ChatGPT

/docs/product-strategy.md

Executive Summary & Positioning

Billigent is positioned as an **AI-powered Clinical Documentation Integrity (CDI) platform** that unifies documentation improvement, denial prevention, and automated appeals in one solution. In both inpatient and outpatient settings, Billigent aims to ensure that *patient care is accurately documented and appropriately reimbursed*. Hospitals today leave millions of dollars on the table due to documentation gaps [1] [2], while facing high denial rates and administrative burdens. Billigent's differentiated approach leverages advanced AI (including large language models and clinical NLP) to **analyze 100% of patient records** and surface missed documentation opportunities in real time. Unlike siloed incumbents, Billigent provides a *single cockpit* for CDI specialists, revenue cycle teams, and clinicians to proactively address documentation issues **before billing**, reducing downstream denials and revenue leakage. The platform is **HIPAA-compliant by design**, integrating seamlessly with EHRs via FHIR APIs and emphasizing transparency and clinician trust. Billigent's positioning centers on delivering **measurable financial ROI** (through improved case mix index, captured complications, and prevented denials) while also enhancing clinical outcomes (by accurately reflecting patient severity and quality metrics) [3] [4]. By targeting general inpatient medicine and high-impact specialties like cardiology, critical care, and surgery (where documentation gaps most often translate to significant revenue impact [5] [6]), Billigent establishes a strong wedge: **start where the financial and clinical stakes of documentation are highest, then expand enterprise-wide**.

Market Opportunity & Trends

Healthcare providers are under growing pressure to improve documentation for both reimbursement and quality reporting. The U.S. CDI software and services market is already **multi-billion dollar** in scale and growing at roughly **8% CAGR** [7], driven by several converging trends:

- **Rising Denial Rates and Revenue Leakage:** Hospitals are facing increasing claim denials – initial denial rates have climbed to ~12% of claims (up from 10% two years prior) [8]. Over **30% of denials are unequivocally avoidable**, yet 65% of denied claims are never resubmitted, representing lost revenue [9] [10]. Missed documentation (e.g. not documenting a comorbidity or complication) directly contributes to these avoidable denials and underpayments. For example, a case of heart failure without documented complications yields ~$3,971, whereas capturing a related comorbidity raises reimbursement to ~$5,769, and a major complication to ~$8,866 [11] – a 2-3× payment difference for the same clinical scenario. In sum, incomplete documentation can skew hospital finances and quality metrics (e.g. overstating mortality rates when comorbidities aren't documented [12] [13]).

- **Regulatory Complexity (ICD-10, DRGs, HCCs):** The shift to **ICD-10** brought far greater specificity requirements, making documentation more complex and increasing the risk of omissions [14]. Medicare's payment models (MS-DRGs for inpatient, HCC risk adjustment for Medicare Advantage) mean that **every diagnosis must be captured annually or during the encounter** to count for reimbursement [15] [16]. CMS and coding guidelines explicitly tie payment to documentation

specificity and completeness [17] [18]. As Medicare Advantage enrollment exceeds 50% of beneficiaries [19], outpatient documentation of HCCs (chronic conditions) each year has become essential. These compliance frameworks (ICD-10, DRG, HCC, plus HIPAA for privacy and CMS rules) create external pressure on providers to adopt CDI tools to avoid noncompliance and revenue loss [20] [21].

- **Pervasive EHR Adoption & Data Overload:** With **96% of hospitals using EHRs** as of 2024 [22], clinicians are documenting more data than ever. However, the sheer volume and complexity of electronic records lead to "data gaps" – 72% of hospitals report gaps in patient data even with EHRs, contributing to errors and missed information [23] [24]. The industry recognizes that manual review by CDI staff cannot scale to cover every chart or find every issue [25] [26]. This makes a strong case for AI automation: *NLP and machine learning can mine the entirety of patient records (notes, labs, radiology, etc.) to identify missing diagnoses or documentation conflicts that humans might overlook* [27] [28]. The digital data explosion, combined with advances in AI, has set the stage for "next-generation" CDI solutions that go beyond retrospective human queries.

- **AI Maturity and Provider Readiness:** Until recently, CDI programs were largely manual or rules-based. Now, modern AI models (including large language models) have demonstrated the ability to interpret unstructured clinical notes and suggest documentation improvements. Early AI-CDI entrants like Iodine and SmarterDx have validated the market's appetite: Iodine's AI-powered CDI is deployed in over 800 hospitals [29], and **100%** of surveyed Iodine customers would purchase their AI CDI again [30]. SmarterDx, a newer AI CDI startup, has rapidly grown by guaranteeing ROI (offering a 5:1 return on investment and contingency pricing) and delivering ~$2M in new revenue per 10k discharges on average [31] [32]. This traction indicates that healthcare providers are increasingly willing to adopt AI solutions that demonstrably boost financial performance. Moreover, providers are demanding **explainability** and workflow-friendly designs in these tools to trust them – e.g. Iodine emphasizes evidence-backed suggestions and has achieved a 96% physician query response rate within 48 hours in one case [33] [34]. In summary, the market is ripe for an AI-driven platform that can integrate into clinical workflows and deliver quick wins.

Given these trends, the **market opportunity** for Billigent is significant. U.S. hospitals collectively lose tens of **billions** annually to poor documentation and denials. The U.S. CDI market alone is projected at roughly $2.5–3 billion in 2024 and growing towards ~$4 billion by 2030 [7] [35]. Beyond hospitals, outpatient clinics and surgery centers represent a growing segment for CDI as value-based care and risk-adjusted payments expand. Billigent can capture this opportunity by addressing the pain points with a unified solution: preventing denials up front, improving documentation completeness in real time, and streamlining the entire mid-revenue cycle.

Differentiation & Competitive Advantage

The competitive landscape is fragmented, with incumbents and recent entrants each tackling pieces of the documentation problem. Billigent's key differentiator is its **end-to-end approach**: it combines capabilities that competitors typically offer only in silos. Specifically, Billigent is the only solution that fully integrates:

- **Proactive CDI + Coding Improvement:** Like Iodine and 3M, Billigent uses AI to identify missing documentation (e.g. indications of diagnoses in notes without corresponding codes) and prioritize cases with the highest impact. But Billigent goes further by providing a *conversational AI assistant* for

CDI that not only flags issues but also helps craft compliant physician queries with cited evidence. The AI suggests specificity improvements in real time (akin to Nuance's CAPD) **while also learning from retrospective patterns** to get smarter over time.

- **Denial Prevention Analytics:** Billigent embeds robust denial analytics and early warning indicators into the CDI workflow. This is an area dominated by players like Optum and FinThrive; Billigent differentiates by marrying denial pattern data with clinical documentation insights. For example, Billigent's dashboard highlights when front-end processes (registration, authorization) are causing recurrent denials [8] [36], so that providers can fix root causes. By integrating these analytics, Billigent positions CDI not just as a coding function but as part of a broader revenue integrity strategy (preventing errors that lead to downstream denials).

- **Automated Appeals & Feedback Loop:** Unlike any existing CDI vendor, Billigent features **AI-generated appeal letters** and an evidence repository for denials management. SmarterDx recently introduced a similar concept (auto-generating lower-dollar claim appeals in minutes) [37] [38], highlighting a whitespace. Billigent's approach uses retrieval-augmented generation (RAG) to compile pertinent documentation and regulatory references to draft a compelling appeal letter for denied claims, which a human can finalize in a fraction of the usual time. This closed-loop system – feeding insights from denials back into documentation improvement – is a unique strength. It means Billigent not only helps capture revenue on the first pass but also recovers revenue that slips through, while learning from those misses.

- **Unified Platform & Data Network:** Billigent's architecture enables *cross-domain intelligence* that single-purpose tools lack. For example, trends from appeals (such as certain diagnoses frequently down-coded by payers) can prompt Billigent's CDI module to flag similar cases pre-bill. The platform's unified data lake and knowledge graph link clinical facts ↔ billing codes ↔ payer rules, something competitors have not fully achieved. Over time, this integrated dataset becomes a network effect: Billigent can potentially learn payer-specific denial patterns across customers (in a de-identified way) to strengthen its AI models – creating a **moat via collective learning**. Incumbents like 3M have large customer bases, but their legacy systems are not built to leverage cloud AI at scale across clients for such network insights.

- **Clinical Depth in High-Impact Specialties:** Billigent is initially focusing on specialties like cardiology, critical care, and surgery, customizing its AI models to those domains. This provides deeper value out-of-the-box in areas where documentation nuances are complex. For instance, Billigent's cardiology module can recognize when an echocardiogram report indicates heart failure with preserved EF but the note doesn't specify acute vs. chronic – prompting a clarifying query. Or in surgery, Billigent can detect mention of post-op complications in nursing notes that aren't coded. By excelling in these high-impact scenarios, Billigent establishes credibility and a "wedge" into organizations, differentiating from one-size-fits-all solutions. Over 12% improvement in surgical cardiology CC/MCC capture has been demonstrated with focused CDI interventions [5] [6], underlining the value of specialty-aware AI. Billigent doubles down on that principle.

Finally, Billigent differentiates on **user experience and trust**. Our UX is designed with clinicians in mind: it offers clear, evidence-backed suggestions (with citations to sources in the patient chart or guidelines), and it operates within clinicians' existing workflow (e.g. EHR integration or single sign-on). Competing AI solutions often face physician pushback due to "alert fatigue" or black-box recommendations. Billigent's transparent

and collaborative approach ("AI copilot" rather than clippy auditor) will engender higher physician adoption and compliance with queries. This human-centered design, combined with a strong emphasis on compliance (AHIMA query guidelines, HIPAA safeguards built-in), positions Billigent as a **trusted partner** to hospitals, not just another software vendor. In summary, while competitors address fragments of the CDI and denials problem, Billigent's unified, AI-driven platform is differentiated in its breadth, technical innovation, and alignment with clinical workflows – a combination poised to redefine mid-revenue cycle management.

Market Sizing & Growth Strategy

Billigent's initial target market is the ~5,200 hospitals in the U.S., especially the ~3,000+ hospitals with established CDI programs (nearly 95% of large hospitals have some form of CDI today). This includes academic medical centers and large health systems where complexity and volumes are highest. We also target integrated delivery networks that encompass outpatient clinics and physician groups, given the expanding emphasis on outpatient CDI for HCC capture [16]. The serviceable obtainable market in year one are hospitals using legacy CDI software (3M, Nuance) that are seeking AI enhancements, as well as those who have not invested in advanced CDI tools yet (still manual) – together representing a multi-billion dollar replacement/new market opportunity.

Our go-to-market wedge is to win CDI leadership and CFO support by demonstrating **tangible ROI in under 12 months**. Case studies like Iodine's show ~$15 million incremental revenue in year one for a health system using AI CDI [39] [33]. Billigent aims to deliver similar or better outcomes, by not only increasing CC/MCC capture (thus raising case mix index) but also preventing costly denials and overturning those that do occur. We plan to offer outcome-based pricing (e.g. a performance guarantee or shared-savings model) to reduce adoption risk for customers – a strategy proven by SmarterDx's rapid growth with contingency pricing [40] [41].

In terms of growth, after initial footholds in inpatient CDI, Billigent will expand modules for **outpatient clinics and specialty workflows**, leveraging our platform's flexibility. With Medicare and commercial payers pushing more care (and risk) to outpatient settings, the need to capture chronic conditions and comorbidities in documentation is growing. Billigent's long-term vision is to be the **enterprise clinical intelligence layer** that sits atop the EHR: continuously analyzing documentation for completeness, compliance, and optimization across all care settings. This positions us in a category beyond traditional CDI – what we term *Clinical Revenue Integrity*. The total addressable market thus extends to ambulatory surgical centers, post-acute care (documentation impacts CMGs in rehab, for example), and even provider groups concerned with risk adjustment coding.

By articulating this broad vision but executing narrowly at first (in high-impact use cases and specialties), Billigent will capture significant market share. The competitive landscape is active – e.g. 3M and Optum have large footprints, and startups like SmarterDx have momentum – but no single player yet solves the full spectrum of needs with AI. By the time incumbents modernize or point-solution startups diversify, Billigent can establish itself as the **new standard** for AI-driven documentation and revenue integrity. We will measure our success not just in software license growth, but in the aggregate **financial impact delivered to clients** – e.g. dollars of additional reimbursement captured, denial dollars avoided, reduction in days to payment, etc. This impact-centric approach will fuel strong word-of-mouth and referenceability in the provider community, further accelerating adoption.

Sources (Product Strategy)

[^1]: Change Healthcare *2022 Revenue Cycle Denials Index*. (Primary research on denials: initial denial rate ~12% in H1 2022 vs 10% in 2020; ~17% regional peak; ~31% of denials deemed avoidable, 43% of avoidable denials were non-recoverable) [8] [9].

[^2]: CAQH CORE *Health Care Claims Issue Brief* (2023). (Industry brief highlighting claims data standardization gaps; notes Medicare Advantage risk adjustment requires capturing HCCs annually) [15] [19].

[^3]: HFMA *Claim Integrity Task Force Standard Metrics* (2022). (Defines six key denial KPIs: Initial Denial Rate, Denial Write-off %, Time to Appeal, Time to Resolution, Overturn Rate, etc., providing benchmarks for revenue cycle performance) [42] [43].

[^4]: HFMA *MAP Keys – Revenue Cycle* (Accessed 2023). (Benchmarking program that puts CDI metrics in context with broader revenue cycle KPIs like Clean Claim Rate, DNFB, A/R days. Emphasizes holistic visibility from documentation through appeals) [9] [10].

[^5]: **HIPAA Security Rule §164.312 (Technical Safeguards).** (Federal regulation requiring access controls (unique user IDs, role-based access), audit controls (activity logs), integrity protection, authentication, and transmission security (e.g. encryption) for ePHI) [44] [45]. These mandates influence Billigent's design for audit trails and data security.

[^6]: CMS & CDC *ICD-10-CM Official Guidelines for Coding and Reporting, 2025*. (Coding guidelines emphasizing documentation specificity. E.g. selection of principal diagnosis and secondary diagnoses must be supported by the provider's documentation. Drives DRG assignment and hospital Case Mix Index [42].)

[^7]: AHIMA/ACDIS *Guidelines for Compliant Physician Query Practice* (2022). (Defines non-leading query standards, documentation of query interactions, and the importance of maintaining query audit logs [46] [47]. Billigent's query workflow adheres to these best practices to ensure compliance and physician trust.)

[^8]: Osborne, T. "Measuring the Cost of Denials." *RevCycleIntelligence/OS Healthcare* (2021). (Study citing average **cost per denial ~$118** and that **~65% of denied claims are never reworked** [9] [10]. Illustrates the financial losses from inefficient denial management, reinforcing the need for prevention and automation.)

[^9]: MarkNtel Advisors. *US Clinical Documentation Improvement Market Outlook 2025-2030* (Mar 2025). (Market research report estimating the U.S. CDI market at **$2.52B in 2024**, projected to reach ~$3.95B by 2030 (~7.9% CAGR) [7] [35]. Attributes growth to rising chronic disease burden, regulatory requirements, and AI adoption in CDI [48] [20].)

[^10]: Health Catalyst. **"Data-Driven CDI Program Increases Revenue…"** – Allina Health case study (2018). (Case study where a health system used analytics to improve documentation: achieved **12.1% increase in surgical cardiology CC/MCC capture** and 6.3% in medical cardiology, translating to "millions in additional reimbursement" [5] [6]. Also provides an example of HF payment differences: heart failure with MCC pays >2× vs without CC [11].)

[^11]: Iodine Software – *Client Impact Interview* (HMA AI Catalyst, 2025). (Reported **$15M additional revenue in one year** at a health system using Iodine's AwareCDI AI, by increasing query rate from 20% to 37% of cases and speeding physician responses [39] [33]. Demonstrates the potential ROI from AI-driven CDI and faster, targeted queries.)

[^12]: **KLAS Research – 2023 CDI Software Performance Report**. (Independent industry survey: Iodine AwareCDI ranked #1 Best in KLAS for CDI in 2022 and 2023 with a score of 90.0 [49] [50]. Notably, 100% of Iodine customers interviewed would buy again [30], and AwareCDI helped >800 hospitals achieve a median 134% productivity lift in CDI workflow, resulting in **>$1.5B additional appropriate reimbursement annually** across those providers [29].)

[^13]: Transformation Capital. **"How SmarterDx Became a Generational Healthcare AI Company"** (May 2025). (VC perspective on SmarterDx: highlights **5:1 ROI guarantee** and contingency pricing (no upfront

cost) [31] [51]. Notes SmarterDx's flagship "PreBill" AI reviews *100% of discharges*, finding on average **$2M new net revenue per 10k discharges** [32]. By Q1 2025, over 50 health systems had adopted it, collectively uncovering "hundreds of millions" in revenue that would've been missed [52] [53]. Illustrates market validation for Billigent's all-charts, outcomes-based approach.)
[^14]: SmarterDx – *Hospitalogy feature* (Feb 2025). (Details how the startup's AI identifies **missed or under-coded diagnoses** to improve both revenue and quality metrics [54] [4]. Emphasizes that it focuses on **accuracy, not upcoding**, even recommending down-codes if appropriate [55]. Also launched **SmarterDenials** to auto-generate appeal letters in minutes, making appealing low-value denials cost-effective [56] [38]. This affirms the viability of Billigent's appeals automation and accuracy-first philosophy.)
[^15]: Nuance Communications Press (Oct 2019). *"Dragon Medical Advisor Clinical Guidance"*. (Nuance's CAPD solution is used by **500,000+ clinicians in 90% of U.S. hospitals** [57], providing in-workflow documentation advice. Proven to significantly **reduce retrospective queries** by delivering real-time specificity prompts to physicians [58] [59]. Validates physician-facing guidance as a way to improve documentation completeness at the point of care, a concept Billigent incorporates via its CAPD component.)

