The Liability Transfer

How Enterprise AI Contracts Assign Responsibility

Richard Foster-Fletcher
MKAI Working Paper, July 2026
research@mkai.org

Abstract

This study reads the public terms and policies for seven enterprise artificial intelligence services: Microsoft 365 Copilot, Gemini for Google Workspace, ChatGPT Enterprise, Claude for Work, Amazon Bedrock with Guardrails, Salesforce Einstein and Agentforce, and Glean. It asks three questions. Do the documents make the customer or user responsible for artificial intelligence output? Do the same documents describe controls that may affect the output before the user sees it, such as filters, instructions, safety checks, retrieved material, tools that scan prompts or responses, data masking, or source marks? Does the wording that assigns responsibility mention those controls?

The answer is consistent across the seven vendors. The terms make the customer responsible for the AI output. Elsewhere, these same contracts disclose that the output is affected by vendor controls such as filters, system instructions, safety features, retrieval rules, data masking, or classifiers. However, the clauses assigning user responsibility completely ignore these controls.

Terms used in this paper

Customer: The organisation that buys or uses the enterprise service under the public terms.

User: The person who enters a prompt or uses the artificial intelligence service.

Output: The answer, text, image, code, or other material returned by the service.

Controls: Settings or systems that can affect the output. These include filters, instructions, safety checks, retrieved material, tools that scan prompts or responses, data masking, and source marks.

Responsibility wording: The part of the public terms that says who is responsible for output, for using output, for checking output, or for claims arising from output.

The Vendors

The seven services are:

- Microsoft 365 Copilot
- Google Gemini for Workspace
- OpenAI ChatGPT Enterprise
- Anthropic Claude for Work
- Amazon Bedrock with Guardrails
- Salesforce Einstein and Agentforce
- Glean

Method

This study uses only public documents. It excludes private contracts, sales materials, demonstrations, and product claims outside the listed public documents.

On 4 July 2026, we read an independent register of 37 public documents across the seven vendors. Thirty-six were read in full. The Google Cloud Data Processing Addendum was read through its main sections, but the copy used ended inside the closing security appendix.

For each vendor, we asked three questions:

1. Does a public document make the customer or user responsible for AI output, use of output, review of output, or claims arising from output?
2. Does a public document describe controls that may affect the output, such as filters, instructions, safety checks, retrieved material, tools that scan prompts or responses, data masking, or source marks?
3. Does the responsibility wording mention those controls?

A document counted only where it was part of the public terms, an acceptable-use policy, a data-processing addendum, or documentation that the public terms point to and make part of the agreement. Product marketing did not count.

Results in one table

The table reports the three questions in plain form. The detailed source list appears in Appendix A, and the vendor-by-vendor evidence appears in Appendix B.

Vendor	Responsibility assigned?	Controls described elsewhere?	Responsibility wording mentions controls?
Microsoft 365 Copilot	Yes	Yes	No
Google Gemini for Workspace	Yes	Yes	No
OpenAI ChatGPT Enterprise	Yes	Yes	No
Anthropic Claude for Work	Yes	Yes	No
Amazon Bedrock with Guardrails	Yes	Yes	No
Salesforce Einstein and Agentforce	Yes	Yes	No
Glean	Yes	Yes	No

The result is the same for every vendor. The terms make the customer or user responsible. The documents also describe controls that affect the output. However, the responsibility wording itself does not mention those controls.

Findings

Finding one: every vendor assigns responsibility for output

All seven vendors place responsibility for output, use of output, review of output, or claims arising from output on the customer or user. The wording varies. Some documents speak directly about responsibility for output. Others place output inside a broader term such as customer content or customer data, then make the customer responsible for that material. The result is the same: responsibility for the output sits with the customer or user.

Finding two: every vendor also describes controls that may affect output

All seven vendors describe controls that can affect output. The controls are described in different ways.

Microsoft refers to content filters, system instructions, and safety systems. System instructions are built-in directions coded into the service.

Google refers to source citations, filters, instructions, and other tools.

OpenAI refers to citation, filtering, safety features, restrictions, GPT configurations, and content pulled from connected applications. Retrieval rules mean the service pulls information from approved sources before producing an answer.

Anthropic says it may block or modify model outputs.

