KNOSTIC

LLM Data Leakage Detection and Response for Enterprise AI Search tools

Solution Brief — How Knostic assesses and remediates LLM data exposure via Copilot, Glean and other AI Search tools.

KNOSTIC

01 Executive Summary

AI-enabled chatbots and enterprise search tools, such as Copilot for Microsoft 365 and Glean, offer significant productivity gains by accelerating the discovery of hard-to-find institutional knowledge.

Although LLMs increase productivity, they also increase the risk of data leakage. LLMs regularly overshare and violate the principle of “need-to-know”, which increases security, privacy, and compliance risks. While slowing or halting LLM deployment might mitigate risks, doing so could cause the organization to fall behind technologically and impair the business’s market position

Traditional data permissions and access controls are insufficient for managing LLMs, as these models frequently overshare sensitive data and can infer confidential information even from limited data. Most organizations also lack granular authorization-level capabilities that fit the needs of LLMs.

Knostic helps organizations continuously identify and remediate data exposure and leakage in LLM-powered enterprise search tools, such as Copilot for Microsoft 365 and Glean. Acting as a safety net, the platform flags sensitive or private information that may be inadvertently exposed or at risk of leaking. Knostic offers automation to assist in adjusting permissions based on these findings, aligning access within users' need-to-know boundaries.

02 The Enterprise LLM Challenge

Mind the (LLM) gap

Adopting LLM tools, such as Copilot for Microsoft 365 and Glean, delivers significant productivity gains by helping users to connect the dots across disparate pieces of institutional data and information.

Although LLMs increase productivity, they also exacerbate data leakage problems. LLMs regularly overshare and violate the principle of need-to-know. The implications? A substantially elevated risk of exposing sensitive (or unauthorized) content.

In the LLM era, reliance on traditional data permissions and information classification is insufficient to meet security needs:

1. **Sensitive data is frequently overshared**, due to misaligned and out-of-date permission settings, over provisioned entitlements and/or failing labeling efforts
2. **LLMs are inference engines** - even if users have precisely the correct permissions on the specific underlying folders and files, LLMs can infer sensitive knowledge from the available data.
3. **Authorization-level capabilities** - Enterprises (and most IAM systems) largely do not have authorization capabilities that fit the needs of LLMs. Meaning, the LLM is unable to properly determine which data and information is allowable for each user.

These issues are exacerbated by the "**flat knowledge problem**," which parallels the challenge of flat networks. Similar to how on a flat network a computer can connect to any other, causing security risks, LLMs can connect every user to potentially any data. For example, someone from engineering could obtain confidential HR data with only one query.

Business and Security Impact

Without addressing these core need-to-know issues with LLMs, an enterprise would be exposed to security, privacy, and compliance risks:

- There is a high risk that previously hidden data will be exposed beyond users' need-to-know.
- Users can bypass current permission mechanisms through the LLM
- Privacy and compliance violations become more likely.
- Slowing or halting LLM deployment plans due to the previously mentioned security risks can impact the competitiveness of the business.

03 Addressing LLM Data Leakage

To begin addressing LLM need-to-know concerns, organizations need visibility into exposed data. An organization can perform an assessment that will subsequently help build a need-to-know authorization scheme.

The necessary steps of the assessment process include:

1. Map business topics accessible to users. We use a list general to any company, such as HR, Finance, and legal. Then we generate a list of topics based on the industry and company specific topics.
2. Identify instances of oversharing, where users have access to information beyond their need-to-know boundaries.
3. Trace the exposed topics back to specific data sources, such as files or databases, where the exposure originated.
4. Remediate the data leakage by adjusting file or DB permissions.

04 Knostic's Solution to the Enterprise LLM Challenge

Knostic helps organizations continuously identify and remediate data exposure and leakage in LLM-powered enterprise search tools, such as Microsoft O365 Copilot and Glean.