/docs/competitive-analysis.md

1. Landscape Overview

The clinical documentation and mid-revenue-cycle technology landscape consists of several segments, each with distinct players. Broadly, we can categorize competitors as follows [60]:

- **CDI Platform Leaders:** Full-suite inpatient CDI workflow solutions often integrated with coding. *Representative:* **3M** (360 Encompass) and **Dolbey** (Fusion). These incumbents offer end-to-end CDI and computer-assisted coding (CAC) in one platform, with large install bases in hospitals. Strengths include mature coding engines and integration; weaknesses are legacy interfaces and slower innovation in AI.

- **Predictive AI CDI Specialists:** Solutions focused on AI-driven identification of undocumented diagnoses and smart case prioritization. *Representative:* **Iodine Software** (AwareCDI). These emerged in the last decade to augment CDI teams with machine learning. Strengths: high accuracy in finding gaps, proven ROI; limitations: narrower scope (typically not addressing denials or physician workflow directly).

- **Physician Documentation (CAPD) Tools:** AI solutions embedded in the physician's workflow to ensure documentation completeness in real-time. *Representative:* **Nuance** (Dragon Medical Advisor, DAX). These aim to prevent documentation issues at the source by prompting clinicians. Strengths: deeply integrated with EHR, reduce need for retro queries [58] [59]; weaknesses: limited to what can be resolved by point-of-care prompts, no broader rev cycle analytics.

- **Denials Analytics & Prevention:** Platforms analyzing denial trends, providing rules-based edits, and front-end revenue cycle controls. *Representative:* **Optum** (Revenue Integrity/Claims Manager), **FinThrive** (Denial Management). Often tied to clearinghouses or large RCM vendors, they focus on root cause analysis of denials and workflow tools to prevent/manage denials. Strengths: enterprise-

scale data, strong payer rule libraries; weaknesses: typically do not improve clinical documentation or engage physicians (they address coding/billing errors and authorization issues).

- **Appeals Services & Automation:** Vendors (often service-oriented) that assist with managing denied claims and writing appeal letters. *Representative:* **CorroHealth** (appeals outsourcing, tech-enabled services), **Revecore** (complex claims). These aren't pure software plays – they combine workflow tools with expert services. Strength: deep expertise in complex cases; weakness: not scalable AI products, and usually post-denial reactive focus.

- **Revenue Integrity & Underpayment Recovery:** Adjacent players that identify missing revenue outside of documentation (e.g., contract underpayments). *Representative:* **Cloudmed** (now part of R1 RCM). Overlaps with denial prevention but more about payment variances. Not direct CDI competitors but competing for hospital budget/attention in revenue recovery.

Billigent spans multiple categories – it's an AI-driven CDI platform that also covers denial prevention and appeals automation – so our competitive analysis must consider all these angles. Below, we focus on the key competitors that Billigent is most likely to encounter in deals, especially **3M, Optum, Iodine, Nuance, and SmarterDX**, with additional context on others.

2. Competitor Feature Matrix

To compare capabilities, we assess competitors across core dimensions: **Denial Prevention Analytics**, **CDI Query Workflow (including AI assistance)**, **Appeal Automation**, **Analytics Depth**, and **AI Explainability/ Transparency**. We also note each competitor's notable differentiators and gaps. (Capabilities marked as "full", "partial", or "none" indicate the presence/maturity of that functionality in the competitor's offering.)

Vendor	Denial Prevention	CDI Query (AI)	Appeal Automation	Analytics Depth	AI Explainability	Notable Differenti
3M (360 Encompass, M*Modal)	**None** – minimal proactive denial tools (focus on coding/CDI)	**Full** – robust CDI workflow with NLP (auto-suggest queries, CAPD nudges)	**None** – no automated appeals (manual workflow)	**High** – extensive coding & CDI analytics, quality indicators [61] [62]	**Low** – AI suggestions are rule-based, limited transparency to end-users	*Unified co CDI platfo* market le in CAC; CA integratio physicians M*Modal

Vendor	Denial Prevention	CDI Query (AI)	Appeal Automation	Analytics Depth	AI Explainability	Notable Differenti
Nuance (Dragon Medical/CDI)	**None** – does not target denials (pairs with others)	**Partial** – offers CAPD for physicians (Dragon Medical Advisor) to prompt for specifics, and a physician-facing query workflow (Nuance Clintegrity CDI)	**None** – no appeals automation	**Moderate** – moderate analytics, mostly documentation quality metrics (O:E LOS, CMI, etc.) [65]	**Low** – AI is a black-box to users, minimal explanation beyond highlighting text [66]	*Physician-f "ambient" documenta capture* (D and real-ti advice; str speech recognitio huge user (500k+ clinicians)
Iodine (AwareCDI Suite)	**None** – no front-end denial focus (partners with R1 for services)	**Full** – AI-driven CDI case prioritization and query suggestion (concurrent review AI)	**None** – does not address appeals	**High** – strong analytics on CDI performance; predictive insights on documentation impact	**Low** – AI recommendations are not deeply transparent to users (scores given, but "black box" ML)	*Proven AI a scale*: 800- hospitals, in KLAS CD solution 2022-23; **predictive prioritiza** of cases; significant productivi gains (2×+ output) [29]
SmarterDX (Smarter PreBill)	**Partial** – identifies documentation issues that could cause payer denials (and quality metric impact) but not full pre-service denial checks	**Full** – AI "second-pass" review of 100% charts for missing diagnoses; flags to CDI team; also includes HCC outpatient focus	**Partial** – new *SmarterDenials* module auto-drafts appeal letters for certain denial types [56]	**Moderate** – good dashboards for revenue and quality impact, but narrower scope (focus on documentation misses)	**Medium – high evidence attribution** (shows exactly what data point triggered a suggestion), building physician trust [28]	*ROI-guara model* (continger pricing, 5: ROI) [31]; analyzes **1 of discha** with AI; fa deployme (cloud-bas minimal IT effort) [68]

Vendor	Denial Prevention	CDI Query (AI)	Appeal Automation	Analytics Depth	AI Explainability	Notable Differenti
Optum (RCM Analytics)	**Full** – comprehensive denial management and prevention tools (eligibility checks, edits, root-cause analysis)	**None** – no CDI query product (Optum focuses on coding edits, not documentation advice)	**None** – no automated appeals (services available separately)	**High** – enterprise-grade analytics, benchmarking across huge data sets (payer and provider data)	**Low** – traditional BI and rules, little AI transparency needed (not an AI CDI tool)	*Scale & pa insight*: pa UnitedHea vast claim dataset; integrated front-end and A/R workflow
FinThrive (nThrive)	**Full** – end-to-end denial prevention and claim scrubbing solutions	**None** – no CDI query offering (FinThrive more focused on billing/ claims)	**None** – no, aside from workflow to track appeals	**High** – strong analytics, especially around claim integrity and denials	**Low** – not focused on AI explanations (rule-based engine)	*Revenue cy breadth*: offerings f scheduling billing; de prevention expertise rules, etc.)
Cloudmed (R1)	**Full** – specializes in identifying underpayments and preventable denials, using analytics and expert audits	**None** – not a CDI tool	**None** – focuses on underpayment recovery, not documentation or appeals	**High** – deep analytics on revenue leakage (contract compliance, etc.)	**Low** – primarily services + analytics, no AI UX	*Underpayr focus* and complex c expertise; compleme CDI by recovering revenue outside document
CorroHealth (Versalus, Chartwise)	**None** (offers denial/appeal services rather than preventive tech)	**Partial** – has a traditional CDI software (acquired ChartWise), but not AI-driven; offers physician query workflow and some NLP	**None** (appeals handled as a service by staff, not automated by software)	**Moderate** – reports on query rates, DRG outcomes; service-heavy	**Low** – limited AI in product; mostly human expertise	*Blended se + tech*: can provide outsource specialists denial exp one-stop s for HIM services

Vendor	Denial Prevention	CDI Query (AI)	Appeal Automation	Analytics Depth	AI Explainability	Notable Differenti
Dolbey (Fusion CAC/ CDI)	**None** – no dedicated denial prevention (integrates with billing edits maybe)	**Full** – integrated CDI with CAC; NLP-assisted coding and some query support	**None** – no appeals functionality	**Moderate** – solid coding analytics, some documentation metrics	**Low** – legacy NLP, not focused on explainability	*Combined CDI workfl* well-know speech recognitio coding too stable and effective

(Table sources: vendor product literature and industry reports as cited. "None/Partial/Full" designations are based on available product capabilities in 2024-2025.)

Key Takeaways from Matrix

- **No Single Competitor Does It All:** As seen above, **no competitor currently offers full denial prevention + CDI + appeals automation together**. 3M comes closest on CDI + coding integration, Optum/FinThrive on denials, SmarterDx on CDI + some appeals, but each has clear gaps. This validates Billigent's strategy to be the first comprehensive solution.

- **Strengths to Emulate, Gaps to Exploit:** 3M's dominance comes from workflow integration (coders and CDI in one platform) – Billigent will compete by similarly offering a seamless experience, but with modern AI. Iodine's success shows the importance of demonstrable ROI and accuracy; Billigent must match their AI performance while highlighting our broader scope. Nuance's ubiquity with physicians underscores that clinician adoption is crucial – Billigent's physician-facing components must be intuitive and valuable to gain buy-in where Nuance is entrenched. **SmarterDX and Iodine confirm the demand for AI; 3M and Optum confirm the need for end-to-end solutions**. Billigent is effectively positioned at the intersection.

- **Emerging Entrants (SmarterDx) vs Incumbents:** SmarterDx's growth (backed by physicians, ROI guarantee) signals how new entrants can rapidly win hospital customers by addressing pain with low friction [31] [68]. However, SmarterDx's narrow focus (chart review for missed diagnoses, and very recent expansion to appeals) leaves a gap in proactive prevention and deeper CDI workflow integration. Incumbents like 3M and Optum have broad reach but are slower-moving and not AI-focused. This dynamic gives Billigent a *timing advantage*: by the time incumbents add similar AI or startups try to broaden out, Billigent can establish itself as the go-to integrated solution.

3. Competitive Profiles

3M Health Information Systems (360 Encompass & M*Modal)

Market Position: 3M HIS is the long-standing market leader in coding and CDI software in hospitals. Their 360 Encompass platform is widely used for computer-assisted coding (CAC) and CDI, often bundled with 3M

consulting services. They also acquired M*Modal's speech recognition/CAPD, which extends their reach into physician documentation (e.g. 3M™ CDI Engage One for real-time physician prompts).

Strengths:
- **Integration:** 3M offers a one-stop platform covering inpatient coding, quality metrics (PSIs, HACs), and CDI workflow [61] [74]. This integration means a coder and CDI specialist can work in the same system, see the same documentation and queries, and avoid siloed processes. Hospitals appreciate fewer interfaces and unified reporting.
- **NLP & Rules Engine:** 3M's platform uses an NLP engine (honed from years of coding suggestions) to auto-suggest diagnoses and even a "working DRG" for each case [75] [62]. It can flag if a case has clinical indicators for a condition without a diagnosis (their version of a CDI alert) [64] [76]. It also continuously updates with new coding rules, ensuring compliance with the latest ICD-10 changes.
- **Quality Focus:** 3M uniquely embeds quality metrics checks (e.g. it will flag potential Patient Safety Indicators or hospital-acquired conditions in documentation) [61]. This helps CDI prioritize documentation that affects value-based purchasing and publicly reported outcomes – a differentiator 3M leverages as hospitals increasingly care about documentation for quality, not just reimbursement.
- **Scale and Trust:** 3M's products are used in thousands of hospitals; it's a known quantity with decades of experience. IT departments and HIM departments trust 3M to deliver reliable software and stay current with regulations.

Weaknesses:
- **Lack of Upstream/Downstream Reach:** 3M's focus is squarely on coding and CDI. They do not provide denial prevention analytics on front-end processes (they assume if coding and documentation are correct, denials go down, but they don't analyze reg/auth errors, etc.). They also don't assist with denial appeals beyond providing data for a human to use. This leaves a *gap that Billigent fills* by covering the full cycle.
- **Legacy Technology & UX:** While effective, 3M's software has a reputation for an outdated interface and being somewhat clunky. New AI-based competitors often showcase more modern, user-friendly UIs. Additionally, 3M's NLP, while competent, is largely rule-based with limited machine learning, meaning less adaptability to new patterns compared to Billigent's AI.
- **Physician Adoption:** 3M's CAPD via M*Modal does put nudges in front of physicians, but adoption and satisfaction of these nudges vary. Some clinicians find them intrusive or not clinically sophisticated enough (e.g. basic reminders for unspecified diagnoses). Nuance's Dragon still dominates physician documentation space. If physicians ignore 3M's prompts, the benefit is lost.

Competitive Outlook: 3M will likely defend its turf by emphasizing the completeness of its platform and perhaps adding more AI features (they have started touting "advanced ML" prioritization [77]). For a customer considering Billigent, we must demonstrate that we can either integrate with 3M (coexist) or replace it with minimal disruption and significant upside. On coexistence: Billigent could feed its findings into 3M's worklist if needed or operate in parallel initially. On replacement: emphasize intuitive UI, cloud-based agility (vs 3M on-prem), and our ability to *reduce the total workload* (3M's system still relies on lots of manual CDI review, whereas Billigent automates much of it). **3M's greatest vulnerability is its lack of real AI and inability to address denials/appeals** – Billigent should highlight those gaps.

Optum (Optum360 / Enterprise CAC and Denial Solutions)

Market Position: Optum (the tech arm of UnitedHealth) offers a variety of revenue cycle tools. In CDI, they are not a leading player for software – they had an older product "Optum CDI 3D" (largely rules-based, not

widely adopted compared to 3M/Nuance/Iodine). Optum's strength is in claims analytics, edits (they own a popular claims edit engine), and denial management services. They often bundle solutions in large payer-provider contracts.

Strengths:
- **Denial Management & Data:** Optum publishes an annual **Denials Index** (absorbing Change Healthcare's index after acquisition) and thus has industry-leading data on denial trends [78]. Their solutions can pinpoint top denial reasons and even predict which claims are likely to be denied using historical data [72]. This insight helps hospitals focus on process fixes (e.g. if a hospital sees many registration errors denials, Optum tools highlight that).
- **Front-End Integration:** Optum's solutions intercept issues early – e.g. insurance eligibility checks, authorization compliance, and coding edits before claims go out. They reduce denials by catching mistakes prior to submission. This is complementary to CDI (which focuses on clinical content); Optum shines on the administrative side of prevention.
- **Scale and Payer Perspective:** Being part of UnitedHealth, Optum has unparalleled payer knowledge. They know what payers scrutinize, and they incorporate that into their products (for instance, analytics to identify patterns of down-coding or medical necessity denials by certain payers). Also, Optum can leverage its vast client network for benchmarking – offering hospitals comparisons to peers for denial rates and such.

Weaknesses:
- **No AI CDI Capability:** Optum's CDI offering (if any currently) is minimal. Clients using Optum for denials often use another vendor for CDI. Optum hasn't invested in the kind of AI-driven documentation analysis that Iodine or Billigent have. This is a glaring gap – having great denial analytics tells you *what went wrong*, but not necessarily *how to improve clinical documentation to fix it*. Billigent can partner with or replace Optum by covering that blind spot.
- **Physician Workflow Absent:** Optum tools are generally used by back-office rev cycle staff, not by physicians or even CDI nurses at the point of documentation. That means they don't directly solve physician documentation behavior issues. At best, they report "medical documentation requested" as a top denial reason, but they don't provide a physician query interface. This is where Billigent provides a solution – bridging clinical and financial workflows.
- **Complexity and Cost:** Optum's solutions, while powerful, can be complex to implement (enterprise software, often requiring integration with many hospital IT systems). They also tend to be expensive and come in large bundles. Smaller or mid-size organizations might find them less accessible. Billigent's cloud-native, modular approach can be more attractive in comparison (faster to deploy, pay-for-value).

Competitive Outlook: Optum will likely continue focusing on denial prevention and may partner for CDI rather than build it. In competitive situations, if a prospect already uses Optum for denials, Billigent can position as complementary: we feed improved documentation (thus cleaner claims) into their process, potentially reducing the volume of denials that Optum's system has to manage. Conversely, if replacing Optum's analytics, Billigent can showcase comparable denial analytics and emphasize synergy – one platform linking documentation errors to denial outcomes. Optum's brand is strong, but its siloed approach and lack of clinician-facing tools give Billigent a strong narrative.

Iodine Software (AwareCDI)

Market Position: Iodine is the pioneer of AI in CDI and the current upstart leader in CDI tech, often mentioned as the top innovative choice for hospital CDI departments. They have been **Best in KLAS** in CDI

software for two years [49], reflecting high customer satisfaction. Iodine's flagship AwareCDI uses machine learning to prioritize cases and identify "likely undocumented" diagnoses for CDI specialists to review.

Strengths:
- **AI Accuracy & ROI:** Iodine's ML models (trained on large datasets of clinical records and past queries) are very adept at finding cases with opportunity. For example, their clients report significant lifts in capture of major comorbid conditions (MCCs) and resulting revenue increases. One health system credited Iodine with a **$7M+ annual reimbursement uplift** after full deployment (from case studies). The transparency of ROI is a major selling point – Iodine can show improvements in metrics like CMI or query rate pre vs. post.
- **Prioritization Efficiency:** Instead of burdening CDI staff with reviewing every case, Iodine triages charts so specialists focus on those with the highest impact missing documentation. Customers have seen CDI productivity more than double (134% increase in cases reviewed per specialist) [29]. This efficiency gain is crucial given CDI staffing constraints. Essentially, Iodine helps hospitals "do more with the same staff" by ensuring they spend time on the right charts.
- **Client Loyalty:** Iodine's KLAS scores indicate excellent client support and product performance – 100% of interviewed customers said they would repurchase [30]. This suggests Iodine has achieved trust and delivered on promises, a high bar in the health IT space. They've rapidly expanded; as of early 2023 they serve over 800 hospitals [29], including many academic medical centers. Being a young company, their tech is modern (cloud-based, quick updates) which clients appreciate compared to older on-prem systems.

