

# Perceptyx AI: Model Governance & Overview

**Driving Employee Experience through Ethical, Secure, and Actionable AI.**

This document outlines key information about the AI models utilized to power Perceptyx’s products. Our data science team customizes commercially available LLMs from reputable providers to align with the specific requirements of our offerings.

## Our AI Model Portfolio

We utilize four distinct types of AI to ensure accuracy, privacy, and meaningful insights:

- **Discriminative AI (Proprietary):** AI Models that annotate employee comments with emotion, intent, noteworthiness, sentiment, and themes.
- **Generative AI (Opt-in Only):** Comment Copilot and AI Coach are powered by Open AI (GPT-4.1, GPT-4o mini, and GPT-5).
- **Privacy AI:** Automated Named Entity Recognition (NER) to detect and mask sensitive data before processing.
- **Rule-Based AI:** Expert-defined logic for consistent, automated decision-making.

## Intended Use & Access

**Purpose:** To enhance employee listening, provide actionable feedback insights, and support behavioral change through personalized coaching.

- **Primary Users:** Central business teams, HR Admins, and Senior Leaders.
- **Guardrails:** Role-based access control (RBAC). AI does not make decisions about individuals or impact rights outside the organization.
- **Human-in-the-loop:** Designed to be intuitive; users can pre-filter data and use curated prompts to ensure relevance.

Use Case	Description
<b>Employee Listening</b>	Enhancement through tools like Comment Copilot, Narrative Analysis, Conversational Listening, AI Coach, and Intelligent Nudges.
<b>Actionable Insights</b>	Providing insights from feedback using sentiment, theme, intent, and emotion modeling.
<b>Personalized Support</b>	Supporting coaching and behavioral change through AI-driven recommendations.
<b>Organizational Analysis</b>	Analyzing comment data from various sources to identify strengths and areas for improvement.

**Intended Users:** Central business teams, Admins, and Senior Leaders who aim to improve employee experience.

**Non-Intended Use:** Not intended to impact individuals outside the user organization or their rights. AI is not used to create content or make recommendations about individuals within the organization.

## Privacy, Ethics & Bias Mitigation

Perceptyx adheres to a "Privacy-by-Design" philosophy.

- **No Customer Data for Training:** Perceptyx does not use your specific data to train our proprietary models or OpenAI's models.
- **Bias Reduction:** The Data Science team measures and minimizes bias during AI model training.
- **Anonymization:** Proprietary training data is aggregated and anonymized to protect individual identity.

## Training Data

- **Discriminative AI:** Separate proprietary and anonymized datasets, aggregated from open-source datasets and pre-cleaned, bootstrapped in-house comment datasets.
- **Generative AI:** In-house experts refine responses and engineer prompts based on workforce industry best practices. Customer data is NOT used for training (Perceptyx or OpenAI).

## Performance Factors

- Variations in input training data, model parameterization, training regimen, and objective function.
- Demographic imbalances are addressed via pre-processing data augmentations (e.g., pronoun replacement) and in-processing procedures (e.g., over/under sampling) to measure and reduce bias.

## Quality & Reliability

We treat AI performance as a rigorous engineering discipline:

- **Rigorous Testing:** Our AI models are evaluated on a separate validation set to promote accuracy and generalizability. Performance is measured using various metrics, including precision, recall, F1 score, accuracy, and ROC/AUROC curves.
- **Monitoring:** We run internal quality tests to monitor accuracy, improve system performance, and refine responses over time.
- **Hallucination Control:** We minimize AI "hallucinations" through structured guidance, curated behavioral science research, and state-of-the-art model selection.

## Responsible AI Principles

<b><i>Principle</i></b>	<b><i>Our Commitment</i></b>
<b><i>Fairness and Bias Prevention</i></b>	Measures to measure and minimize bias; use of data augmentations and over/under sampling to address demographic imbalance. Models trained on augmented datasets with artificial noise to increase robustness.
<b><i>Explainability and Transparency</i></b>	Inputs/outputs and model performance metrics are logged. Data Science team can trace and explain results of all Perceptyx AI systems. AI features are labeled within the platform. All AI features are clearly labeled and GenAI is strictly opt-in
<b><i>Accountability and Privacy</i></b>	No customer-specific data used in model training. Adherence to GDPR and other data privacy laws. Generative AI features are opt-in. Customers may choose to disable AI products. ISO 27001 and SOC 2 Type 2 compliant.
<b><i>Safety and Security</i></b>	Regular static/dynamic code reviews and external penetration testing. Regular privacy and security audits. All AI systems tested against held-out datasets.
<b><i>Guardrails and Ethical Use</i></b>	Cross-functional collaboration on AI principles. Guardrails implemented (e.g., pre-set prompts, misuse prevention) to promote relevance, accuracy, and ethical use.
<b><i>Sustainability</i></b>	Real-time monitoring, comprehensive system instrumentation, and regular assessments of system variance to maintain sustainable operations.

**Note:** AI-generated content is intended to summarize data and coach managers—it is never used to create content or recommendations regarding specific individuals.