The platform serves as a safety net by identifying and flagging sensitive, or private information that is exposed and could potentially leak. Then, we provide automation for adjusting permissions based on these findings and the users' need-to-know boundaries

Our offering:

- Readiness Assessment
  - Exposure mapping
  - Investigation capabilities
- Remediation capabilities
- Program building
- Runtime monitoring

Readiness Assessment

The readiness assessment identifies initial gaps in the organization's readiness to adopt enterprise LLM-based search tools, based on business topic exposure mapping and a permissions gap analysis.

To identify exposures, the need-to-know boundaries of various user profiles are assessed. These boundaries are defined by an examination of what business context is relevant to their role. The assessment takes place across three topic tiers:

1. **General:** Topics regarded as sensitive regardless of the organization. E.g. compensation, sales numbers, IP, API keys, and audit-related considerations such as HIPPAA, PCI, GDRC-related content.
2. **Industry:** Topics sensitive to a specific industry. E.g. buy and sell side separation in banking.
3. **Company:** Topics customized to customer needs specific to that company.

Through these tiers, the organization learns what business topics are accessible from each vantage point, starting with the least-privilege access or basic user accounts (e.g., the "everyone" group).
Profiles are provided by the company in the following three formats:

1. Real user accounts.
2. Accounts that copy user profiles entitlements.
3. Generic profiles that mimics a role, e.g. QA, Marketing, Financial Analyst.