Weaknesses:
- **Limited Scope (No Denials or Appeals):** Iodine deliberately focused on the CDI workflow and did it extremely well. However, they do not address other parts of revenue integrity: no module for denial management, no physician-facing CAPD (they rely on CDI staff to generate queries via their interface). Thus, Iodine stops at the point of recommending a query; it doesn't ensure the physician responds or integrate that with denial outcomes beyond some reporting. Billigent can exploit this by offering an *integrated continuum*.
- **Physician Resistance if Overused:** Iodine's model can generate a lot of query recommendations. Some clients have noted that if all those suggestions were acted on, physician query volume could spike – potentially causing fatigue. Iodine has guidelines to mitigate this (and many suggestions might be bundled into one query), but it's a dynamic to manage. Billigent, by incorporating physician-centric design and perhaps more real-time engagement, might navigate physician relationships more smoothly.
- **Data and Explainability:** Iodine provides a confidence score for each case and highlights some evidence, but the inner workings of its predictions are not fully transparent. In the era of explainable AI, Billigent can differentiate by providing clearer rationale for each suggestion (e.g., "The patient's lactic acid is elevated and WBC >12; sepsis not documented – consider querying"). If Billigent's AI is more of a "copilot" that cites chapter and verse, it could win trust faster. Iodine is moving in this direction, but historically it was a bit of a black box ("this chart has a 80% probability of missing something").
- **Outpatient/HCC still nascent:** Iodine has an outpatient solution (likely leveraging similar tech for clinics and Medicare Advantage HCC capture), but it's not as widely adopted as their inpatient offering. This means for organizations looking to unify inpatient and outpatient CDI, Iodine is still proving itself outside the hospital setting. Billigent can aim for parity in both realms from the start, giving us an edge in integrated delivery networks.

Competitive Outlook: Against Iodine, Billigent must respect their incumbency in AI CDI. They will be our toughest head-to-head competition in many CDI department-led deals. However, Billigent's strategy should be to broaden the conversation: involve the CFO, involve the VP of Revenue Cycle, where our end-to-end

scope (denials and appeals) resonates more than Iodine's narrower CDI focus. If a prospect already has Iodine, Billigent could potentially complement (e.g., use Billigent for denials/appeals and maybe outpatient, leaving Iodine for inpatient for now). Over time, though, Billigent aims to displace by matching Iodine on core CDI excellence *and* offering more. Price-wise, Iodine is premium; Billigent could be competitive or justify a similar price by more value delivered. Iodine is also now partnering (e.g. with R1 RCM) for services – Billigent might partner with consulting firms to counterbalance. Importantly, **many Iodine users still rely on 3M or homegrown systems for coding** – Billigent can position as a platform to consolidate these functions. Expect Iodine to potentially add denial analytics or partner up; our advantage is being built from day one as integrated.

Nuance (Microsoft) – Dragon Medical & CAPD

Market Position: Nuance (now part of Microsoft) is not primarily a CDI software vendor in the traditional sense, but it's impossible to ignore because of its ubiquity in clinical documentation via **Dragon Medical One** (speech recognition) and the newer **Dragon Medical Advisor (DMA)** CAPD tool. Nuance also had legacy CDI software (Clintegrity, formerly J.A. Thomas consulting) but their strategy has shifted to front-end documentation improvement and ambient clinical intelligence (like DAX, which creates notes from conversations). Essentially, Nuance touches physicians and documentation creation; their goal is to "get it right the first time" so downstream CDI is less needed.

Strengths:
- **Physician Adoption and Workflow Integration:** Nuance's speech recognition is used by a huge share of providers. Over half a million clinicians use Dragon, and it's deeply integrated into EHRs. This gives Nuance a channel to deploy CAPD tips without extra logins or new interfaces – a big plus. Clinicians seeing inline advice while they dictate or type is powerful (if the advice is good). Also, Nuance's DAX ambient solution, while not directly CDI, shows their forward-looking approach to reduce physician burden while capturing details. A physician might trust a Nuance suggestion more as it's within a familiar tool.
- **Impact on Query Volume:** By providing instant feedback (e.g. "document the type of CHF" or "patient on insulin – document diabetes type"), Nuance's CAPD has shown to significantly reduce the need for retrospective queries. In client testimonials, hospitals reported "significantly fewer queries from CDI and coders" once DMA was live [58] [59]. This proactive approach can improve physician responsiveness and overall documentation quality metrics like Severity of Illness and Risk of Mortality scores.
- **Microsoft Backing and AI R&D:** Now under Microsoft, Nuance is likely to leverage Azure's AI advances. They have the resources to incorporate large language models, possibly offering more sophisticated, context-aware suggestions (e.g. as GPT-like models become available internally). Microsoft is also positioning Azure as a health data platform, and Nuance sits on that, which could lead to strong integration (e.g., Teams integration for clinical documentation, etc.).

Weaknesses:
- **Narrow Focus (No Full CDI Program Tool):** Nuance's solutions help physicians but don't provide a comprehensive CDI workflow for CDI specialists or detailed revenue cycle analytics. Many hospitals actually use Nuance CAPD in tandem with another CDI tool (like 3M or Iodine). Nuance doesn't prioritize the financial reporting side – their metrics are more clinician-focused (like documentation completeness, physician engagement stats) rather than direct reimbursement impact. Billigent can fill that gap by providing the backend analytics and accountability that Nuance lacks.
- **Alert Fatigue and Relevance:** CAPD alerts must be finely tuned – too many or irrelevant alerts and physicians will tune out. Nuance's knowledge base is broad, but some organizations have struggled with

high volumes of alerts that irritate doctors (for example, constant prompts for acuity or linking diagnoses). If not managed, this can cause backlash. Billigent's strategy of targeted, high-yield suggestions with context may fare better. Also, Nuance's rules might not catch complex scenarios that require synthesis (where an LLM might do better).
- **No Denial or Post-Visit Component:** Once the physician completes the note, Nuance's job is done. It doesn't tie into whether the claim was paid or denied. Thus it cannot learn from mistakes that slipped through, nor assist in recovering those. Billigent's closed-loop covering post-bill (denials) is a contrast. Additionally, Nuance doesn't address coder workflows (assigning codes, etc.) – it's very physician-centric. Billigent offers a more team-based approach (physicians + CDI + coders on one platform).

Competitive Outlook: Nuance (with Microsoft) is a potential competitor mainly in the sense that it could evolve into offering more AI for documentation. For example, Microsoft could integrate GPT-4 into Nuance to do advanced CDI tasks in real-time. However, any such solution would likely still focus on physician documentation stage. Billigent should consider integration with Nuance rather than head-on replacement when possible (e.g., we could take input from Nuance that a physician declined to specify something, and ensure a CDI follow-up occurs – a complementary workflow). In selling against Nuance's CAPD, emphasize Billigent's superior clinical understanding (if using a large model) and the advantage of having context from the entire record (Nuance's real-time prompts might not synthesize as much context). Also, emphasize **traceability**: Billigent can show exactly why it suggests something and track it to outcomes; Nuance's suggestions disappear after the note is signed, with less accountability. Since Microsoft is a giant, Billigent should integrate with Azure tech where beneficial (we already plan to use Azure OpenAI and Cognitive Services). If a prospect worries "Microsoft will do this," we point out that Billigent is already leveraging Microsoft's best AI (via Azure) but with our unique healthcare-specific corpus and design – plus we cover the revenue cycle end-to-end, which Microsoft/Nuance currently do not.

SmarterDx

Market Position: SmarterDx is a newer startup (founded 2020) that directly competes in the AI CDI space. They position themselves as an "AI second reviewer" of all charts prior to billing (hence the product name "PreBill"). They gained attention with strong ROI claims and a contingency pricing model (they only get paid when they find new revenue) [40] [41]. They've also begun to tackle automated appeals (with their "SmarterDenials" module) in late 2024. With backing by prominent VC firms and a story of physician-founders, they are a close analog to what Billigent aspires to do, albeit on a smaller scale so far.

Strengths:
- **Comprehensive Chart Review with Minimal Effort:** SmarterDx's value prop is that their AI reads **every element of the patient's record** – notes, labs, meds, etc. – and ensures the final coding "tells the full story." They highlight finding diagnoses that were missed by human reviewers, thereby capturing revenue that would have been lost [28] [55]. Because it's fully automated in scanning charts, hospitals can use it as a safety net without massively increasing CDI staff. Many clients run it on all discharges to catch anything their own team didn't.
- **Business Model & ROI Focus:** By offering a 5:1 guaranteed ROI and only charging a percentage of what they help recover, SmarterDx greatly lowers the barrier for hospitals to try it [31] [51]. This approach has been very persuasive, especially in times when hospital margins are tight – it's essentially a "no-risk" proposition that any CFO would consider. It also aligns vendor incentives with the client's outcomes, which builds trust. They tout on average $2 million net new revenue per 10k discharges [32], a figure that resonates strongly.

- **Appeals Automation (Innovation):** With SmarterDenials, they identified a niche others ignored: many low-dollar denials go unappealed because staff time is limited. By automating appeal letter writing (using AI to draft letters in minutes), they make it feasible to contest those denials [56]. Early reports from pilots show this can recoup revenue that hospitals usually write off. It's a clever extension and aligns with Billigent's vision that every denied dollar has a story that AI can help tell in appeals.
- **Physician-Founder Credibility:** The founders being hospitalist physicians who experienced documentation pain lends them credibility with clinical leadership. It signals that the product is built with understanding of clinical workflows, not just by tech people. They often emphasize "we're not upcoding, just getting documentation right" [55] to alleviate compliance fears, an important message that resonates because it's clinician-led.

Weaknesses:
- **Limited Track Record & Scale:** As of early 2025, SmarterDx has 50+ health system customers [79], which is impressive for its age but still a fraction of the market. They lack the installed base and long-term reference history that a 3M or even Iodine has. Billigent, though new as well, can capitalize on any early growing pains SmarterDx might have (e.g., if their algorithms need refinement or if their support infrastructure struggles as they scale). Essentially, SmarterDx is blazing the trail; Billigent can learn from their missteps or customer feedback.
- **Product Breadth vs. Depth:** SmarterDx's core is similar to Iodine's – AI chart analysis for missing diagnoses. They have now two products (PreBill and Denials). But they do not offer the broad workflow tools around those. For instance, they flag issues, but how the CDI team triages and fixes those still relies on maybe spreadsheets or existing systems. Billigent plans to incorporate a full task management and communication loop (worklists, query generation, tracking). In appeals, SmarterDx generates a letter, but likely doesn't interface deeply with hospital denials management systems or track appeal status end-to-end. Billigent can differentiate by offering more **workflow integration** – a more "system of record" for these processes, not just an AI suggestion engine.
- **Competition and Exit:** News in mid-2025 is that SmarterDx was acquired by a private equity firm to be merged with others under "Smarter Technologies" [80]. Such consolidation could either strengthen them or distract them. If integration with other products (maybe coding or analytics companies under the same umbrella) isn't smooth, it could slow their innovation. Billigent, being independent and laser-focused, can potentially out-innovate a SmarterDx that's busy merging. There's also the possibility that as part of a larger company, they might shift to more traditional pricing or lose the startup nimbleness that gave them edge.

Competitive Outlook: Billigent and SmarterDx will often be compared head-to-head by customers looking for an AI CDI solution. We have similar value propositions. To win, Billigent should highlight: (a) **Integrated platform** vs SmarterDx point solutions – Billigent is built from ground up to connect documentation to denials to appeals, whereas SmarterDx is adding modules (with possibly separate UIs). (b) **Real-time concurrent use** – SmarterDx is generally post-discharge (a "second pass"). Billigent can intervene **concurrently** (during the patient stay or before bill drop) as well as retrospectively. This means we can improve documentation in time for initial billing more often, not just retrospectively catch it (which might require rebill or addendum). Hospitals prefer issues be caught early (concurrent CDI) to avoid rework. If Billigent can do that, it's an edge. (c) **Technology** – Without disparaging, we can indicate our use of state-of-the-art language models that can read and reason in charts, possibly giving more nuanced suggestions than SmarterDx's algorithms (their marketing mentions "2,200 algorithms" [28] which suggests more rules-based heuristics). Billigent's AI might be more flexible, for example handling novel scenarios or combining data points (hallmark of large neural models). (d) **Customization and partnership** – as newcomers, both companies need to show willingness to adapt to client needs. Billigent can differentiate by offering more

user control (e.g., thresholds for alerts, custom rules, etc.) and a more open platform approach (maybe exposing our API or integration options more than SmarterDx). Price-wise, competing with a contingency model is tough; Billigent might consider similar risk-share to neutralize that advantage or at least provide a strong ROI guarantee.

In summary, SmarterDx validates the market and some of Billigent's approach. They are an important competitor to monitor, but there is room to outmaneuver by going broader and deeper in capabilities.

4. Billigent's Opportunity ("Wedge")

Analyzing the competition reveals clear **whitespace opportunities** for Billigent. These represent the "wedge" by which we can enter and expand in the market:

- **Unified Solution (Breaking Silos):** As noted, none of the competitors spans CDI, denials, and appeals together. Billigent's wedge is to be the first **unified platform** for the entire documentation quality loop [81] [82]. This is extremely appealing to hospitals that currently juggle multiple vendors or processes. By pitching "one tool to replace three," we wedge into budget discussions across departments (HIM, Case Management, Finance). For example, a hospital might have a CDI tool, a separate denial analytics tool, and outsource appeals – Billigent can consolidate those, saving cost and improving effectiveness through integration. This unity also builds a competitive moat: once Billigent is embedded as the central platform linking clinical documentation to financial outcomes, it's hard for point solutions to dislodge us.

- **AI Explainability & Trust:** All competitors use AI or algorithms to some extent, but virtually all are lacking in providing *transparent, explainable outputs*. Billigent is uniquely focusing on an **"AI explainability layer"** as a core feature – every recommendation or query our AI suggests will come with the evidence and reasoning clearly shown [82] [83]. This directly addresses a pain point: others produce scores or alerts but often without sufficient context, which leaves clinicians skeptical. By building trust through transparency, Billigent can wedge into organizations that were hesitant to adopt AI CDI due to "black box" concerns. For instance, if a hospital looked at Iodine but worried the AI wouldn't be accepted by physicians, Billigent can be offered as the more physician-friendly alternative because of our explainability and human-in-the-loop design (all AI outputs can be reviewed by a human, and we log and justify everything). In an RFP, this could be a deciding factor that sets us apart. **Trust is a wedge** – especially in healthcare, where a more explainable AI will be preferred over a marginally more accurate but opaque AI [84] [85].

- **Appeals Automation (Untapped ROI):** Billigent's inclusion of automated appeal drafting meets a largely untapped need. Most providers either ignore low-value denials or spend costly nurse time on high-value ones. By leveraging AI (using the same data we've gathered for CDI) to instantly generate appeals letters with citations to medical record and medical policy, we give hospitals a way to recover revenue they'd otherwise write off [81] [82]. This is a compelling wedge in discussions with CFOs: "Not only will we improve your documentation going forward, we'll also help you claw back money you're currently leaving on the table." SmarterDx is the only competitor starting to do this, but it's early – we can quickly establish leadership in AI appeals. Our differentiator might be handling *complex* appeal scenarios too (e.g. include relevant journal references or Medicare rules automatically). By being first-to-market in robust appeals AI, we gain a wedge into the denials management teams of hospitals, who historically have not been served by CDI vendors.

- **Cross-Department Value (Clinical + Financial Leadership Buy-in):** Many of our competitors pitch either to CDI managers (Iodine, 3M) or to CFOs (Optum, etc.), but rarely both effectively. Billigent's broad value proposition can engage both clinical documentation leaders and finance leaders in one narrative. For example, **Case 1**: We talk to the CDI director about how we'll make queries more efficient and accurate (appeals to quality of documentation) and simultaneously talk to the CFO about how we'll reduce denial write-offs and increase revenue capture (financial outcomes). Neither 3M nor Optum can check both boxes with one product. This dual-value wedge is powerful for getting consensus in buying committees. It also insulates us: even if a hospital initially just wants a CDI tool, we can later upsell the CFO on the denials module, etc. Conversely, we might land via the CFO's interest in reducing denials, then expand usage to the CDI team. The ability to start in one area and expand is a strategic wedge that pure-play tools don't have.

- **Modern Architecture & Speed:** From a technical standpoint, Billigent being cloud-native and API-friendly is a wedge against legacy incumbents when it comes to implementation speed and data integration. Hospitals frustrated with their older systems might be swayed by how quickly Billigent can connect to their EHR (FHIR APIs) and start delivering results – potentially within weeks, versus months for on-prem installs or lengthy data prep for others. Our use of up-to-date tech (containers, serverless scaling, etc.) can also mean better performance and frequent updates, which we should highlight if it comes up. While this is more of a secondary differentiator for non-IT stakeholders, it can win the support of the hospital CIO/CTO in a decision, tipping the scales in our favor when evaluating new vs. legacy solutions.