Amazon refers to available filters, other tools, and source or origin marks applied to output.

Salesforce refers to tools that scan inputs and outputs for safety concerns, system policies, data masking, and the use of trusted data to support answers.

Glean refers to content filters, safety systems, and retrieval limited by access permissions.

Finding three: the responsibility wording leaves the controls out

This is the central finding. The documents do not hide the controls. They name them elsewhere. The responsibility wording still makes the customer or user responsible without mentioning those controls. That pattern appears in all seven vendors.

Each vendor places the control descriptions in a different document. Microsoft and Google put their clearest references inside conditions attached to promises to defend output-related claims. Anthropic puts its clearest reference in its Usage Policy. Salesforce puts its controls language in documentation treated as part of the public terms. Glean combines a condition attached to a promise to defend claims with retrieval limited by access permissions. The relevant sentence sits in a different document at each vendor. The pattern remains the same.

Vendor-by-vendor reading

Microsoft 365 Copilot

The responsibility evidence comes from the Enterprise AI Services Code of Conduct, the Product Terms, and the Microsoft Customer Agreement.

The same public terms refer to content filters, metaprompts, safety systems, and Microsoft powers to limit responses on particular topics. In Microsoft wording, metaprompts are built-in instructions for the service.

The responsibility wording does not mention those controls.

Google Gemini for Workspace

The responsibility evidence comes from the Google Workspace Service Specific Terms and the Google Cloud Terms of Service.

The promise-to-defend wording refers to source citations, filters, instructions, and other tools that Google makes available.

The responsibility wording does not mention those controls.

OpenAI ChatGPT Enterprise

The responsibility evidence comes from the OpenAI Services Agreement and its treatment of input and output as Customer Content.

The Service Terms refer to citation, filtering, safety features, restrictions, GPT configurations, and content retrieved from connected applications.

The responsibility wording does not mention those controls.

Anthropic Claude for Work

The responsibility evidence comes from the Commercial Terms of Service and the wording that makes the customer defend certain claims.

The Usage Policy says Anthropic may block or modify model outputs, and it bars attempts to bypass product restrictions or guardrails.

The responsibility wording does not mention those controls.

Amazon Bedrock with Guardrails

The responsibility evidence comes from the AWS Responsible AI Policy, the AWS Service Terms, and the AWS Customer Agreement.

The Service Terms refer to available filters and other tools, and also refer to source or origin marks applied to output.

The responsibility wording does not mention those controls.

Salesforce Einstein and Agentforce

The responsibility evidence comes from the Einstein Platform Notices and License Information and the Artificial Intelligence Acceptable Use Policy.

Document treated as part of the public terms refers to tools that scan inputs and outputs for safety concerns, system policies, data masking, and the use of trusted data to support answers.

The responsibility wording does not mention those controls. The Salesforce finding depends on two links: the Main Services Agreement defines Documentation, and a Trust and Compliance document points to the listed Einstein Trust Layer material.

Glean

The responsibility evidence comes from the Glean AI Addendum, the Acceptable Use Policy, and the Terms of Service.

The AI Addendum and Acceptable Use Policy refer to content filters, safety systems, and retrieval limited by access permissions.

The responsibility wording does not mention those controls.

What the finding means

The terms assign responsibility after the service has produced output. The same set of public documents says the output may have passed through filters, system instructions, scanning tools, data masking, retrieved material, or source marks. The responsibility wording does not mention those controls.

This point is narrow. This paper does not decide whether the wording is valid, fair, enforceable, or commercially normal. It also does not decide whether the controls improve or weaken output. It records the public wording. The customer or user is made responsible. The controls are described somewhere else. The responsibility wording leaves those controls out.

A companion study, *The Override Rank Cannot Reach*, reads the public administration and compliance documentation for the same seven vendors. It found that executive rank has no documented prompt-time effect on these controls. Together, the two studies say this: the customer or user can be made responsible for output that passed through controls they did not set, while rank gives no documented way to affect those controls when the service is in use.

Limitations

The document set is public and dated. It captures the documents listed as accessed on 4 July 2026. Public terms change often. This study does not claim that the same wording remained in place after that date.

Private agreements may differ from the public documents. Order forms, negotiated riders, enterprise addenda, or private master agreements may change the wording. Those documents were outside the public document set.

