and recovery from an interrupted action.

Separate model behaviour from controls enforced by the application. Tool permissions, approval requirements and limits on an operation should not depend solely on the model following instructions. Record who can change these controls and how the change is reviewed. Where a limitation is accepted, define the permitted use and the person responsible for revisiting it. An unresolved risk should not disappear merely because the pilot is popular.

Review changes after deployment

Model upgrades, new source collections and added tools can materially change the system's behaviour. Route these through a proportionate review process with evaluation evidence and rollback planning. Maintain a way to disable a problematic capability while preserving unaffected workflows where possible.

Review incidents and user feedback for patterns, including information exposure, unsupported answers and excessive review effort. Decide when those findings require a new evaluation case, a design change or a narrower operating scope. Keep the evidence readable by business owners as well as engineers. This resource supports practical governance discussions and does not itself certify compliance with a particular law, standard or procurement requirement.

Can the system change its own permissions?

Permissions belong to the application’s security model. A generated instruction must not grant authority or bypass an approval requirement.

Read the current resource online

https://cobnex-review.rgcsekaraa.workers.dev/trust-hub/responsible-ai/