In conclusion, Billigent's competitive wedge lies in *combining the strengths of multiple solution types while eliminating their weaknesses*. We bring the **predictive AI smarts** of Iodine/SmarterDx, the **workflow integration** of 3M, the **physician-facing guidance** of Nuance, and the **denial expertise** of Optum – all in one. Each competitor we analyzed has a blind spot, and Billigent's strategy is to have **no significant blind spots**, or at least far fewer. This holistic approach, underpinned by cutting-edge AI with human-centric design, positions Billigent not just to compete, but to redefine the space. Our differentiation is not merely incremental improvement; it's a step-change: *from fragmented point solutions to an intelligent, learning system that continuously improves documentation and revenue integrity across the care continuum*. That is our wedge and ultimately our moat.

Sources (Competitive Analysis)

[^1]: 3M Health Information Systems – *360 Encompass System Overview* (2019). (Describes 3M's NLP-driven CDI worklist prioritization: flags cases lacking CC/MCC, cases with clinical indicators but no diagnosis, etc. [86] [64] . Emphasizes single platform for coding+CDI+quality and CAPD integration [87] [88] .)
[^2]: Nuance Communications – *Dragon Medical Advisor Press Release* (Oct 2019). (Nuance's CAPD provides in-workflow documentation guidance; used in 90% of US hospitals by 500k+ clinicians [89] . Shown to reduce retrospective queries by capturing comorbidities in real time [58] . Highlights physician quote praising improved documentation and fewer follow-up queries [67] [90] .)
[^3]: Iodine Software – *KLAS 2023 Award Release* (Feb 2023). (AwareCDI ranked #1 CDI software with score 90.0 [91] ; 100% of customers would buy again [92] . Over 800 hospitals use it, achieving median 134% productivity lift in CDI and >$1.5B total additional reimbursement annually [29] . Validates Iodine's strong performance and ROI.)
[^4]: SmarterDx – *Hospitalogy Article* (Feb 2025) by Blake Madden. (Explains SmarterDx's value prop: AI reviews every patient chart to capture missed diagnoses, improving revenue *and* quality metrics [4] [28] .

Notes SmarterDx guarantees 5:1 ROI and $2M new revenue per 10k discharges [93] [94]. Also details **SmarterDenials** module launching 2024 for AI-generated appeal letters in minutes [56].)
[^5]: Transformation Capital – *SmarterDx Profile* (May 2025). (PE investor write-up: SmarterDx's PreBill AI ingests ~30k data points per chart and finds missed revenue; contingency pricing with **5:1 ROI** promise [31] [51]. On average yields **$2M net new per 10k discharges** [32]. Over 50 health systems using it by 2025, with "hundreds of millions" in aggregate revenue uncovered [52]. Confirms strong results and adoption.)
[^6]: **Competitive Matrix Sources:** Internal analysis of vendor materials and industry reports (2024). Key sources include 3M 360 Encompass CDI Fact Sheet [63] [75], Nuance DAX/Advisor product info [95] [96], Optum denial management brochure [72], Iodine client case studies [33], SmarterDx case examples [70], FinThrive and Dolbey product descriptions. This synthesis informed the capabilities table, with assumptions cross-validated by user comments in KLAS and ACDIS forums (e.g., common knowledge that 3M and Dolbey lack appeals modules, etc.).
[^7]: KLAS Research – *"Clinical Documentation Improvement 2023"* (Aug 2023). (Market research showing Iodine leading in CDI software satisfaction, 3M and Nuance trailing in user-reported innovation. Noted pain points: many CDI vendors lack robust physician engagement and denial integration. Used for qualitative competitive context.)
[^8]: ACDIS/AHIMA – *Industry survey on CDI* (2022). (Provided insight that ~30% of hospitals planned to invest in AI for CDI in next 2 years, and biggest unmet needs were outpatient CDI and metrics linking CDI to denials. Supported identification of whitespace in outpatient and closed-loop analytics.)

/docs/prd.md

Product Name: *Billigent Clinical Intelligence Platform*
Version: 1.0 (MVP) – *In Development*
Date: August 14, 2025
Owner: Product Management (Billigent Inc.)

1. Market Needs & Pain Points

Healthcare organizations face critical needs at the intersection of clinical documentation, revenue capture, and compliance. Below we outline the key market needs Billigent addresses, each grounded in evidence and given a severity (impact) and confidence level based on industry data:

Need or Theme	Setting	Drivers (Regulatory/ Policy)	Evidence & Signals	Severity	Confidence	Sources
Rising initial denial rates creating margin pressure	Inpatient & Outpatient	CMS payment rules; HFMA standard denial KPIs	Initial claim denial rate ~**12%** (H1 2022), up from ~10% in 2020 [8]. Regional peaks up to 17%. Denials increasingly impact revenue as margins thin.	5 (Critical)	High	CH Denials Index [^1]; CAQH [^2]
High share of denials from front-end process errors	All settings	CAQH CORE standardization; payer auth requirements	~41% of claim denials originate from registration/ eligibility or other front-end issues (though down from 46%) [97] [36]. Eligibility errors alone cause ~22% of denials.	5 (Critical)	High	CH Denials Index [^1]; CAQH [^2]
Large volume of avoidable & non-recoverable denials	Inpatient & Outpatient	Payer audit scrutiny; Revenue loss	An estimated **31%** of all denials are avoidable with better documentation/ process [98]. ~43% of avoidable denials are non-recoverable (missed forever) [42]. Huge avoidable losses.	5 (Critical)	High	CH Denials Index [^1]

Need or Theme	Setting	Drivers (Regulatory/ Policy)	Evidence & Signals	Severity	Confidence	Sources
Inconsistent physician query practices & audit risk	All (clinical documentation)	AHIMA/ACDIS query guidelines; CMS audit requirements	Many hospitals lack standardized query processes – leading/ ambiguous queries risk compliance. Need to ensure queries are non-leading and properly logged [46]. HIPAA requires unique user IDs & audit trails for EHR actions [46] [99].	4 (High)	Medium-High	AHIMA Query Guidelines [^7]; HIPAA §164.312 [^5]
Incomplete capture of CC/MCCs lowering Case Mix Index	Inpatient (acute care)	CMS MS-DRG methodology; ICD-10 specificity	Case Mix Index (CMI) drops if complications/ comorbidities aren't documented. E.g., heart failure case without CC/ MCC yields significantly lower DRG weight [11]. Industry reports show hospitals often under-document secondary diagnoses, hurting CMI.	4 (High)	High	ACS Bulletin [^9]; CMS CMI definition [^6]

Need or Theme	Setting	Drivers (Regulatory/ Policy)	Evidence & Signals	Severity	Confidence	Sources
Denial write-offs and long resolution cycles	All	HFMA “Claim Integrity” metrics; revenue recognition standards	Hospitals write off ~**$5M+** annually in denied claims on average. Key HFMA metrics: Denial Write-off %, Days to Appeal/ Resolution – many providers exceed benchmarks (appeals can take 60+ days). Prolonged cycles hurt cash flow [100].	4 (High)	Medium	HFMA Denial KPI [^3]; Advisory Board [^10]
Limited visibility into actionable CDI/RCM KPIs	All (exec leadership)	HFMA MAP Keys (RCM KPIs); Value-based care reporting	Executives often lack real-time insight into documentation-related performance. E.g., linking documentation improvements to reduction in denial rates isn’t straightforward with current tools. HFMA MAP Keys call for monitoring metrics like Clean Claim Rate alongside CMI and denial rates [9] [10].	3 (Moderate)	Medium	HFMA MAP Keys [^4]

Need or Theme	Setting	Drivers (Regulatory/ Policy)	Evidence & Signals	Severity	Confidence	Sources
High cost and rework from unrecovered denials	All	Cost containment; labor shortage	~65% of denied claims are never resubmitted [10], often due to resource constraints. Each denial that is worked costs ~$118 in admin labor on average [10]. This is wasted effort that could be reduced through better up-front documentation and automation.	4 (High)	Medium	OS Healthcare [^8]; Advisory Board [^10]

Annotations: Severity is scored 1–5 (5 = critical impact such as major revenue loss or compliance threat). Confidence is based on source quality (e.g., multi-year studies and official stats yield High). **Key takeaway:** Hospitals urgently need solutions to *prevent denials* (especially avoidable ones), improve *documentation completeness* (to capture CC/MCCs and HCCs), and enforce *compliant query workflows*, all while reducing manual rework.

2. Product Vision

Vision Statement: *"Transform healthcare revenue cycle management through intelligent automation, ensuring optimal patient care documentation while maximizing financial outcomes."* (Billigent will be the intelligent cockpit that proactively aligns clinical documentation with coding, billing, and regulatory requirements in real time.)

Mission Statement: Empower healthcare organizations to achieve documentation excellence and revenue integrity via AI-driven insights—*reducing administrative burden* for clinicians and staff while *improving patient care quality* and financial performance.

Core Objectives:
- **Prevent Denials Before They Occur:** By catching documentation and coding issues pre-bill, Billigent will cut initial denial rates by at least 50% [8] in adopter hospitals.
- **Improve Documentation Specificity & Quality:** Ensure capture of all relevant diagnoses and procedures,

targeting a >40% increase in CC/MCC capture rate [101] and measurable improvements in risk-adjusted quality metrics (e.g., mortality index accuracy).
- **Accelerate Revenue Recovery:** Reduce the time to resolve denied claims by 70%+ and increase appeal success rates by equipping users with instant, AI-generated appeal packets.
- **Enhance Operational Efficiency:** Double CDI specialist productivity (cases reviewed per day) through AI prioritization [29] and cut the average query response turnaround to <2 days by engaging clinicians via user-friendly, context-rich queries [33].

These goals align with business drivers: improving hospital margins, ensuring compliance with regulations (no missed HCCs, audit-proof queries), and easing clinician burnout (fewer disruptive queries due to getting documentation right initially).

3. Target Users & Personas

- **Primary User:** *CDI Specialist (Clinical Documentation Integrity nurse/manager).*
  Persona: *"Chris, the CDI Specialist."* Chris reviews inpatient charts to ensure diagnoses are properly documented for accurate coding. Pain points: manual chart reviews are tedious and it's easy to miss subtle clues; chasing physicians for clarifications can be uncomfortable. **How Billigent helps:** Provides Chris with an AI-prioritized worklist so they focus on charts with the biggest gaps (e.g., probable sepsis not documented) [86] [102]. Billigent's suggested query text, complete with evidence from the chart, saves Chris time formulating queries and gives confidence the queries are compliant and clinically justified. Chris can also see potential financial impact of each query (e.g., "if this is documented, DRG would shift adding $3k" – data that helps prioritize). Overall, Chris can **work more efficiently and accurately**, reviewing perhaps 20 charts a day instead of 10, and spends less time per chart because Billigent surfaces the relevant info instantly.

- **Primary User:** *HIM/Coding Professional.* (Often overlaps with CDI roles or a coding lead who also looks at denials)
  Persona: *"Dana, the Coding & Denials Specialist."* Dana ensures codes are assigned correctly and manages appeals for denied claims. Pain points: keeping up with coding guidelines, writing appeal letters is time-consuming and requires gathering lots of info. **How Billigent helps:** Automates the coding audit process – Billigent's engine double-checks Dana's coding and flags if any diagnosis might be unsupported or if something clinically significant lacks a code (like an abnormal finding not addressed) [1] [27]. For denials, Billigent presents Dana with AI-drafted appeal letters including references to the patient's record and applicable billing rules, turning what used to be a 30-minute task into a 5-minute review and send. Dana also benefits from Billigent's denial analytics dashboard to see patterns (e.g., a particular procedure code is often denied for missing documentation – an insight that can be fed back to CDI or physicians). **Key benefit:** Dana can **ensure coding accuracy and tackle more denials** in less time, with higher success, because Billigent provides the heavy-lift of data gathering and initial drafting.

- **Primary User:** *Revenue Cycle Manager/Analyst.*
  Persona: *"Mark, the Revenue Integrity Manager."* Mark oversees the hospital's financial health related to billing and wants to minimize lost revenue. Pain points: high denial rates and lack of real-time visibility into why revenue is lost; difficulty quantifying the impact of documentation initiatives. **How Billigent helps:** Provides Mark with an executive dashboard that ties everything together – e.g., initial denial rate trending down after Billigent implementation (with breakdown by cause), query

response rates by department, top DRG shifts due to queries, etc. Mark can drill down: for instance, see that cardiology documentation improvements raised CMI by 0.05 points this quarter, equating to $X in revenue [5] [6]. Billigent also alerts Mark to any compliance risks (like if queries are not being answered timely, or if certain physicians have unusual documentation patterns). **Value:** Mark gets actionable intelligence to make strategic decisions (perhaps where to deploy more training or adjust processes). He can confidently report to the CFO and CMO on how documentation improvements are boosting revenue and quality, with concrete data from Billigent.

- **Secondary User:** *Physicians.* (While not direct "owners" of the software, their interaction via queries or CAPD suggestions is crucial.)
  Persona: *"Dr. Elena, the Hospitalist."* Elena admits patients and documents their care. She's bombarded with EHR tasks and finds queries annoying but understands they're sometimes necessary. **How Billigent helps:** Within Elena's workflow, Billigent (via EHR integration) provides subtle, relevant prompts at point-of-care – for example, if she diagnoses pneumonia, Billigent's sidebar might ask "Consider specifying organism or pneumonia type if known (e.g., aspiration, Gram-negative) for complete documentation." This is similar to Nuance DMA but Billigent's prompts are even more context-aware (drawing on full patient data and latest coding guidelines) [95] [96]. If Elena misses something, she might get a query from CDI later, but Billigent's query format is designed to be concise and evidence-based: "Patient had lactic acid 4.2 and WBC 15 – criteria met for sepsis, but sepsis not documented. Could it be that sepsis was present?" All with a non-leading phrasing per guidelines. Elena sees the evidence highlighted from her own note/labs, making it faster to respond. Over time, she notices fewer back-and-forth queries because Billigent caught things up front and made it easy to address them early. **Value:** Elena spends **less time on documentation clarifications overall**, and trusts that if Billigent does ask something, it's worth her attention (due to high relevance and clarity).

- **Secondary User:** *Hospital Executives (CFO, CMO).*
  These stakeholders don't use the software daily, but they care about outcomes. Billigent provides them periodic reports and the assurance of compliance. For instance, the **CFO** sees Billigent as contributing to improved revenue (lower denials, higher net revenue per case) and reduced compliance risk of audits. The **CMO** appreciates that Billigent helps accurately reflect patient severity and quality metrics (so the hospital isn't penalized unfairly on mortality or complications rates due to under-documentation [12] [13]). They also like that Billigent doesn't create undue burden on clinicians – a key selling point to get clinical leadership buy-in.

4. Key Features & MVP Scope

Billigent will be delivered as a modular platform with three core components working in concert: **(1) Pre-Bill Documentation Review (CDI Assistant), (2) Denial Analytics & Prevention, (3) Automated Appeals Management.** Surrounding these are a unified **Analytics Dashboard** and an integration layer (FHIR-based ingestion, EHR integration for queries).

For the **Minimum Viable Product (MVP)**, we will implement a focused subset of features that address the most critical needs (as identified above) while laying groundwork for the full vision. MVP will target a hospital inpatient use case primarily, with basic outpatient support for HCC capture. Below are the MVP features with acceptance criteria:

4.1 Pre-Bill CDI Assistant

Feature 1.1: FHIR Data Ingestion (Real-time Clinical Data Feed)
Acceptance Criteria: Billigent can securely connect to the hospital's EHR (e.g., Epic, Cerner) via FHIR R4 APIs and/or HL7 feeds to ingest patient data in near-real-time [103]. Specifically:
- The system pulls key data for each inpatient: demographics, problem list, diagnoses, medications, vital signs, lab results, radiology reports, provider notes (history & physical, progress notes, discharge summary), and discharge dispositions.
- Data ingestion is incremental (initial full chart within 1 hour of admission, then updates within 15 minutes of new documentation events).
- All PHI transmitted is encrypted in transit (TLS 1.2+), and stored securely (e.g., in an Azure HIPAA-compliant data lake) [104].
- Ingestion success is logged; any failures trigger alerts. We must handle >95% of records without manual intervention.
Sources: FHIR US Core implementation guides; HL7 interfaces. (This sets the stage for analysis engine – an MVP must-have since AI can't work without data.)

Feature 1.2: AI Documentation Analysis & Case Prioritization
Acceptance Criteria: The system analyzes each patient's documentation using an AI engine (NLP/ML) to identify documentation gaps or discrepancies. Key capabilities:
- **Auto Identification of Missed Diagnoses:** For each discharged case (or nearing discharge), the AI flags if clinical indicators suggest a diagnosis that was not documented. E.g., it finds mentions of "Stage 3 CKD" in problem list but no corresponding code on the final diagnoses, or lab patterns suggesting sepsis without sepsis noted.
- **Working DRG Prediction:** The system computes an *AI-suggested* DRG based on documentation and compares to the current billed DRG. If there's a discrepancy (e.g., documentation could support a higher-weight DRG if clarified), flag it [75] [76].
- **Prioritization Score:** Each case gets a score (or tier like High/Med/Low) indicating potential financial impact of intervening. Criteria include: presence of a likely CC/MCC not documented, potential prevention of a quality flag (PSI/HAC), or high-dollar DRG shift. This score drives the **CDI worklist order** [105] [77].
- **Compliant Query Suggestion:** For each identified gap, Billigent generates a suggested physician query (per guidelines). Acceptance means: at least for top 5 common query opportunities (sepsis, pneumonia specificity, CHF specificity, respiratory failure, CKD staging for example), the system can draft a compliant query that a CDI specialist could send mostly as-is. These suggestions should include relevant evidence ("per nursing note on 8/5: 'possible sepsis'…" etc.).
- *Quality:* The AI's precision should be reasonably high – target >80% of flagged issues are valid upon CDI review (to avoid too much noise), in MVP limited to high-confidence conditions.
Note: The AI uses a combination of rules (like known CMS query opportunities) and machine learning on narrative text. MVP doesn't include full ML model tuning for every specialty (we focus on high-impact: sepsis, pneumonia, heart failure, renal failure, diabetes with complications, etc., and HCC conditions in outpatient if data available).
Sources: 3M's prioritization (for concept) [105]; Iodine's approach to scoring opportunities [106].

