emtelligent®

Turn Unstructured Clinical Data into Real-World Evidence at Scale

Billions of unstructured clinical notes — no structure, no clarity. The challenge: rip the signal from the noise, surface the hidden clinical truth, and ignite research breakthroughs, measurable quality gains, and population health outcomes.

Program Highlights

5.1B
Clinical Notes

256M
Clinical Parameters Extracted

98.4%
Confidence across 6 Therapeutic Areas

Respiratory, Circulatory, Autoimmune, Neurology, Gastroenterology, Oncology

79.2%
Increase in the Amount of Saleable Data

A Hidden Opportunity

A healthcare data reseller identified an opportunity to increase the value of its dataset offered to customers on a subscription basis. Buyers in payer, provider, life sciences, and device markets need structured datasets, not PDFs and prose. To do that, they needed to convert **5.1 billion unstructured notes** — spanning thousands of clinicians and formats — into structured data customers can use.

An initial pilot processed 43,000 notes to extract key respiratory parameters (e.g., FEV1, FEV1/FVC, CAT score) with 99.3% accuracy, giving the client confidence that emtelligent was the right partner for large-scale extraction. In less than six months, emtelligent implemented its Medical Language Engine, a purpose-built healthcare data extraction platform, to process the dataset and extract similar clinical parameters across six therapeutic areas. The program scaled to production with weekly data deliveries and clinical validation for quality control, producing millions of structured data points for the customer's commercial solutions

The downstream impact of enhanced structured data is tangible. Accurate data extraction improves healthcare and business outcomes, driving revenue growth and strengthening long-term customer relationships.

The downstream impact of enhanced structured data is tangible. Accurate data extraction improves healthcare and business outcomes, driving revenue growth and strengthening long-term customer relationships.

Commercial Impact

With structured clinical measurements in place, the customer's solution catalog gains the clinical detail buyers expect. Healthcare teams powered by these datasets will benefit from:

More Precise Population Health: Identifying cohorts using detailed clinical characteristics improves program targeting.
Streamlined Regulatory Reporting: Comprehensive, auditable reports can be assembled for compliance and accreditation.
RWE Depth: Pharma and device teams will access more complete clinical pictures for post-market surveillance and outcomes studies.
Accelerated Research: Trials and observational studies accessing a larger share of relevant patient data start faster.
Automation Efficiency: Automatic extraction of structured data prevents costly manual chart reviews.

The Approach

The engagement transformed narrative notes into structured clinical measurements, aligned with the formats and definitions used by the data reseller's customers in their studies and healthcare programs. Models were specialized for six therapeutic areas:

- Respiratory
- Circulatory
- Autoimmune
- Neurology
- Gastroenterology
- Oncology

Extraction logic separated current findings from historical and family-history references, preserved timing critical for longitudinal use, and removed duplicates to avoid inflation. Each measurement carried a confidence score to support thresholding by use case.

The Result

The table below highlights the outcome of the extraction engagement — demonstrating significant model rigor, not just volume. emtelligent's Medical Language Engine reads both straightforward and complex physician notes and handles multiple values, irregular formatting, or embedded tables, all while keeping context intact. The solution separates current findings from historical or family mentions, preserves temporal context for measurement comparisons, and extracts several measurement types from a single sentence.

Each measurement includes a confidence score by therapeutic area, with deduplication to prevent double-counting and clinician spot-checks to verify accuracy and usefulness.

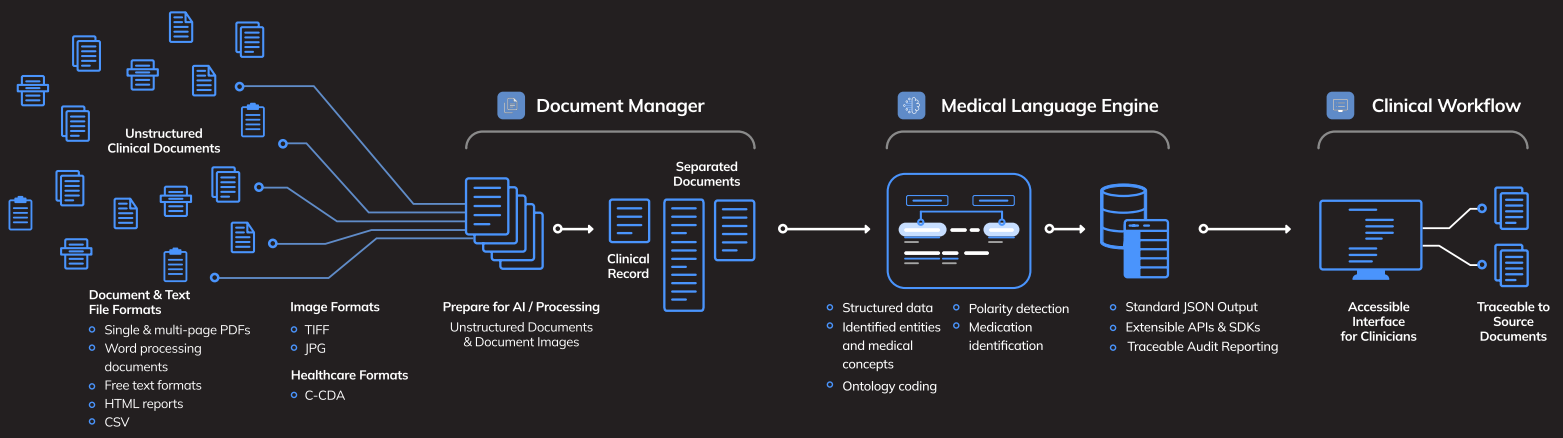
Therapeutic Area	Example Clinical Parameters	Parameters Extracted	Average Confidence
Oncology	TNM staging, ECOG, Karnofsky, tumor biomarkers, mutation results, treatment response	144.8M	0.98
Gastroenterology	IBD activity indices, liver fibrosis/staging, endoscopic severity, nutritional indicators	58.7M	0.99
Circulatory	LVEF, echocardiographic parameters, NYHA class, CHADS2, CHA_2DS_2-VASc	42.1M	0.99
Respiratory	FEV1, FVC, FEV1/FVC (pre/post), FeNO, CAT/ACT, GOLD	7.5M	0.94
Neurology	MMSE, MoCA, UPDRS, MSFC, headache impact/disability	1.98M	0.98
Autoimmune	DAS28, RAPID3, CDAI, SLEDAI, joint counts, inflammatory markers	0.961M	0.88

The Result: significant measurement extraction across therapeutic areas and consistently high confidence — evidence that the extractions are both comprehensive and reliable.

Processing of clinical notes occurred in emtelligent's US-based Enhanced Cleanroom environment using emtelligent's HITRUST-certified medical language engine, with weekly data deliveries to the client via secure file transfer.

emtelligent®

AI-powered solutions that transform unstructured data into actionable insights at enterprise scale.



Contact Us

Kim Perry
Chief Growth Officer

emtelligent.com

312.307.8402

kim@emtelligent.com

emtelligent®

About emtelligent

Based in Vancouver, British Columbia, emtelligent helps healthcare professionals work more efficiently and improve patient care.

Built by medical experts for medical experts, its Medical Language Engine and suite of clinical and search tools automate the extraction and analysis of data across systems, specialties, and populations, turning complex, unstructured medical text into actionable insights.

Through partnerships with health networks, imaging facilities, research institutions, payer organizations, and technology innovators, emtelligent increases safety, boosts operational efficiency, and elevates care quality across North America.