The study reads text, not what happens inside live systems. A product may have controls that are not described in the public terms, and a public term may describe controls differently from how they work in a particular deployment.

The Salesforce result depends on documentation treated as part of the public terms. The Main Services Agreement defines Documentation, and the Trust and Compliance material links to the Einstein Trust Layer material used here. A narrower reading that excludes that help documentation would produce a different Salesforce result.

One person applied the reading rules. The original research notes record a check of twenty-eight quoted extracts. Twenty-seven matched the source. One Google Workspace condition attached to a promise to defend claims was corrected to the wording at source. The individual list of those twenty-eight checks was not retained, so this version does not claim a quote-by-quote verification record.

Suggested citation

Foster-Fletcher, R. (2026). The Liability Transfer: How enterprise artificial intelligence terms assign responsibility in one place and describe controls elsewhere. MKAI Working Paper. Available at mkai.org.

References

Elish, M. C. (2019). Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. Engaging Science, Technology, and Society, 5, 40-60. doi:10.17351/ests2019.260.

Foster-Fletcher, R. (2026). Blamefall: Why mandated AI concentrates liability downward. Available at https://fosterfletcher.com/blamefall/

Foster-Fletcher, R. (2026). The Override Rank Cannot Reach. MKAI Working Paper.

Lawton, T., Morgan, P., Porter, Z., Hickey, S., Cunningham, A., Hughes, N., Iacovides, I., Jia, Y., Sharma, V., and Habli, I. (2024). Clinicians risk becoming "liability sinks" for artificial intelligence. Future Healthcare Journal, 11(1), 100007. doi:10.1016/j.fhj.2024.100007.

Porsdam Mann, S., Earp, B. D., Nyholm, S., Danaher, J., Moller, N., Bowman-Smart, H., Hatherley, J., Koplin, J., Plozza, M., Rodger, D., Treit, P. V., Renard, G., McMillan, J., and Savulescu, J. (2023). Generative AI entails a credit-blame asymmetry. Nature Machine Intelligence, 5, 472-475. doi:10.1038/s42256-023-00653-1.

Appendix A: source list

All documents in this appendix were accessed on 4 July 2026. The list contains thirty-seven public documents. Thirty-six were read in full. The Google Cloud Data Processing Addendum was read through its main sections, with the copy used for this study ending inside the closing security appendix.

Microsoft

- Product Terms, Universal License Terms for Online Services (acceptable-use policy section; Generative AI terms; Customer Copyright Commitment) (Contract document (product terms; includes acceptable-use policy and generative artificial intelligence terms); Publication effective 1 July 2026; Read in full). microsoft.com/licensing/terms/product/ForOnlineServices/all
- Product Terms, Microsoft 365 product entry (Microsoft 365 Copilot section) (Contract document (product terms); Publication effective 1 July 2026; Read in full). microsoft.com/licensing/terms/productoffering/Microsoft365/EAEAS
- Product Terms, Glossary (Metaprompts and related definitions) (Contract document (product terms; definitions); Publication effective 1 July 2026; Read in full). microsoft.com/licensing/terms/product/Glossary/all
- Enterprise AI Services Code of Conduct, v4.0 (treated as part of the Product Terms) (Document treated as part of the public terms; 1 May 2026; Read in full). learn.microsoft.com/en-us/legal/ai-code-of-conduct
- Microsoft Customer Agreement (public UK sample) (Contract document (customer agreement); March 2023; Read in full). microsoft.com/licensing/docs/customeragreement

  Note: Public only as regional samples dated March 2023; UK sample read. Staleness and public-baseline caveat.
- **Products and Services Data Protection Addendum** (data-processing addendum; 22 May 2026; Read in full). microsoft.com/licensing/docs/view/Microsoft-Products-and-Services-Data-Protection-Addendum-DPA

