

July 2025

## SieraPlex System Overview

A technical architecture guide for partners,  
advisors, and enterprise clients



# Table of Contents

SieraPlex System Overview ..... 3

1. Purpose and Scope..... 3

2. System Architecture ..... 3

2.1. SiFT (Siera Finite-State Transducer) ..... 3

2.2. SieraNorm ..... 3

2.3. SieraMat (Materiality Extractor) ..... 3

2.4. Stream Configurator ..... 3

2.5. Weekly Report Generator..... 4

2.6. Delivery and Logging..... 4

3. Data Inputs and Storage ..... 4

4. Event Tiering Logic..... 4

5. Automation vs. Operator Control ..... 5

6. Example: OilCo ESG Risk Stream..... 5

7. Roadmap & Integration Potential..... 5

# SieraPlex System Overview

A technical architecture guide for partners, advisors, and enterprise clients

## 1. Purpose and Scope

This document provides a technical and conceptual overview of the SieraPlex system — a fully automated media intelligence platform that transforms unstructured news into structured, role-specific insight streams. The system combines deterministic event detection with AI-driven salience modelling, designed for decision-makers in corporate strategy and planning, financial, sales & business development, legal & compliance and ESG roles.

We detect about 50 000 000 events per year.

## 2. System Architecture

SieraPlex is composed of six main components, operating in a modular, end-to-end pipeline:

### 2.1. SiFT (Siera Finite-State Transducer)

- Written in C for high-speed deterministic processing
- Classifies over 430 business event types from article snippets
- Supports high recall/precision use cases (90%+ for trained company sectors)

### 2.2. SieraNorm

- Extracts corporate events from news
- Lexicographical system to identify and disambiguate subject companies and extract their stock prices
- Company sector and industry taxonomy allows easy defaults of competitors
- Assigns date, company ID, event type, source
- Handles overlapping events, for instance, “Exxon buys 1000 electric vehicles from General Motors” would create two events: One denoting Exxon buying and the other GM selling
- The knowledge graph allows reporting on the supply-chain interdependencies
- Time-zone mapping

### 2.3. SieraMat (Materiality Extractor)

- Flags events as material based on:
  - Media elevation (vs. historical baseline)
  - Article count thresholds
  - Sector-specific impact indicators
  - Stock price volatility

### 2.4. Stream Configurator

- Uses a local LLM to generate event inclusion policies based on:
  - Purpose (e.g., risk, ESG, innovation)



- Company or Sector
- User Role
- Generates a stream profile per (company/sector × role × purpose), stored in MySQL
  - Tier 1 (always include)
  - Tier 2 (include if moderate elevation)
  - Tier 3 fallback (high elevation + high volume)
  - Material override rules by amending MySql tables

## 2.5. Weekly Report Generator

- Loads stream config and filters last 7 days of events
- Applies stream thresholds and material override rules
- For selected events:
  - LLM filters for saliency from user perspective
  - LLM generates 100-word summary
  - LLM fact-checks summary against cached article content
- Outputs Markdown/HTML ready for MailChimp or API
- Can be run at any time, but also scheduled as chron-job

## 2.6. Delivery and Logging

- Report delivered weekly via email or dashboard
- Logs include:
  - Events selected per stream
  - LLM outputs (summary, saliency filter, fact-check)
  - Stream version ID and config snapshot

# 3. Data Inputs and Storage

- **Articles:** Ingested daily, snippet form (title + lead paragraph)
- **Events:** Structured output from SiFT → SieraNorm
- **Company/Sector Metadata:** Enriched from internal and external sources
- **Stream Configs:** Stored in MySQL with job function + purpose metadata
- **Reports:** Generated weekly, stored by stream ID and week

# 4. Event Tiering Logic

Tier	Inclusion Rule	Typical Events
T1	Always include	FIN_guidancecut, LIT_lawsuit, DIS_accident



Tier	Inclusion Rule	Typical Events
T2	Include if media elevation $\geq 30$ or $\geq 3$ articles/24h	ENV_pollution, OPS_strike
T3	Include if elevation $> 100$ AND $\geq 10$ articles/24h	All other events
Material Override	Always include if material (custom rules)	e.g. recent DIS_damageenvir in Bakken

## 5. Automation vs. Operator Control

- The system is fully automated by default
- Operators can:
  - Override stream config (e.g., promote or suppress event types)
  - Add company-specific material triggers
  - Approve summaries (optional, not required)
- Each report is auditable and reproducible from config + logs

## 6. Example: OilCo ESG Risk Stream

- **Purpose:** Corporate risk insight
- **Role:** ESG officer
- **Recent triggers:** ENV\_pollution, DIS\_lawsuit, SOC\_protest
- **Report includes:**
  - 6 events from past 7 days
  - Peer comparison on pollution event frequency
  - Summaries generated and fact-checked by LLM

## 7. Roadmap & Integration Potential

- Daily report variants
- Dashboard visualisations with stream logic exposed
- CRM integrations (e.g. Salesforce, HubSpot)
- Lead generation overlays for professional services
- Industry-wide benchmarking streams (for public marketing)

SieraPlex is designed not just to detect events — but to **deliver signal** to the right people at the right time, with the option to add a consultative lens.