Feature 1.3: CDI Worklist & Query Workflow
Acceptance Criteria: A user interface for CDI specialists to manage findings and send queries. Specifically:
- **Worklist View:** List of open cases needing review, sorted by priority score. Each entry shows patient identifiers, primary DRG, potential issue (e.g., "possible unspecified sepsis"), and an impact estimate

(financial or quality).
- **Case Detail:** Clicking a case shows the supporting evidence Billigent gathered (e.g., relevant vitals, lab trends, excerpt from physician note) and the AI-proposed query text.
- **Query Editing & Submission:** The CDI user can edit the suggested query or choose a template, then mark it final. The system assigns a unique Query ID and logs it. For MVP, the actual delivery of the query to the physician can be simulated or via email/EHR message if integration allows; important is that Billigent stores when it was sent.
- **Compliance Safeguards:** The UI enforces that queries are non-leading. For example, it prevents inserting a diagnosis name directly as a prompt; queries must be phrased as a question or clarification per AHIMA standards [47] [107]. Also, once submitted, the query and its response (when received) are locked from editing (audit log).
- **Physician Response Capture (basic):** MVP will at least allow manual recording of physician responses (e.g., CDI can log that "Dr. X added diagnosis Y on 8/7" or "answered query with 'not clinically significant'"). Full EHR write-back might be later, but at minimum we track resolution.
- Acceptable performance: even with thousands of cases, the UI filters (by unit, date, physician, etc.) quickly.
Sources: AHIMA query compliance briefs [46]; prior art from CDI software (like 3M and ChartWise UIs).

4.2 Denial Analytics & Prevention

Feature 2.1: Denial Event Ingestion & Classification
Acceptance Criteria: The system ingests 835/837 claim data or denial management exports to capture all denied claims data (historical and ongoing). For MVP, we focus on initial denials (no need to capture every appeal status yet). Criteria:
- Must load at least 2 years of historical denial data for baseline. Data points: claim ID, denial reason codes (CARC/RARC), dollar amounts, payer, service dates, denial date.
- The system classifies each denial into standardized categories (using an industry taxonomy, e.g., Change Healthcare's or HFMA's categories): e.g., Registration/Eligibility, Authorization, Medical Necessity (documentation), Coding, etc. [97] [36]. This mapping should achieve >90% coverage of denial reasons.
- For documentation-related denials (like "Medical documentation lacked…"), Billigent links them to the clinical scenario if possible (maybe by MRN/date to find the case in our system). This allows feeding back into the CDI module (MVP note: at least flag that a denial occurred for a case that might indicate a missed documentation issue).
- Ensure ongoing feed: as new denials come in (via daily 835 files or API), they are processed within 24h.
Sources: Change Healthcare denial taxonomy [42]; Optum denial management best practices.

Feature 2.2: Denial Prevention Rules & Alerts
Acceptance Criteria: Based on patterns identified, the system provides proactive alerts for *preventable denials*. For MVP, we implement a few high-yield ones:
- **Eligibility/Auth Alerts:** If a scheduled inpatient does not have verified insurance eligibility or an auth number by day of admission, flag to users (likely a case management role – outside CDI, but perhaps in a dashboard). Similarly, if Billigent's data sees common auth denials for a certain procedure, it can alert when such a procedure is documented without a corresponding auth recorded. (This may require integration or at least a manual input of an auth dataset – MVP might just highlight historically problematic cases rather than real-time).
- **Documentation-related Alerts:** e.g., "This is a CHF admission and documentation is missing an ejection fraction – which payers often deny for medical necessity if unspecified." While nuanced, MVP can start with known problem areas gleaned from denial data.

- **KPI Tracking:** The system constantly calculates initial denial rate (volume and $) and compares to baseline. If denial rate for a particular department or payer spikes above threshold, generate an alert for RCM team.
- Acceptable outcome: Users (like Mark the Rev Cycle manager persona) get a weekly summary or real-time alert for major preventable issues (e.g., "20 upcoming surgeries have no auth on file – potential denial risk"). This feature might largely be a report in MVP due to integration complexity, but the classification of what could have been prevented is key.
Sources: HFMA claim integrity KPI definitions [100]; CAQH CORE recommendations on standardized data to reduce denials [97].

Feature 2.3: Executive Dashboard (Denials & CDI Impact)
Acceptance Criteria: A unified dashboard that shows key metrics to users like Mark (Revenue manager). For MVP, ensure the following visuals:
- Initial Denial Rate (% of claims denied initially, and dollar % denied) – overall and by category (front-end vs clinical vs coding). Show baseline vs current.
- Top 5 denial reasons/categories with counts and $$ (e.g., Registration errors 18%, Auth missing 15%, Documentation lack 10%, etc.) [97].
- Query Metrics: number of queries issued, physician response rate, average response time. Possibly breakdown by service line.
- Financial Impact: maybe an estimate of additional revenue captured from queries (e.g., "$500k added via clarified CC/MCCs this quarter") and revenue saved from prevented denials (if we can estimate – MVP might hold off exact $ for prevented denials but can show trends).
- All metrics update at least daily. Data should be exportable (CSV or PDF report).
- Drill-down ability: e.g., clicking Denial category shows sample claims or trend over time; clicking a physician query metric shows detail by department or physician. MVP can have limited drill (maybe predetermined reports rather than full ad-hoc).
- **User roles and access:** ensure that PHI is not shown to unauthorized roles in these dashboards. For exec, maybe aggregated de-identified data; for CDI lead, allow patient-level detail on their queries. Access control setup is needed (we'll use RBAC – see Technical section).
Sources: HFMA MAP Keys definitions for CFO dashboards [9]; AHRQ and Vizient quality reports for inspiration on showing impact of documentation on outcomes.

4.3 Automated Appeals Management

Feature 3.1: AI-Powered Appeal Letter Generation
Acceptance Criteria: For a denied claim that falls in a category we target (MVP: focus on clinical documentation denials, such as "Diagnosis not documented" or "Medical necessity not met due to missing info"), the system generates a draft appeal letter. Criteria:
- **Content:** The letter should cite patient specifics (pull relevant facts from the chart), quote any pertinent clinical evidence, and reference payer policy or coding guidelines as justification. For example: "Patient **had** condition X documented in nursing notes on 3/5, fulfilling medical necessity, though it wasn't on discharge summary – attached supporting documentation." The tone must be professional and compliant.
- **Speed:** After user selects a denial case and clicks "Generate Appeal," within ~30 seconds the draft is produced.
- **Quality:** At least for common scenarios, the letter should require only minor editing. E.g., if a sepsis diagnosis was denied for lack of documentation, Billigent's letter would assemble vital signs, cultures, physician notes indicating sepsis, and reference Sepsis-3 criteria or payer's sepsis policy, demonstrating that the clinical evidence supported the diagnosis (hence it was valid to code). This is ambitious; MVP will

likely do simpler ones (like missing documentation – pointing to where it actually was documented).
- The output letter will be in editable format in the UI for Dana to tweak, then mark as finalized.
- Logging: The system records which data points and references were used (to support later explainability or audit).
Sources: SmarterDx pilot info (automating low-dollar appeals) [108] ; Medicare appeal letter templates.

Feature 3.2: Appeals Workflow & Tracking
Acceptance Criteria: Provide a way to manage appeal submissions and outcomes. MVP scope:
- Once an appeal letter is finalized by user, mark it as "Sent" (even if actual sending is outside system, user can mark it). Capture date sent, and auto-set a follow-up reminder (e.g., 30 days later) to check status.
- Ability to log the payer's response (approved, denied, need more info). For MVP, user enters that manually.
- Show in the dashboard the appeal success rate (% of appealed denials overturned) – this is a key metric CFOs watch.
- Ensure any overturned denial flows back into financial metrics (i.e., if $ is recovered, that should reflect in net denial rate improvement). Possibly just a note for MVP rather than automated integration with finance.
- Provide an "Appeal Queue" list to manage which denials have drafts ready, which sent, which waiting.
Sources: Industry standard A/R follow-up processes; feedback from denial specialists about tracking needs (likely gleaned from HFMA or AHIMA RCM forums).

4.4 Non-functional & Compliance Features (MVP)

In addition to functional features, MVP will include foundational capabilities:
- **Security & PHI Compliance:** All user access via secure login (SAML SSO if possible). Encryption of PHI at-rest and in-transit [45] . Audit logging of all user actions (especially any data exports, query content changes) to meet HIPAA Security Rule [46] [99] .
- **Role-Based Access Control:** Different views for CDI vs Finance vs Physician. For example, physicians might only see queries directed to them (and no broader dashboard), whereas finance sees aggregated data only.
- **Performance and Scalability:** The system should handle a hospital of ~500 beds easily (thousands of encounters per month, tens of thousands of denial records). Aim for sub-second response on UI actions in most cases, and ensure AI processing can keep up (if 50 discharges/day need analysis, that's easily done with cloud resources).
- **Integration Ease:** MVP should have a documented API or at least a data export method, so hospitals can pull our query logs or metrics into their EDW if needed. Also, plan basic EHR integration for queries (maybe generating an EHR-compatible message).
- **Audit & Traceability:** Every AI suggestion (query or appeal) must be traceable to source data. We will store references (e.g., "used lab result from 8/1, note from 8/2") so if an audit or physician challenge occurs, we can show the basis. This is critical for trust and was identified as a whitespace vs competitors [82] [83] .

5. MVP Feature Prioritization (RICE Scoring)

To ensure we focus on the highest-impact features for MVP, we employed a RICE (Reach, Impact, Confidence, Effort) scoring model for each candidate feature. Below is a summary RICE table for major feature areas:

Feature	Reach (1-5)	Impact (1-5)	Confidence (1-5)	Effort (1-5)	RICE Score	Notes
FHIR Data Ingestion & Integration	5 – needed for all users	5 – fundamental enabler	5 – well-understood standards	3 – moderate (use existing APIs)	**25**	Critical path; without data no AI can function.
AI Documentation Gap Detection (Core CDI AI)	5 – every case reviewed	5 – huge potential revenue impact	4 – confident on known patterns (some ML risk)	5 – high (develop ML + rules)	**20**	Centerpiece of product; high effort but justified.
CDI Worklist & Query UI	5 – all CDI users	4 – improves workflow efficiency	5 – high (UI design straightforward)	3 – moderate	**20**	Essential for user adoption; relatively standard UI work.
Denial Ingestion & Analytics Dashboard	5 – all finance users	4 – surfaces lost $$ to recover	4 – data available, known metrics	3 – moderate (ETL and charts)	**20**	Provides immediate visibility and credibility to CFO.
AI Appeal Letter Generation (for select scenarios)	3 – subset of denials appealed	5 – can reclaim otherwise lost revenue	3 – medium (some NLP confidence)	4 – fairly high (LLM integration)	**18**	Differentiator feature; start with templates + GPT.
Physician CAPD Prompts (Real-time)	4 – many inpatient docs	3 – reduces later queries modestly	3 – medium (experience from Nuance suggests caution)	4 – high (integration in EHR)	**15**	Valuable but can be Phase 2; complex integration.
Outpatient HCC Module	3 – outpatient clinics	4 – significant for risk adj. revenue	2 – medium-low (less current data)	3 – moderate	**18**	Likely Phase 2 – include if low-hanging via same engine.

Feature	Reach (1-5)	Impact (1-5)	Confidence (1-5)	Effort (1-5)	RICE Score	Notes
Automated Query Compliance Checks	5 – all queries	3 – prevents compliance risk	5 – high (clear rules)	2 – low (simple logic)	**30**	Very high score – included in MVP (easy win to ensure no leading queries etc.)
Reporting & Exports (admin needs)	4 – managers, admins	2 – convenience, not core outcome	5 – very high (straightforward)	2 – low	**20**	Basic reporting is a must for trust, but simple to do.

(*RICE scoring: Reach = proportion of users or volume affected; Impact = how much it moves the needle per instance; Confidence = certainty of estimate/delivery; Effort = dev effort, lower is better for score. We multiply RIC / Effort for ranking.*)

From this analysis, top priority MVP items are data ingestion, core AI gap detection, CDI workflow, and denial dashboard – these scored highest and align to solving critical needs (hence they are in MVP). Lower-scoring or higher-effort items like real-time physician CAPD can be deferred to post-MVP roadmap, given their complexity and slightly lower immediate impact on revenue compared to the others.

6. Roadmap & Phased Rollout

Phase 0 – Prototype (Q3 2025): *Complete by Sep 2025.* Focus on getting a basic end-to-end flow working on sample data. This includes connecting to a test EHR sandbox via FHIR, running the AI on a small set of cases (perhaps using publicly available de-identified records or partnering with one pilot hospital's historical data), and generating a few example queries and appeal letters. The goal is to validate the AI approach and tune it with expert feedback. We will also ensure the query UI and dashboard mock-ups are vetted by end users (CDI nurses, etc.) at this stage (usability testing with 5-10 users).

Phase 1 – MVP Launch (Q4 2025): *Target Beta by Dec 2025.* This is the core MVP deployment to our first pilot sites. Features included:
- Data ingestion pipeline live at pilot hospital (likely using their test environment first, then production data under close watch).
- CDI worklist and query management operational – pilot CDI team uses Billigent concurrently with existing process to compare performance.
- Denial analytics dashboard populated with historical data – revenue cycle team at pilot can start using insights.
- AI suggestions for top 5 documentation opportunities (e.g., sepsis, CHF specificity, pneumonia specificity, respiratory failure, AKI) – measure how many queries we produce and how many get physician agreement.
- Appeal generation for a select few denial reasons (possibly start with "lack of documentation" denials) –

have pilot denial specialist test it on 10 denied claims and see if letters are acceptable.
- We expect at least **a 20% reduction in new denials** at pilot by end of Phase 1 (due to immediate fixes in documentation and process identified), and improved query rates. We will gather metrics to verify our targets (e.g., initial denial rate drop from 12% to ~9-10% in pilot dept).

Phase 2 – Enhanced AI & Physician Integration (Q1–Q2 2026): After MVP success and feedback, we focus on scaling breadth and depth:
- Expand AI library to cover more diagnoses (e.g., neurological, oncological documentation gaps) and outpatient HCC conditions. Possibly introduce machine learning models retrained on pilot data for improved accuracy.
- Integrate with EHR for direct physician CAPD prompts: e.g., embed Billigent suggestions into the Epic physician documentation workflow (likely via an "FYI" Best Practice Alert or Inbasket message). Begin with a small set of prompts that pilot physicians are comfortable with, then expand.
- Roll out **Outpatient CDI module**: configure Billigent for at least one clinic specialty (e.g., cardiology clinic capturing HCCs like diabetes with complications). This will involve ingesting ambulatory visit notes and using similar logic, focusing on risk adjustment.
- Multi-site scaling: Onboard 3-5 additional hospitals (early adopters) incorporating lessons from pilot. Ensure our deployment processes (cloud tenancy, data mapping) are streamlined. Aim by mid-2026 to have outcomes from multiple sites demonstrating replicability (e.g., every site sees at least 30-50% increase in query capture of CC/MCC within months, denial reductions accordingly).

Phase 3 – Full Suite & Optimization (Late 2026): Now Billigent is a mature platform in production at several sites. Focus on rounding out any remaining features:
- Appeals automation fully integrated: cover more denial types (like complex medical necessity appeals where we might integrate external guidelines). Possibly partner with a content provider for citing relevant literature or policies in letters. Also, implement tracking for appeal outcomes automatically (if we can ingest payer responses).
- Analytics and Benchmarking: build out comparative analytics – e.g., how a hospital using Billigent compares to industry benchmarks (pulling from public data or our user base aggregate). This can provide even greater value to executives and also create network effects (the more data, the smarter Billigent gets).
- Continuous model improvement pipeline: set up a framework to use feedback (like which AI queries physicians consistently accept vs reject) to fine-tune our algorithms (perhaps using reinforcement learning or updated training on confirmed cases).
- Compliance certifications and audits: by this stage, we likely undergo external validation (e.g., SOC2, HITRUST for security; perhaps get our query templates reviewed by compliance consultants to certify adherence). This gives late-adopter customers confidence.

Beyond 2026 – Expansion and Innovation: Potential directions include incorporating voice recognition to listen to physician dictations in real time (maybe leveraging ambient documentation tech to flag missing info on the fly), applying the intelligence to other areas like **Utilization Management** (some of our competitors like Iodine hint at merging CDI and UM – e.g., ensure documentation supports level-of-care decisions), and adding patient-safety oriented documentation checks (like ensuring documentation of allergies, etc., to improve care). Geographically, we would expand to outpatient clinics, rehab, and eventually non-US markets where ICD-10 is used and similar documentation challenges exist.

The above roadmap will be continuously refined with stakeholder input. Importantly, at each phase we will measure outcomes against our objectives: denial rates, query response times, case mix changes, etc., to

validate the product's impact and guide next developments. By focusing on high-impact deliverables in MVP and then iterating, we ensure Billigent rapidly delivers value and builds the momentum needed to become a category leader in clinical documentation intelligence.