Google

- **Google Cloud Acceptable Use Policy (serves Workspace)** (Policy (acceptable-use policy); 13 October 2025; Read in full). workspace.google.com/terms/use_policy/
- **Generative AI Prohibited Use Policy (treated as part of Service Specific Terms 12.3)** (Policy treated as part of the public terms; 17 December 2024; Read in full). policies.google.com/terms/generative-ai/use-policy
- Google Workspace Service Specific Terms (Contract document (service-specific terms); 5 February 2026; Read in full). workspace.google.com/terms/service-terms/
- Google Cloud Terms of Service (integrated Workspace terms; serves as enterprise terms and customer agreement) (Contract document (serves as both enterprise terms and customer agreement); 1 June 2026; Read in full). workspace.google.com/terms/premier_terms/
- **Cloud Data Processing Addendum** (data-processing addendum; Current; effective on acceptance; Partial: truncated inside closing security appendix). cloud.google.com/terms/data-processing-addendum

  Note: The 37th document. The copy used for this study ended inside the closing security appendix after all main sections had been captured. This is the only document not read in full.
- Generative AI Services covered by the promise to defend claims (treated as part of Service Specific Terms 12.10(a)) (Document treated as part of the public terms (promise to defend claims terms); 22 April 2026; Read in full). cloud.google.com/terms/generative-ai-indemnified-services

OpenAI

- **Usage Policies** (Policy (acceptable-use policy); Effective 29 October 2025; Read in full). openai.com/policies/usage-policies
- OpenAI Services Agreement (enterprise terms and customer agreement) (Contract document (serves as both enterprise terms and customer agreement); Effective 1 January 2026; Read in full). openai.com/policies/services-agreement
- Service Terms (section 3: ChatGPT Enterprise) (Contract document (service terms); Updated 2 June 2026; Read in full). openai.com/policies/service-terms
- **Data Processing Addendum, v.010126 (read from OpenAI content-network document)** (data-processing addendum; 1 January 2026; Read in full). cdn.openai.com/pdf/openai-data-processing-addendum.pdf

  Note: Data-processing page serves an execution form; full addendum read from OpenAI's own content network.

Anthropic

- **Usage Policy (self-described Acceptable Use Policy)** (Policy (acceptable-use policy); Effective 15 September 2025; Read in full). anthropic.com/legal/aup

- Commercial Terms of Service (enterprise terms and customer agreement) (Contract document (serves as both enterprise terms and customer agreement); Effective 17 June 2025; Read in full). anthropic.com/legal/commercial-terms
- Service Specific Terms (section A: Claude for Work) (Contract document (service-specific terms); Effective 8 June 2026; Read in full). anthropic.com/legal/service-specific-terms
- **Data Processing Addendum** (data-processing addendum; Effective 24 February 2025; Read in full). anthropic.com/legal/data-processing-addendum

Amazon

- **AWS Acceptable Use Policy** (Policy (acceptable-use policy); 1 July 2021; Read in full). aws.amazon.com/aup/
- **AWS Responsible AI Policy (treated as part of Service Terms 1.23)** (Policy treated as part of the public terms; 13 January 2025; Read in full). aws.amazon.com/ai/responsible-ai/policy/
- AWS Service Terms (section 50; 50.12 Amazon Bedrock) (Contract document (service terms); 30 June 2026; Read in full). aws.amazon.com/service-terms/
- AWS Customer Agreement (Contract document (customer agreement); 13 March 2026; Read in full). aws.amazon.com/agreement/
- **AWS Data Processing Addendum** (data-processing addendum; No date stated in text; Read in full). d1.awsstatic.com/legal/aws-dpa/aws-dpa.pdf

  Note: No effective date stated in the document text.

Salesforce

- **Acceptable Use and External-Facing Services Policy** (Policy (acceptable-use policy); 8 July 2025; Read in full). salesforce.com (legal agreements page, document)

  Note: Acceptable-use page returned an empty body; operative policy read from the linked document.
- **Artificial Intelligence Acceptable Use Policy** (Policy (acceptable-use policy); 11 December 2025; Read in full). salesforce.com (legal agreements page, document)
- Main Services Agreement (enterprise terms and customer agreement) (Contract document (serves as both enterprise terms and customer agreement); 1 June 2026; Read in full). salesforce.com/.../salesforce_MSA.pdf
- **Data Processing Addendum** (data-processing addendum; Revision April 2026; Read in full). salesforce.com (legal agreements page, document)
- Einstein Platform Notices and License Information (document treated as part of the public terms) (Document treated as part of the public terms; 22 May 2026; Read in full). salesforce.com (Trust and Compliance Documentation, document)
- Einstein Platform Security, Privacy and Architecture (document treated as part of the public terms) (Document treated as part of the public terms; 22 May 2026; Read in full). salesforce.com (Trust and Compliance Documentation, document)
- Salesforce Trusted AI Resources (document treated as part of the public terms) (Document treated as part of the public terms; May 2026; Read in full). salesforce.com (legal misc, document)
- Einstein Trust Layer (Salesforce Help; Documentation under Main Services Agreement section 1) (Document treated as part of the public terms (product help treated as part of defined term); Undated; current; Read in full). help.salesforce.com/s/articleView?id=ai.generative_ai_trust_layer.htm

