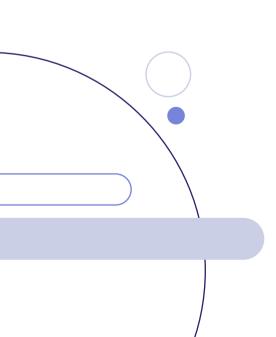




Introduction

Protecting employee privacy is a crucial aspect of your role as an HR professional, particularly when utilizing AI tools to process and analyze data.

This guide outlines which types of employee data are personally identifiable and which combinations of seemingly harmless data points can also expose identity. It includes what employee information is safe to use when working with AI systems and best-practice tips for protecting employee privacy when using AI.





Understanding Employee Data Categories

DATA TYPE	DESCRIPTION	EXAMPLES	AI USE CONSIDERATIONS
Directly identifiable (PII)	Data that can identify a person immediately	Name, employee ID, work email, photo, address	Never input directly into AI tools unless anonymized and approved. Always remove or mask these details.
Indirectly identifiable	Data that might seem harmless on its own but can reveal identity when combined with other details, known as the mosaic effect	Department + job title + location; age + tenure + performance rating	Use only in aggregated form. Review combinations carefully — small team sizes or unique roles increase risk.
Sensitive personal data	Information that requires extra protection under privacy laws, including anything revealing personal attributes, beliefs, or health	Health status, disability, ethnicity, gender identity, religion, political opinions, sexual orientation	Treat with highest caution. Only analyze or share when aggregated, anonymized, and explicitly justified.
Inferred data (AI- generated)	Data created by AI from existing inputs: insights, predictions, or classifications about a person	"Attrition risk score," "promotion readiness," "sentiment trend"	Handle like sensitive data. Always validate with human review before use in decisions.
Nonidentifiable / aggregated data	Fully deidentified data that cannot be traced back to an individual	Company-wide engagement score; average salary by department	It is safe for reporting and AI use. Still ensure anonymity is preserved during uploads or exports.

What Happens When Employee Data Is Combined?

Even when individual data points appear harmless, combining them can identify someone. This is called the **mosaic effect**—small fragments that, when joined, reveal the whole picture. This can unintentionally infringe on privacy laws.

Example combination	Level	Why it matters	What to do
Department + job title + office location	High	Can single out one person in small teams or unique roles.	Aggregate to group level before analysis.
Age + gender + years of service	Medium	May identify individuals in specific demographic groups.	Round or band data into ranges (e.g., "30–40" instead of "32").
Engagement score + tenure + team size	Medium	In small departments, scores can reveal who gave feedback.	Report only for groups of 5+ people.
Health condition + department	Very high	This directly exposes sensitive personal data.	Never share unless anonymized and approved by governance.

What Data You Can Use in AI Tools (and What You Shouldn't)



Safe to use

Aggregated or anonymized data sets

Public or generic content (e.g., role descriptions, policy summaries)

Synthetic or sample data



Use with caution

Department-level trends

Demographic data with small sample sizes

Inferred data about individuals (AI predictions)



Never use without anonymization or approval

Names, ID numbers, personal emails

Health, medical, or family-related details

Salary, disciplinary, or performance records

Practical Steps for Protecting Employee Data When Using AI

Anonymize first.

Remove names, emails, and identifiers before upload.

4 Keep it minimal.
Only use what's essential for the AI's task.

Aggregate wherever possible.

Report by group or team, not individual.

Document your actions.

Record what was removed or masked before sharing.

Check tool settings.

Turn off data retention, storage, or model training.

Ask before you share.

If the purpose or audience isn't clear, stop and confirm.

Pause-or-Proceed Questions Before Using AI to Process Data