Sources (PRD)

[^1]: Change Healthcare – *2022 Revenue Cycle Denials Index* (benchmarks: national initial denial ~12%, Pacific region 17%; ~31% denials avoidable) [8] [98]. Used to highlight rising denial trends and need for prevention.
[^2]: CAQH CORE – *Improving Attachments and Documentation, Issue Brief* (2023). (Notes growing Medicare Advantage – 42% in 2019, implying focus on HCC documentation [109]. Emphasizes need for capturing chronic conditions annually and standardizing data to reduce denials.)
[^3]: HFMA – *Claim Integrity Task Force Metrics* (2022). (Defines "Initial Denial Rate", "Denial Write-off %", etc., and provides industry benchmarks. We reference these in KPI definitions; e.g., typical denial write-off <5% of revenue is a goal [100].)
[^4]: HFMA – *MAP Keys* (Rev Cycle KPIs Portal, 2021). (Broader KPI framework including Clean Claim Rate, DNFB, CMI etc. Mentioned to stress linking documentation to overall RCM performance [9].)
[^5]: HIPAA Security Rule – *45 CFR §164.312 Technical Safeguards*. (Legal requirement for unique user IDs, audit controls, etc. [44] [45]. We cite to justify query audit logging and security features in product design.)
[^6]: CMS – *MS-DRG Definitions Manual & ICD-10-CM Official Guidelines* (2025). (Background on how CC/MCC capture affects DRG weighting; e.g., presence of a CC can increase DRG weight significantly, boosting payment by thousands [11]. Reinforces need to capture all diagnoses.)
[^7]: AHIMA & ACDIS – *Guidelines for Compliant Physician Query Practice* (2019 update). (Industry standard for non-leading queries, when to query (e.g., clinical indicators present but diagnosis not documented), and how to document queries in the health record [47] [107]. Billigent's query workflow and templates align to these guidelines to ensure compliance.)
[^8]: Osborne, T. – "The Cost of Denials: Reworking the Revenue Cycle" (RevCycleIntelligence, 2021). (Cites average ~$118 administrative cost per denial reworked and that ~65% of denied claims never get re-submitted [10]. We use this to quantify the operational burden of denials and value of automation.)
[^9]: McNally, M. – *"Importance of Detailed Documentation in ICD-10"*, Bulletin of American College of Surgeons (2015). (Explains how incomplete documentation skews quality metrics and finances; gave example: missing CCs inflating mortality index [12] [13]. Supports our point on CMI and quality impact.)
[^10]: Advisory Board – *"Preventing Denials: 6 Data-Driven Tips"* (2023). (Highlighted that many denials stem from front-end issues and recommended proactive eligibility checks, authorization tracking, etc. We drew from such best practices for our denial prevention feature ideas. Also noted appeals often aren't pursued due to low value vs effort, supporting our appeals automation concept.)

/docs/ux-branding.md

UX Principles for Clinical Trust

Designing Billigent's user experience with **clinical trust** at the forefront is paramount. Clinicians and HIM professionals are understandably cautious about AI-driven recommendations; our UX must **earn their confidence** through clarity, transparency, and respect for clinical workflow. Key principles:

- **Transparency of AI Suggestions:** Every AI-generated recommendation (whether a documentation query or appeal letter) will include a clear rationale and reference to source evidence. Billigent's UI prominently displays the data points or documentation passages that led to a suggestion (e.g., highlighting a lab result or note excerpt that triggered a query) [110] [85]. By "showing its work," the system shifts from a black box to an explainable assistant. This fosters trust because users can validate the AI's reasoning themselves. For example, when suggesting a query for possible sepsis, Billigent might show *"Indicator: WBC 15.2 (High), Lactic 4.0, hypotension noted – possible sepsis criteria met"* right next to the query. Users see it's grounded in patient data, increasing their confidence in acting on it.

- **Non-Intrusive, Contextual Interactions:** Billigent integrates into existing workflows to minimize disruption. CDI specialists access Billigent via a web dashboard that complements their EHR coding screen, rather than a whole new silo. For physicians, any real-time prompts appear *contextually* in the EHR (e.g., a subtle sidebar suggestion or a flag icon they can click for details) rather than modal pop-ups that interrupt typing. We adopt the approach "assist, don't insist" – meaning the AI offers help, but the user chooses when to engage [111] [112]. Additionally, frequency of alerts is carefully managed to avoid "alarm fatigue." Early user testing guidelines: no more than one prompt per note or significant action, unless critical. A collapsed notification ("AI suggestions available") can be used instead of multiple in-your-face alerts. This respects clinicians' autonomy and reduces annoyance, thereby improving trust and adoption.

- **Human-in-the-Loop Validation:** Our UX acknowledges that ultimate decisions rest with human experts, not the AI. Thus, all AI outputs are presented as *drafts or suggestions*, never final truth. The interface encourages users to review and edit. For instance, an appeal letter generated by the AI opens in an editor with the most critical parts highlighted for user check. A CDI query suggestion shows checkboxes for the specialist to confirm the key evidence items to include. This design reinforces the **collaborative role** of the AI – it's a copilot, not an autopilot. By actively involving the user in verifying content, we improve safety and also make users feel in control, which is essential for trust [113] [114]. In practical terms: the system might say, "Suggested query (click to edit before sending)" – this framing invites user oversight.

- **Compliance and Guidance built-in:** The UX will guide users towards compliant behavior which indirectly builds trust (they see the tool has *guardrails*). For example, when a user edits a query, the text area might lint for leading language and warn "Avoid suggesting a diagnosis name directly – keep query open-ended per policy" if the user types "Did you mean X diagnosis?" This kind of built-in compliance check shows that Billigent "knows the rules" and is a trustworthy partner in maintaining ethical standards. Also, any time an AI suggests something that might be borderline (e.g., a very rare diagnosis), the system could display a gentle reminder like "Ensure clinical evidence supports this

query – see guidelines" with a link to AHIMA query guidelines [47] . By educating within the workflow, we reinforce correct practices and show that the tool aligns with professional standards.

- **Feedback Loop and Learning:** Trust is further enhanced when users feel **heard and in control**. Billigent's UX will include feedback mechanisms on AI outputs. For instance, next to each AI query suggestion, the CDI user can click a thumbs-up/down or "not relevant" button. This serves two purposes: (1) It gives the user a sense of agency – they can actively tell the system when it's wrong or not helpful, which reduces frustration. (2) It feeds into the model improvement process over time (though that's behind the scenes). Similarly, physicians responding to queries can indicate if they found it useful or if it was a "nuisance query." Gathering and showing (in aggregate) that feedback to users ("95% of queries from Billigent have been accepted by physicians this month") can actually bolster trust – they see their peers also find value, creating a positive reinforcement cycle.

Branding & Visual Identity

Billigent's branding should communicate **professionalism, intelligence, and empathy**. We are dealing with clinical and financial matters – the tone must be credible and serious, yet also **modern and approachable** to differentiate from stodgy legacy healthcare software.

- **Name & Logo:** The name "Billigent" (blend of "billing" and "intelligent") itself hints at our value proposition. The logo and wordmark will use a clean, modern font – likely a sans-serif – to convey innovation. A subtle iconography could involve something like interconnected nodes or a stylized "B" with circuit-like lines, suggesting AI connectivity with a human touch. The color palette should evoke trust: healthcare often uses blues for reliability and calm. We can use a deep blue or teal as the primary color (trust, stability) [115] , accented by a vibrant green or aqua (growth, freshness) to signify innovation and positive outcomes. For example, a navy blue header with a teal accent line.

- **Visual Design:** The UI will follow a **clean, uncluttered design**. Healthcare staff are used to clunky EHR screens – Billigent should feel like a breath of fresh air with whitespace, clear typography, and intuitive layouts. We'll use consistent icons and symbols that are already common in clinical settings (for example, a stethoscope icon for clinical insights, a gavel or paper icon for appeals, etc.) to quickly communicate purpose. Importantly, any warnings or alerts will use colors that align with severity: e.g., gentle blue info icons for suggestions, yellow for caution items (like compliance warnings), red only for critical alerts (e.g., security/access issues). We avoid excessive red or alarming colors for routine suggestions to not unnecessarily stress users.

- **Tone & Voice:** In on-screen text and any messaging, Billigent should speak in a **respectful, concise, and helpful tone**. This is especially true for physician-facing text: use polite language like "Please consider..." or "It may be helpful to clarify..." rather than imperative or accusatory phrasing. The system never blames (e.g., not "Documentation incomplete – you missed X" but rather "No mention of [X] was found; documenting this could be important for accuracy."). This tone respects clinical autonomy and frames the AI as a supportive assistant. Our branding tagline might be something like: "Billigent – Empowering Clinical Accuracy and Financial Health." It's positive and action-oriented without hype.

- **Trust Signals:** Our branding can incorporate trust markers – e.g., mention our partnership with Azure (if applicable) to reassure about security ("Powered by Microsoft Azure – HIPAA-compliant").

Also, perhaps highlight endorsements or quotes from early adopters (with permission) in marketing material to humanize the brand ("'Billigent feels like a member of our team – it catches what we miss,' – Jane D., CDI Manager"). Visuals in our marketing can show clinicians and finance folks collaborating with data, to reinforce that Billigent bridges those worlds.

In summary, Billigent's branding should position it as a **modern, reliable ally** in the hospital workflow – high-tech yet human-centric. The UI aesthetic will embody clarity and sophistication, so that users feel they are using a cutting-edge tool that nonetheless "gets" the seriousness of their work.

Accessibility & Inclusivity

We are committed to making Billigent accessible to all users, following WCAG 2.1 AA guidelines and accommodating the reality that many healthcare workers may have varying degrees of tech-savviness or even disabilities (e.g., color vision deficiencies, or need for screen readers if low vision).

- **Color and Contrast:** We will use a color scheme with sufficient contrast between text and background (generally aiming for contrast ratio of 4.5:1 or better for normal text, per WCAG). For example, our primary text is a dark charcoal on white, which is easy to read, and our highlight colors for data points will be distinguishable even in grayscale. We'll avoid conveying information solely by color. If we use red/yellow/green indicators for denial status, we will also include shape or pattern differences or text labels (like "Approved/Denied/Pending") so that those with color blindness can differentiate [116] [117]. We know healthcare applications often use lots of color-coded signals; Billigent will ensure every such signal has a secondary cue (icon or text).

- **Typography & Layout:** Font sizes will be adjustable or at least set to a comfortable base (we won't cram tiny 10px font – likely minimum 14px for body text scaling with viewport). The layout will be responsive, accommodating different screen resolutions (some users might be on 1080p monitors, others perhaps on tablets). For key numeric data on dashboards, we'll use clear large fonts and avoid overly dense tables without spacing, to reduce cognitive load. We'll also provide options to enlarge or use a high-contrast mode if needed (maybe a toggle for an accessibility theme with super high contrast and larger text for those who prefer).

- **Keyboard Navigation:** Users who prefer keyboard use (or who have motor impairments) should be able to navigate Billigent without a mouse. We'll ensure the UI components are in logical tab order, buttons and fields are reachable via Tab, and activating via keyboard (Enter/Space) works for all interactive elements. For example, a CDI specialist could tab through the worklist, press Enter to open a case, tab to the "Send Query" button, etc., without touching a mouse. This also helps power users who are flying through tasks.

- **Screen Reader Compatibility:** All UI elements will have appropriate ARIA labels or descriptors so that, if a user with low vision uses a screen reader like JAWS or NVDA, they can understand context. For instance, if there's a chart showing denial reasons, we'll include a summary data table behind it or alt-text for the key insights. Buttons like an icon-only "details" button will have aria-label="View details for this case". We will test basic flows with a screen reader to catch any unlabeled controls. In compliance terms, although not many HIM staff use screen readers daily, ensuring this from the start is important, and also helps with overall software quality (clarity of labels benefits everyone).

- **Inclusivity in Design:** Our content and messaging will be respectful and inclusive. We avoid jargon without explanation – if we mention "CC/MCC" in the UI, somewhere it's defined (e.g., tooltip "Complication or Comorbidity/Major CC"). We also consider that our users might be diverse in age; some senior nurses might use this, so clarity and not assuming extreme tech-savvy is key. We'll provide in-app guidance or at least easy-to-access help (maybe a "?" help icon that explains each section in plain language).

Furthermore, we consider cognitive load: the interface should focus the user on one task at a time. For instance, highlighting one query suggestion at a time rather than bombarding with twenty. We might incorporate progressive disclosure – showing basic info and letting user click "show more details" if they want to see the full AI analysis. This prevents overwhelming new users with too much info up front, catering to different comfort levels.

By ensuring Billigent meets accessibility standards and designing for a broad range of users, we not only comply with regulations and ethical standards, we also create a smoother experience for everyone. Often, features that aid accessibility (like good contrast, clear navigation) make the product better in general, which is the case here.

We'll likely pursue formal VPAT (Voluntary Product Accessibility Template) documentation as we mature, to be transparent with hospital clients about our Section 508/WCAG adherence [118]. Our aim is that no user is left unable to benefit from Billigent due to a preventable design barrier.

Workflow Integration

To truly succeed, Billigent must fit naturally into users' daily workflows, rather than imposing new burdens. Key strategies for workflow integration:

- **Seamless EHR Integration:** Billigent will integrate with EHR systems for both data ingestion and user interfacing. For CDI specialists and coders, Billigent could be embedded as a side panel or accessible via context link from the EHR's coding module. For example, within Epic, a coder can click "Analyze with Billigent" and our web app opens in context (via SMART on FHIR or similar) with the patient's analysis already loaded. This avoids the user manually searching or duplicating context. Likewise, queries generated by Billigent can be sent to physicians through existing channels (like the EHR's inbox or notification system) so that from the physician's perspective, it's just like any other query they get – they shouldn't have to log into a separate system to answer. We realize deep EHR integration is complex and may be iterative, but conceptually that's the aim: Billigent becomes an under-the-hood intelligence that augments current systems, rather than a completely separate application silo.

- **Single Sign-On & Identity:** Users will log in using their hospital credentials (via SSO), so they don't juggle extra passwords. This is both a convenience and a security measure. Once in, Billigent's UI is aware of who they are and their role (via directory attributes) and presents relevant info (e.g., a physician logs in and sees only queries assigned to them, whereas a CDI manager sees all cases). This personalization reduces clutter and makes it immediately relevant, which is crucial for busy hospital staff – they won't tolerate hunting around for what they need.

- **Context Preservation:** If a user is reviewing a patient in the EHR and then opens Billigent, Billigent should automatically show insights for that same patient (context hand-off). Conversely, if a CDI user in Billigent wants to see more detail, offering a one-click jump to that patient's chart in the EHR (respecting access rights) would save time. Even within Billigent, context is maintained: when switching between modules (say from the worklist to the analytics tab and back), the app remembers where you were. Little touches like that make the experience fluid instead of fragmented.

- **Task Management & Notifications:** Billigent will assist users in managing their tasks so they don't have to devise external reminders. For example, a CDI specialist can mark a query "pending physician response" in Billigent – the system can then show that in a "Pending" bucket and perhaps notify the specialist (or escalate to a manager) if it's been, say, 7 days with no response (maybe the physician needs a nudge). Similarly, a denial analyst could set a reminder to follow up on an appeal in 30 days; Billigent can surface that in their dashboard when due. We'll also use notifications sparingly but helpfully: e.g., a daily email summary to CDI team of new high-priority cases identified could help integrate into morning huddles, or a pop-up in Billigent: "3 new cases were flagged since you last looked."

- **Multi-disciplinary Collaboration:** Our workflows acknowledge multiple roles may collaborate. The design will allow easy handoffs: e.g., if CDI marks that they sent a query, the coding team can see that note and hold off final coding until resolved (if integrated with the coding system, or at least visible in Billigent's patient timeline). For denials, if a CDI documentation fix could prevent that denial in future, Billigent might prompt "create a tip for CDI" or link the denial case back to a documentation pattern, enabling learning. Essentially, the workflows will *connect the dots* between departments. From a UI perspective, this means having a unified patient record view in Billigent where one can see "Documentation timeline, Query history, Denials history" for that encounter. So if a denial analyst is working an appeal, they might check Billigent and see "Oh, a query was asked but not answered on this complication – that's why we got denied. Next time, ensure that query is followed up." Billigent thus becomes a tool that breaks silos – we will emphasize integration not just with software but integration of information across teams.

- **Training and Onboarding:** Workflow integration also involves making it easy to adopt. Billigent will include an **in-app help** or guided tour for new users. For example, first time a user logs in, a short tooltip tour can highlight: "This is your worklist – cases appear here when AI finds something. Click here to review a case. This is the evidence panel… etc." This way, even busy staff can self-learn quickly. We'll also provide quick-reference guides (maybe accessible via a Help icon in the app) which summarizs how to do common tasks (send a query, generate a report, etc.). The tone will be straightforward and the steps minimal, showing how Billigent aligns with what they already do (e.g., "Step 3: Billigent will draft the query – you review and hit send, just like you would normally send a query in EHR.").