  Note: The link to this help article depends on the Main Services Agreement definition of Documentation and on a Trust and Compliance document that lists the article by name. A narrower reader may exclude this item.

Glean

- **Acceptable Use Policy** (Policy (acceptable-use policy); 5 June 2026; Read in full). glean.com/legal/aup
- Terms of Service, v5Jun2026 (enterprise terms and public customer agreement) (Contract document (serves as both enterprise terms and customer agreement); 5 June 2026; Read in full). glean.com/legal/terms

  Note: No negotiated master agreement published; click-through Terms of Service stands as the public customer agreement.
- Glean AI Addendum (treated as part of Terms of Service 1.3) (Document treated as part of the public terms; 5 June 2026; Read in full). glean.com/legal/ai-addendum
- **Data Processing Addendum (Online), v6Mar2026** (data-processing addendum; 6 March 2026; Read in full). glean.com/legal/dpa

Appendix B: evidence summary

This appendix gives the short evidence basis for the three questions. The wording is concise so the main paper remains readable.

Microsoft 365 Copilot

Responsibility: The responsibility evidence comes from the Enterprise AI Services Code of Conduct, the Product Terms, and the Microsoft Customer Agreement.

Controls described: The same public terms refer to content filters, metaprompts, safety systems, and Microsoft powers to limit responses on particular topics. In Microsoft wording, metaprompts are built-in instructions for the service.

What is missing: The responsibility wording does not mention those controls.

Google Gemini for Workspace

Responsibility: The responsibility evidence comes from the Google Workspace Service Specific Terms and the Google Cloud Terms of Service.

Controls described: The promise-to-defend wording refers to source citations, filters, instructions, and other tools that Google makes available.

What is missing: The responsibility wording does not mention those controls.

OpenAI ChatGPT Enterprise

Responsibility: The responsibility evidence comes from the OpenAI Services Agreement and its treatment of input and output as Customer Content.

Controls described: The Service Terms refer to citation, filtering, safety features, restrictions, GPT configurations, and content retrieved from connected applications.

What is missing: The responsibility wording does not mention those controls.

Anthropic Claude for Work

Responsibility: The responsibility evidence comes from the Commercial Terms of Service and the wording that makes the customer defend certain claims.

Controls described: The Usage Policy says Anthropic may block or modify model outputs, and it bars attempts to bypass product restrictions or guardrails.

What is missing: The responsibility wording does not mention those controls.

Amazon Bedrock with Guardrails

Responsibility: The responsibility evidence comes from the AWS Responsible AI Policy, the AWS Service Terms, and the AWS Customer Agreement.

Controls described: The Service Terms refer to available filters and other tools, and also refer to source or origin marks applied to output.

What is missing: The responsibility wording does not mention those controls.

Salesforce Einstein and Agentforce

Responsibility: The responsibility evidence comes from the Einstein Platform Notices and License Information and the Artificial Intelligence Acceptable Use Policy.

Controls described: Document treated as part of the public terms refers to tools that scan inputs and outputs for safety concerns, system policies, data masking, and the use of trusted data to support answers.

What is missing: The responsibility wording does not mention those controls. The Salesforce finding depends on two links: the Main Services Agreement defines Documentation, and a Trust and Compliance document points to the listed Einstein Trust Layer material.

Glean

Responsibility: The responsibility evidence comes from the Glean AI Addendum, the Acceptable Use Policy, and the Terms of Service.

Controls described: The AI Addendum and Acceptable Use Policy refer to content filters, safety systems, and retrieval limited by access permissions.

What is missing: The responsibility wording does not mention those controls.