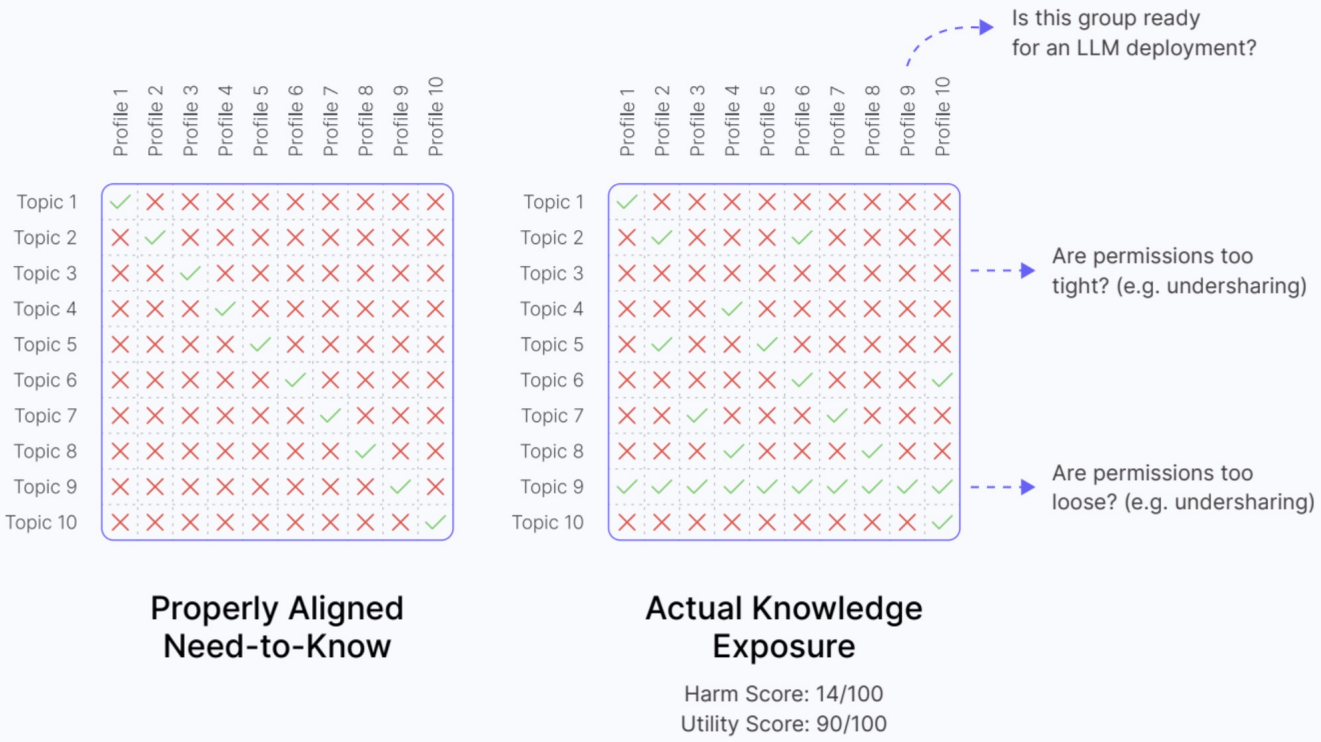
Exposure Map

The first result of the initial assessment is an exposure map which gives the organization visibility into need-to-know boundary violations of each user. We determine whether employees are either over-exposed or under-exposed for their roles. E.g. a QA engineer might have access to payroll information, which could indicate over-exposure. This includes assessing the sensitivity of that information.

Supporting LLM Adoption

Based on the results from the initial assessment and the exposure map, the organization can visualize their standing as they correct permissions and build an adoption plan.

The diagram on the left shows an ideal scenario of ‘perfect’ permissions. For example, if we assume Topic 1 represents Finance and Profile 1 is the CFO, the checkmark indicates that appropriate access has been granted.



Alternatively, the diagram on the right reflects a more realistic view of an organization's exposure:

- **Profile 9** appears well-organized and ready for deployment
- **Topic 9** seems over-permissioned, potentially exposing sensitive data to profiles that exceed their need-to-know access
- **Topic 3** might be under-permissioned, preventing several employees from accessing the information needed to perform their roles.

Investigation Capabilities

In-depth investigation capabilities help assess the sensitivity of findings within each of the three tiers.

Investigation capabilities include the ability to override Knostic’s AI-based recommendation engine, and transparency regarding how those decisions are made.

Remediation

Based on automation policies and human analyst approval, the system provides several remediation paths, either directly or through integrations.

Remediation features include;

- **Violations Triage:** A system for prioritizing and managing violations effectively.
- **AI-Based Recommendation Engine:** A machine learning-driven engine that offers recommendations based on analysis.

Ongoing Assessment and Remediation

As new knowledge is created, as individual roles change, and morphs as people move in, out, up or down in the organization their need to know changes. The remediation progress is tracked by assessing whether the offending content remains accessible.
Findings and recommendations should be clearly presented in a report, as illustrated in the following table.

High Priority Findings

Topic	Location	No Permissions Profile - Has NTK?	{Secondary Role}- Has NTK?
Confidential Agreements	SharePoint LINK	NO	NO
Mergers and Acquisitions	SharePoint LINK	NO	YES?
Applicant Information*	SharePoint LINK	NO	NO
Audit Issues	SharePoint LINK	NO	NO
Settlements	SharePoint LINK	NO	No
Passwords	SharePoint LINK	NO	NO
Legal Disputes	SharePoint LINK	NO	No

Actionable Intelligence

Through Knostic's platform, organizations can:

- Map knowledge boundaries with precision
- Assess data exposure risk through sophisticated probing techniques
- Generate and validate findings through AI-based recommendation engines
- Track remediation progress with clear metrics and reporting

Let's Talk and Share Knowledge

Learn more about how to assess and remediate data leakage from AI powered enterprise search tools at Knostic.ai or connect with us on LinkedIn for the latest research and insights from our R&D teams.

About Knostic

Knostic is the world's first provider of need-to-know based access controls for Large Language Models (LLMs). With knowledge-centric capabilities, Knostic enables organizations to accelerate the adoption of LLMs and drive AI-powered innovation without compromising value, security, or safety.

Knostic's leadership in the field is validated by significant industry recognition:

- RSA Conference 2024 Launch Pad startup competition winner
- First place at the Blackhat 2024 Startup Spotlight competition