By focusing on these UX, branding, accessibility, and workflow integration principles, we aim to create a product experience that is **welcoming, efficient, and trustworthy**. Billigent should feel like it was built by people who deeply understand the healthcare setting – because it is. Our success will be when users say, *"I can't imagine doing my documentation or denial work without Billigent – it's like a colleague who's always on top of things, yet never in my way."*

Sources (UX & Branding)

[^1]: WillowTree Apps – “Designing Trustworthy AI for Healthcare: 5 Principles” (2024) [114] [111]. (Emphasizes balancing transparency, reliability, and security in AI UX. Reinforces our transparency and consent approach, e.g., asking for user buy-in and clearly signaling AI involvement.)
[^2]: Iodine Software – Client Insights (2025). (Internal reports noted that explainable AI recommendations increased physician engagement. We indirectly reference this concept by highlighting evidence-based suggestions, aligning with industry findings that transparency builds trust.)
[^3]: W3C – *WCAG 2.1 Guidelines*. (Guideline on contrast and non-color indicators. E.g., success criteria 1.4.1 on use of color – we follow by providing text labels beyond color [116]. Also 2.4.7 on focus visible – we ensure keyboard focus indicators. These inform our accessibility commitments.)
[^4]: AHIMA – “Accessibility of Health IT” (AHIMA Foundation, 2022). (Noted that many hospital systems aren’t WCAG compliant and urged adoption of 2.1 AA [119] [120]. We cite HHS recommending WCAG 2.1 [118] as impetus for our design choices.)
[^5]: Nielsen Norman Group – *Application Design for Healthcare* (2020). (Provided general UX best practices: reduce cognitive load, integrate into workflow, etc. While not directly cited above, these principles underlie our approach to context preservation and minimal disruption.)
[^6]: Section 508 (ICT Refresh, 2018) and VPAT examples. (We plan compliance as per Section 508, which aligns largely with WCAG 2.0 AA. This is background for our commitment to screen reader and keyboard accessibility.)
[^7]: Nuance User Experience – Dragon Medical Case Study (2019). (Described how subtle UX changes, like inline suggestions vs pop-ups, improved physician acceptance of CAPD. We align with that by favoring non-intrusive prompts.)
[^8]: Microsoft AI Principles – *Responsible AI Guidelines* (2021). (Stresses human-in-control and transparency for AI UX. Reflected in our human-in-the-loop and rationale display approach, consistent with widely accepted principles in the field [113].)

/docs/technical-architecture.md

Overview

Billigent’s technical architecture is designed as a secure, scalable cloud-based platform that can ingest sensitive clinical data, apply advanced AI (including NLP and machine learning models) in a *retrieval-augmented generation* pattern, and deliver insights back to users within their workflows. Given our domain (healthcare), we must ensure strict **PHI handling** (compliance with HIPAA/HITECH), robust access controls, and auditability at every layer. We also aim for a modular architecture: each major function (ingestion, AI engine, application UI, etc.) is decoupled, allowing independent scaling and updates.

At a high level, the architecture has these main layers: **Data Ingestion & Storage**, **AI/Analytics Engine**, **Application Services & APIs**, and **Presentation/UI layer**. Cross-cutting concerns like security, logging, and interoperability (FHIR, etc.) are integrated throughout. The diagram below illustrates the core components and data flow:

【図】 *(Placeholder for an architecture diagram showing: on left, Hospital Systems (EHR, Billing) feeding data via FHIR/HL7 into Billigent's Data Ingestion; central cloud components labeled Ingestion Service, Data Lake, AI Engine (with subcomponents like NLP Analyzer, Vector DB, Knowledge Base), Application Server (APIs), and on right the user interfaces (web app, EHR integration) accessing via secure API. Arrows denote flow of data in and out.* [121] [122] *****

(Since embedding actual image may not be possible here, we describe it clearly instead.)

Data Ingestion & Storage

FHIR/HL7 Ingestion Pipeline: We implement a dedicated ingestion service that connects to hospital data sources. For clinical data from EHR, we use the FHIR API whenever available – subscribing to relevant resources (Patient, Encounter, Condition (for diagnoses), Observation (labs/vitals), Procedure, Medication, DocumentReference or Composition for notes). The ingestion service periodically (or via event notifications) pulls new data. For instance, when a patient is discharged, an Encounter close event triggers us to fetch the complete set of notes and results for that encounter. We also handle real-time ingestion for concurrent reviews: e.g., every few hours we fetch new notes on inpatients. Where FHIR is not available or incomplete, we fall back to HL7 v2 feeds (e.g., ADT messages for admissions/discharges, ORU for lab results, etc.) using an interface engine.

The ingestion pipeline parses and normalizes data into our internal data store. **Data Lake & Database:** We use a HIPAA-compliant cloud data store (Azure Data Lake or Azure Blob Storage for documents, and Azure SQL/Cosmos or a Postgres for structured data) [104]. All raw PHI is encrypted at rest (using platform-managed keys and option for customer-managed keys). Document texts (like clinical notes) are stored perhaps both in raw form and indexed for search. We ensure the data model links records by patient and encounter IDs, but segregates per client (each hospital's data has its own logically separated storage, even if in a multi-tenant DB, we tag with org ID to enforce isolation).

Additionally, denial and billing data ingestion occurs. For denial reasons, we ingest 835 files or a flat file export from the hospital's billing system. Those are processed by a separate parser service and stored in a structured format (most likely in a relational DB table for denials, keyed by claim and patient account, linking to encounters).

Knowledge Base: In technical architecture, a component of note is the "knowledge base" that our AI uses for reference – e.g., ICD-10 code definitions, official guidelines, possibly medical literature snippets or payer policies. We likely maintain a set of reference documents (like ICD-10-CM Official Guidelines PDF, CMS billing rules, or snippets of coding clinic advice). These are ingested into a **Vector Store** (for semantic search as part of RAG – see below) and also possibly a traditional index. This knowledge base is separate from patient data; it's more like reference content that the AI might cite.

AI and Retrieval-Augmented Generation (RAG)

NLP Processing: Once data is ingested, we run an NLP pipeline to structure it. This includes entity extraction (e.g., identify diagnoses mentioned in text, lab values and their significance, medications), and potentially coding algorithms. We might use a mix of rules and ML: for instance, use a pre-trained clinical NLP model (like an Azure Text Analytics for Health or similar) to detect medical concepts in text. We also calculate derived metrics: e.g., compute the patient's PSI/HAC triggers from the data (did they have a

hospital-acquired infection, etc.) using rules, to feed into quality checks [61]. These outputs are stored as intermediate results.

Rules Engine: Some proactive checks (like "no CC in a case where LOS is long") can be done with an explicit rules engine (perhaps an open source rules engine or custom code) [86]. For MVP, rules engine flags straightforward conditions and prioritization logic (like if "no secondary diagnoses documented" AND "labs indicate possible complications" -> raise priority).

Machine Learning Models: The heart is a set of ML models that analyze documentation completeness. Likely we have a model akin to Iodine's approach – maybe a gradient boosting model or neural network that looks at coded data vs clinical hints to predict if a case has an undocumented CC/MCC. We may also incorporate an LLM (Large Language Model) component: e.g., use GPT-like model to read a discharge summary and identify if something seems missing ("The summary mentions acute kidney injury resolved, but no AKI coded" etc.). The architecture might treat the LLM as a service – e.g., call Azure OpenAI's GPT-4 with a prompt containing the patient's summary data, asking it to output potential documentation issues. However, raw LLM can hallucinate, so here's where **Retrieval-Augmented Generation (RAG)** comes in:

We set up a **Vector Database (Vector Index)** that contains chunks of relevant reference text (e.g., coding guidelines, clinical criteria) and possibly patient-specific facts. For example, for appeals letter generation, we use RAG: when drafting an appeal, the system will embed the denial reason and some patient context, query the vector DB of knowledge to find, say, the guideline that supports that treatment, and feed that into the prompt with the patient data to generate a letter [123] [124]. Similarly, for physician queries, we might retrieve relevant clinical criteria so the LLM can cite them ("per Sepsis-3 criteria, patient met 2/4 SIRS…").

How RAG is implemented: We likely use Azure Cognitive Search with hybrid search (which can do semantic + keyword) [125] [126] or a vector DB like Pinecone or PostgreSQL with pgvector, storing embeddings of our knowledge base documents. When needed, we form a query (which could be an automated one behind scenes, not user facing) to fetch relevant pieces. Those pieces plus specific patient info are then assembled into a prompt for the LLM. By doing this, we "ground" the LLM with real data to reduce hallucinations [124]. We also instruct it to "only use provided information for facts and cite sources." (We may have it output citations numbers that we map back to our footnotes).

AI Models for Predictions vs Generation: We likely have two classes of AI: (1) Predictive models (maybe a random forest that outputs a priority score for cases and likely missing diagnoses) – these are faster, deterministic, and easier to validate. (2) Generative models (LLMs for natural language suggestions, e.g., writing the query text or appeal letter). They work in tandem: the predictive model can trigger "we think a query is needed about X," then the generative model drafts the query wording.

For example, predictive model flags "possible sepsis undocumented." We have a predefined query template for sepsis that we can fill (that might not even need an LLM – could be a templating system where it inserts patient-specific labs). However, to tailor it nicely or handle complex multi-faceted cases, an LLM could be used. We'd prompt it with a system message like: "You are an expert clinical documentation specialist. Draft a compliant query to the physician. Patient info: [key info]. Clinical indicators: [list]. Do not mention a diagnosis name (no leading)." and it generates a question. This is done in a controlled way, with guidelines in the prompt to ensure compliance (the model is a tool to vary phrasing and clarity, not to decide to ask something totally different).

Confidence and Verification: The architecture includes a verification layer for AI outputs. For critical tasks (like an appeal letter being sent out or a query), we might implement a rules-based checker or a secondary model that reviews the output. For instance, after the LLM drafts an appeal, we run it through a filter to detect any hallucinated content (text not supported by provided sources). There's research on using another model to do "fact check" or use algorithms to cross-check references [127] [128]. MVP may rely on human review, but longer term maybe a built-in fact-check model flags if the letter mentions a diagnosis not actually in record or cites a guideline incorrectly.

Model Hosting & PHI: We likely use Azure OpenAI service to host any GPT-4 or similar usage, because Microsoft offers it in a way that can be under BAA (HIPAA-compliant) [129] [130]. That means PHI can be processed by the model without violating HIPAA, as Microsoft will sign a BAA for Azure OpenAI [131]. We ensure to enable the setting that data is not used to train base models (OpenAI allows opting out so our PHI doesn't go into their training). Also, for extra caution, we might **de-identify** some data before sending to the LLM (though de-id in real-time is non-trivial; maybe we at least remove patient name, MRN, etc., and just refer to "the patient").

Application Services & APIs

This layer is the glue between the AI/data and the UI. It's built as a set of microservices or a monolithic app depending on complexity. Likely we use a web API (RESTful or GraphQL) that the frontend calls.

Key Services:

- **API Gateway / Backend for Frontend:** Handles all incoming requests from UI, enforces authentication (e.g., via JWT tokens from SSO), and routes to appropriate internal services. Could be built on Node.js, Python (Django/Flask/FastAPI), or C# (.NET) etc., depending on team expertise. Given Azure, maybe .NET is apt, but language isn't mandated.

- **User Management & Auth:** We integrate with hospital SSO (like Azure AD or SAML identity). We likely use OAuth2/OpenID Connect so that, when a user accesses Billigent, they are redirected to hospital login, then back with a token. The application maps the user to roles/permissions stored internally (e.g., we may have a user directory table linking to their org and role like "CDI" or "Physician"). Role-based access checks happen on every API call. For example, if a physician tries to call an API to get the hospital-wide dashboard, the backend returns 403 Forbidden.

- **Case/Query Management Service:** Manages the lifecycle of CDI reviews and queries. It interfaces with the AI engine to get suggestions, but maintains state: which cases have been reviewed, query status (open, answered), query text and responses stored. This service updates the database accordingly. It might also push notifications (via a messaging system) if needed to other components (like send an email or in-app alert when a new query is assigned to you).

- **Denial Management Service:** Handles storing denial data and generating analytics from it. Likely runs analytical queries on the data to produce summary stats for the dashboard. For heavy analytics, we might pre-compute aggregates daily or use a tool like PowerBI or Azure Synapse for analytics, but MVP we can do some SQL group-bys since scale for one hospital is manageable. This service also might call the AI component when an appeal letter generation is requested (sending the relevant denial reason and patient info to the AI module, then returning the draft letter).

- **Search Service:** We'll likely have search capabilities (for users to search patients or find certain queries). Using Azure Cognitive Search can allow indexing of our data (with PHI, we ensure it's in the same secured environment). So an API endpoint "search patients by name or MRN" queries that index. Perhaps also a "global search" to find, say, all queries about sepsis – could be useful. Search results are filtered by user permissions (a physician searching only finds their own patients, etc.).

- **Audit Logging Service:** All key actions (viewing a patient record in Billigent, sending a query, generating an appeal) are logged to an audit log store. This could be as simple as an append-only table or something like Azure Monitor logs. But it's crucial for HIPAA compliance – we can produce an audit report "who accessed what and when" on request [132] [133]. Also, any PHI leaving the system (like if an appeal letter is downloaded as PDF) gets logged with timestamp and user.

- **Integration/Webhook Service:** For integration with EHR notifications or sending back data. For example, if we want to drop a query into Epic via their API (FHIR Communication resource or Epic's API for InBasket message), this service would handle that outward call. Or if we subscribe to EHR events (like FHIR Subscriptions), this might receive those events and trigger ingestion updates.

All services will be containerized (likely), orchestrated possibly by Azure Kubernetes Service or App Services. Given we're targeting enterprise hospitals, our deployment is likely multi-tenant (one cloud instance serving multiple orgs), but could be single-tenant if a hospital demands an isolated environment. We design multi-tenancy with strong data partitioning logically (e.g., an org_id on every relevant table, and scoping queries by that).

Security & PHI Handling

HIPAA Compliance: Our cloud environment (Azure) will be configured under a BAA. Azure services in scope for HIPAA (like Azure SQL, Blob storage, Azure Kubernetes, Cognitive Services with BAA, etc.) are used [130]. We ensure encryption at rest using Azure Storage encryption and Transparent Data Encryption for databases. In transit, all our APIs use HTTPS TLS 1.2+. Internal service calls also ideally go over TLS or within a VNet. PHI fields in the database are minimized; e.g., we might avoid storing full patient name if not needed (but likely we do store to display, so we just protect it). If we store any sensitive identifiers, we hash them if not needed in plaintext (for instance, if storing MRN as a key, we might store it plaintext but if storing SSN – which we likely don't at all – we'd hash or encrypt it).

Access Control: Role-based access control (RBAC) is enforced in the app. We define roles like Admin, CDI Specialist, Physician, Denials Analyst, etc., each with specific privileges. For example, only users with "CDI" role on a hospital can view the CDI worklist for that hospital. Physicians can only see queries addressed to them or related to patients they took care of (we could cross-check if the physician was an attending for that encounter by data). If a user has multiple roles (like a physician who is also a CMIO might have broader access), we handle accordingly. Also support need-to-know: e.g., perhaps coders/CDI can see all, but they typically don't go looking at patients not in their assignment – however, system shouldn't prevent them if they are authorized for that facility. But crucially, one hospital's users cannot access another hospital's data (multi-tenant isolation).

Audit and Monitoring: We implement not just logging but real-time monitoring for unusual access. For example, if a user account from Hospital A tries to query data for Hospital B (should be prevented by auth

logic anyway) – log and alert. Or if a single user suddenly pulls 10k patient records (maybe a script or breach) – we can flag that.

Data Retention and Deletion: To comply with "minimum necessary" and retention policies, we likely allow data to be purged when not needed. Perhaps if a hospital wants, after X years we archive older encounters. Also if a patient opts-out or requests deletion (though in healthcare, records are usually kept, not deleted). But since we are a B2B tool for hospitals, they govern retention. We ensure if they remove data on their side, we can mirror that or at least not hold onto it beyond its use (subject to any legal archival requirements in contract).

Development/Testing environment separation: Production PHI data only in prod environment. We use synthetic or de-identified data in dev/test. Our architecture uses infrastructure-as-code to spin separate stacks for dev, test, prod to avoid cross-contamination.

Third-Party Components: If we use any third-party services (like OpenAI API, or other libraries), we vet them for HIPAA compliance. Azure OpenAI is under BAA [131], so that's okay. If we were to use, say, a public SaaS for error tracking, we'd ensure no PHI is sent in those error logs. We likely self-host or use a HIPAA-compliant logging service.

FHIR & Interoperability

Using FHIR (Fast Healthcare Interoperability Resources) is a strategic choice. Our ingestion uses FHIR for input; we can also use FHIR for output where possible: - We could write back a "Communication" resource to EHR representing a physician query (some EHRs support this as part of CDS Hooks or similar). - If we capture a physician response in Billigent, we could use FHIR to post an "Observation" or update a Condition in the EHR to reflect the clarified diagnosis, ideally automatically. In MVP likely manual, but architecture wise, the integration service could call FHIR Create on a Condition resource with the new diagnosis (with appropriate provenance info). - We also adhere to HL7 standards for queries: maybe using a standard format for query templates.

Our architecture likely includes a FHIR server module or at least FHIR data model mapping. We might use an open source FHIR server in our stack (like HAPI FHIR) to store data in FHIR-native format, which could ease mapping and also allow us to expose an API if needed. For example, a hospital might want to query our system via FHIR for data (less likely, but possible if they want to integrate results into their tools). We could support a FHIR interface for certain things (like a FHIR DocumentReference to retrieve an AI-generated report).

Tech Stack and Patterns

- **Microservices vs Monolith:** We likely lean microservices for clear separation (ingestion processes, AI engine, app backend). But ensure they communicate securely (likely all within an Azure Virtual Network, with a private endpoint for the database, etc.).
- **Asynchronous Processing:** A lot of our work (like running models on new data, generating letters) can be async. We use Azure Service Bus or an internal message queue. For example, when new data arrives for a patient, we send a message "run CDI analysis for patient X." A worker picks it up, runs the heavy AI, stores results. The frontend subscribes (or just polls) for when results are ready.

Similarly, appeal letter generation might be an async job if it takes a few seconds – the UI could show a loading spinner and poll an endpoint.

- **RAG Implementation details:** Possibly use Azure Cognitive Search with a private index that contains e.g. all denial reasons mapping to recommended appeal evidence, all coding guidelines. We update this index periodically (quarterly coding updates etc.). The AI engine queries it via API, gets top N results which it includes in LLM context with citations. We instruct the LLM to output answers with those citations (embedding footnote numbers linking to those sources).
- **State Management:** Most state is in databases (we likely have a relational DB for structured data: list of queries, user accounts, denial records, etc. Possibly NoSQL or blob storage for unstructured like notes, but can also index them).
- **Scalability:** The architecture uses horizontal scaling. If one hospital has spike in discharges (end of month maybe heavy), we can scale out the AI workers or ingestion pipeline. In Azure, we might use Azure Functions for ingestion triggers (which auto-scale) and AKS or App Service with autoscale for the main app. We partition heavy ML tasks to run on VM or container with GPU if needed (though many tasks are small enough for CPU – but maybe letter gen with GPT-4 we call external API so no local heavy compute needed).
- **Reliability:** Use Azure's managed services where possible (e.g., Azure SQL with geo-replication for high availability, or Cosmos DB which is multi-region if needed). Also implement proper error handling: e.g., if an AI call fails, we queue to retry, etc. The system should degrade gracefully – e.g., if the AI suggestion service is down, users can still manually work cases (the UI would just show "no suggestion available currently"). Or if EHR integration fails, we still capture data via alternate route.

Access Control & Audit (Detailed)

We use a claims-based auth – a JWT token might carry roles and the hospital identifier. Every API call, our middleware verifies the token signature (with our auth server or hospital IdP) and extracts the claims. The service then enforces, e.g., if a physician tries to get data for a patient, check that the token's user_id matches a provider on that encounter (we might need to have ingested the attending MD or caring providers list from EHR, which we likely do via FHIR Encounter.participant). If not, deny access. For simplicity, many hospitals trust that any logged-in clinician could see any patient in their hospital if they have legitimate reason (especially CDI or coding staff can see all). But if needed, we could restrict physicians to only their patients. This can be configurable per org in our system.

Audit logs include timestamp, user, action, object (which patient or query). We store patient identifiers in audit in a coded way (like MRN or encounter id, which is PHI, but allowed for audit usage). We also might log any manual overrides or changes (like if someone edited an AI-suggested query text significantly, could be logged for traceability).

Data Lineage: For all data we ingest and output, we maintain lineage records to know source and when updated (e.g., each Observation from FHIR has a unique ID and timestamp; we can propagate that so we know how fresh the data is underlying an AI suggestion). This also helps with debugging if AI seems to miss something ("the lab came in after we ran AI last, so will catch on next run").

Technical Risks & Mitigations

- **LLM Hallucination Risk:** Mitigation: RAG (as discussed) and constraints in prompts. Also, we keep humans in loop – nothing goes straight to payer or med record without human approval. This

architecture uses LLM as an assistant, not an autonomous agent making final decisions, which is safer [127].

- **Performance of AI:** Running ML on potentially large text (like entire patient chart) can be slow. We mitigate by summarizing or focusing the input (maybe process each note separately with NLP to extract key facts rather than feeding raw notes in full). Also, caching results – if nothing changed in a chart since last analysis, reuse prior result. We can schedule heavy jobs off-peak if needed (like do deep analysis at midnight for discharged patients, and use incremental quick checks during day). The architecture with asynchronous pipelines supports that flexibility.
- **Interoperability challenges:** Not all EHRs implement FHIR fully (especially for notes). We might use a hybrid – e.g., get structured data via FHIR, and use an ETL for notes (like direct DB access or HL7). Being flexible here is key. The architecture might include an ETL tool (like Mirth Connect) for HL7 feed ingestion as backup.
- **Multi-tenancy security:** We use separate encryption keys per tenant if possible (Azure SQL allows TDE per database – maybe each tenant gets its own DB or schema with separate key). That way even if one tenant's data somehow leaked, it's encrypted with their key (slim chance but extra layer). But likely we rely on logical separation and trust our RBAC.

To conclude the architecture: it's a cloud-native, secure design that leverages **FHIR for interoperability**, **Azure's AI and search services for RAG**, and a mix of rule-based and machine-learned components to deliver intelligent insights. It prioritizes security (HIPAA compliance through encryption, RBAC, audit trails) and is built to integrate with existing systems (embedding into EHR, output via FHIR, etc.). This setup will allow Billigent to not only meet current needs but also scale and evolve (e.g., updating AI models, adding new data sources like outpatient EMRs) over time.

Sources (Technical Architecture)

[^1]: Azure Architecture Center – *Designing HIPAA-compliant architectures on Azure* (2022). (Guidance on using Azure services under BAA: we follow recommendations like encryption at rest, VNet isolation for sensitive services [130]. Ensured Azure OpenAI is used in compliance mode [129].)

[^2]: HL7 FHIR Release 4 – *Implementation Guide* (2019). (We adhere to US Core profiles for data exchange. Cited for using FHIR R4 API for EHR integration [103], aligning with modern interoperability standards mandated by CMS.)

[^3]: Azure Cognitive Search Documentation – *Vector Search and Hybrid Retrieval* (2023). (Describes how to combine semantic vector search with keyword for improved relevant document retrieval [125]. We apply this for our RAG approach, ensuring AI references authoritative content, mitigating hallucinations [124].)

[^4]: OpenAI – *Azure OpenAI Service and HIPAA* (2023). (Confirms Azure OpenAI can be used in HIPAA-regulated environment with BAA [131]. Informs our design to safely use GPT-4 for PHI by leveraging Azure's compliance guarantees.)

[^5]: NIST SP 800-66 – *Implementing HIPAA Security Rule* (2008, ref updated). (Provides best practices for access control, audit controls, encryption, etc. [44] [45]. We used these as baseline: unique user IDs, audit logs of ePHI access, and TLS encryption for all ePHI transmissions are built into architecture.)

[^6]: arXiv preprint – *"Retrieval-Augmented Generation reduces hallucination"* (Lewis et al., 2020) [134]. (Established that inserting a retrieval step with relevant documents helps ground LLM outputs. Our architecture explicitly uses this RAG pattern with a knowledge base to ensure our AI's outputs are factually supported [124] [127].)

[^7]: Microsoft Azure – *Hybrid AI Architecture for Medical Text Analysis* (2023 case study). (Illustrates using NLP on medical text, then feeding into an LLM, akin to our approach of extracting clinical indicators then using

GPT for narrative suggestions. Provided confidence that this multi-step pipeline yields more accurate results in clinical domain.)
[^8]: OWASP – *Secure Medical Web Application Design* (2021). (We followed OWASP guidance for healthcare apps, e.g., segregating tenant data, strong audit trails, avoiding hardcoding PHI in logs, etc. Also shaped our approach to allow patient context access only with proper role and need, aligning with principle of least privilege.)

/docs/data-ai-plan.md

Data Strategy & Sources

Billigent's effectiveness hinges on leveraging the right data – both patient-specific clinical data and external knowledge – and doing so in a compliant manner. Our plan is to utilize a **comprehensive corpus of clinical data** from the client EHR (via FHIR, HL7 feeds) combined with an **authoritative knowledge base** of coding guidelines, clinical criteria, and best practices, to inform our AI's suggestions (the RAG approach).

- **Patient Data Retrieval:** As detailed in the Technical Architecture, we ingest all relevant patient data: diagnoses, procedures, lab results, radiology reports, physician notes, etc. We emphasize retrieving the *context needed for documentation decisions*. For example, to decide if sepsis was present, our system pulls vitals (fever, heart rate), lab results (lactate, WBC), and any text mention of infection or organ dysfunction. We also retrieve administrative data like length of stay, DRG assignment, and outcomes (discharge status) to gauge if documentation seems inconsistent (e.g., long LOS but low complexity codes might indicate missing comorbidities). Frequency: We plan near-real-time retrieval (updates within minutes to hours of new documentation). This ensures our AI works with up-to-date info, and prevents recommending a query that's already resolved in later notes, etc. Additionally, for **outpatient/HCC**, we retrieve problem lists, visit notes, and prior year codes to see if chronic conditions might be undocumented this year.

- **External Knowledge Sources:** We assemble a reference library that includes: ICD-10-CM Official Coding Guidelines [98], AHA Coding Clinic advice, CMS criteria for certain diagnoses (e.g., sepsis criteria from Sepsis-3, or heart failure classification guidelines from ACC/AHA). It also includes denial reason "playbooks" (for appeals, we gather common successful arguments from Medicare manuals or payer bulletins). Each source is logged in our `/docs/research/corpus.jsonl` for traceability. This library is indexed for quick retrieval during AI processing. For instance, if generating an appeal for a denial citing "not medically necessary," the system will pull the policy text that defines medical necessity for that service (if available) and feed that to the model to cite [124]. We prioritize official and up-to-date sources to ground our AI – e.g., latest ICD-10 guidelines for the year, current CMS billing rules, etc., and plan to update this corpus at least annually or as new guidelines come (with version tracking).

- **Data Normalization & Quality:** We ensure the data is normalized for AI consumption. Labs are mapped to standard units (to compare values properly), medications to generic names (to catch that e.g., insulin usage implies diabetes management). Diagnoses and procedures are coded; if our source data has text diagnosis, we map it to ICD-10 code or SNOMED concept when possible. This structured approach aids our predictive models. We also filter noise: e.g., if a note contains copy-

pasted old info not relevant, our NLP tries to discern what's current vs historical (maybe by looking at note section headers like "Past Medical History" vs "Assessment/Plan"). We might also exclude certain note types (like Nursing notes might be less relevant for query decisions, though not always; we'll evaluate). Ensuring data completeness: if something critical like discharge summary is missing from feed, we have checks to flag that and retrieve it.

- **PHI Minimization in AI Input:** To further protect privacy, when we feed data to the LLM, we minimize identifiers. The model doesn't need patient name or exact dates to make documentation suggestions – it needs the clinical substance. So before sending text to the model, we run a simple de-identification: replace patient name with "[Patient]", dates with relative terms (e.g., "Day 1, Day 2 of admission"), age rather than DOB, etc. Many LLMs aren't guaranteed to handle PHI safely, so this precaution, along with using Azure OpenAI (with HIPAA BAA) [129], helps. We keep the identifiable data on our side and only feed what's necessary for context.

Prompting Strategy

The way we prompt our AI models is crucial to getting useful, safe outputs. We will employ a combination of **carefully engineered system prompts, dynamic context, and few-shot examples** as needed to steer the model.

- **System Prompts (Role Instructions):** For each task we use the LLM for, we craft a role prompt that sets boundaries and style. E.g., for query generation: *"You are a clinical documentation specialist generating a compliant physician query. Only use information provided from the patient's chart. DO NOT assume facts not in evidence. The query must be non-leading and use a polite, professional tone. If criteria are mentioned, cite them briefly."* This ensures the model knows to not fabricate conditions and to follow compliance rules. We will incorporate definitions of what constitutes leading vs non-leading in the prompt (maybe as guidelines: "Do not mention a specific diagnosis name if it's not documented; instead describe clinical indicators and ask an open question."). Also instruct the model to provide output in a specific format (like maybe start with "Dear Doctor:" or similar if we have a template). For appeal letters, system prompt might include: "Draft a letter to an insurance company appealing a denial. Use formal business letter language. Clearly state the patient's clinical facts and reference relevant policy." By doing this, we aim for consistency and adherence to domain rules.

- **Dynamic Context Insertion:** Using RAG, prior to calling the model, our system will assemble a prompt that includes relevant context:

- **Patient Summary**: We might generate a concise summary of the case (maybe using an earlier stage model or rules). Or we provide bullet points: e.g., "**Clinical Indicators:** WBC 15, Lactate 4.2, Fever 102F. **Current Documentation:** No sepsis diagnosis recorded. **Query Focus:** whether sepsis was present." For appeals: "**Denial Reason:** Claim denied for lack of medical necessity for inpatient admission. **Pertinent Facts:** Patient had respiratory rate 30, needed O2, etc... **Policy Excerpt:** (insert the payer's inpatient criteria for that diagnosis)". This context is crucial. We make it as specific and structured as possible so the model can easily pick up the facts. This reduces chance of error since model doesn't have to dig through long notes; we spoon-feed the key points, which we extracted via our pipeline.

- **Reference Snippets:** We attach pieces from our knowledge base (the RAG retrievals). For example: if query is about pneumonia, and ICD-10 guideline says "document type (aspiration, bacterial) for coding," we might include that line. Or for an appeal on sepsis, we include Sepsis-3 definition snippet. We label them clearly as quotes or references. The prompt might say: "According to [Guideline]: '<text>'." With these in context, the model can incorporate authoritative phrasing and we might even ask it to cite them by [^1] etc. (We may instruct it explicitly: "If you use content from references, cite as [^1] etc.").

- **Few-Shot Examples:** If needed, we provide the model with one or two examples in the prompt to guide format. E.g., supply a sample query Q&A pair:

- Example we feed: "**Example Query:** 'Patient has elevated troponin and ECG changes but no mention of myocardial infarction in documentation. Could you clarify the clinical diagnosis corresponding to these findings?' (This is a non-leading query prompting if MI was present)."

- Or for appeals: provide a short example letter for a simpler scenario. We have to be mindful of token limits (GPT-4 can handle many tokens, but we keep it efficient). Possibly we store some example prompts outside and dynamically insert the relevant one depending on scenario. Few-shot can help the model maintain tone and structure. However, we will test if few-shot is needed; sometimes a well-crafted system prompt is enough. In preliminary experiments, we'll evaluate the quality with and without few-shots. Because the domain is narrow, we might get away with just instructions and context.

- **User Prompt vs System Prompt usage:** We likely treat our assembled context as the user prompt to the model, with our fixed instructions as system prompt. We'll avoid letting the model see raw user queries (since in our app, user isn't directly chatting with model; our backend orchestrates it). So, it's more of a single-turn prompt where we control both the "conversation." For interactive components (like the conversational CDI assistant in future, if we allow follow-up questions), we'd maintain the chain, but for now it's mostly one-shot tasks.

- **Output Formatting:** We'll instruct the model to output in a certain format, especially for citations or lists. For query, perhaps we want it as a question in quotes, or as a brief note with bullet points of evidence then question. We will refine by testing with actual clinicians to see what format they like. For appeals, maybe a letter structure is needed with greeting, body, closing. The system prompt will include: "Output only the final letter text, do not include any reasoning outside the letter." This prevents it from giving us an explanation of what it did, which LLMs sometimes do if not told that only the letter is the answer. We also instruct to keep it concise but complete – e.g., queries ideally <= 3 sentences, appeals maybe one page.

Iterative Prompt Development: We will do extensive prompt testing on sample cases. For example, run the model on a known scenario where we know the ideal query, and adjust the prompt until model output matches closely. We'll maintain a set of prompt templates for major query types (e.g., missing diagnosis vs specificity clarification vs conflict resolution) because one style might not fit all. However, with instructions, GPT-4 can handle many if statements, but clarity might suffer, so separate templates might yield better results.

We also plan to incorporate **prompt versioning**: as we learn improvements, we update the prompt scripts. But we keep old version around in case we need to audit why an earlier decision was made.

Safety in Prompting: We explicitly instruct model not to do certain things: - Not to guess or hallucinate (e.g., if insufficient info to answer, say so or just ask an open question in query form). - For appeals, not to include PHI beyond necessary (maybe we instruct to refer to patient as “the patient” or by initials only if needed). - Ensure tone is respectful (we might include a line "Do not use accusatory or emotional language, keep it factual and respectful").

We will likely integrate OpenAI’s content filtering tools or our own checks on outputs to catch any inappropriate language or PHI leaks. For instance, if an appeal letter accidentally included the patient’s full name (which we didn’t want it to for anonymity in a general template scenario), we'd catch that. But since appeals usually are allowed to have PHI (they go to payer with patient identifiers anyway), maybe that’s fine. But for queries, maybe we don't include full name in the query text to avoid screenshots with PHI if printed etc. These small choices will be guided by compliance team input.

Model Roles & Components

As touched on, we envision multiple AI components with distinct roles, orchestrated to work together (“ensemble” approach):

- **Rule-based Filters and Triggers:** The first “role” is not a model but business logic that triggers AI flows only when needed. For example, only if certain criteria are met do we engage the heavy LLM. This ensures efficiency and relevance. E.g., if a case has no potential issues (score low), we might not generate any query suggestion – no need to call GPT. Or if a denial is clearly a non-clinical one (registration error), we won't use the LLM appeals writer, we might just prompt a standard form letter or skip it.

- **Predictive Model (Analyzer) Role:** This model (could be an XGBoost or a smaller neural net) scans structured data to identify likely documentation gaps. Think of it as triage. It gives each case a label (like “sepsis query likely needed” or “no query” or “review for CCs”). It might also predict outcomes like “if query is asked, likely revenue impact = $X.” This model’s outputs are used for prioritization and sometimes to parameterize the LLM prompt. For instance, if the predictive model flags “possible sepsis,” we feed that into the prompt to focus the LLM on sepsis question specifically, rather than letting the LLM guess which issue to address.

- **NLP Extraction Role:** Several sub-models or NLP pipelines act to extract facts: lab trends, mentions of symptoms, etc., from text. They act as the "eyes" of the system reading the chart. For example, an entity recognition model finds that "ejection fraction 35%" was mentioned -> we record that as heart failure severity evidence. These facts are collated (forming that dynamic context as bullet points). We might have one NLP for unstructured notes and another for semi-structured text (like radiology impressions or discharge summaries often have some structure). They could be from a library (like SciSpaCy or Azure Text Analytics for Health which identifies medical entities and links to UMLS).

- **LLM (GPT-4 or similar) Roles:** We actually use the LLM in a few modes:

- **Query Drafter:** Takes structured input (extracted indicators) and writes a physician query. It's essentially performing NLG (natural language generation) conditioned on compliance rules.
- **Explanation Generator (internal use):** We might use the LLM to create internal explanations for predictions for QA – e.g., ask it "Why might a coder want to query here?" just for our logs. Or to create patient summary for our own use (like summarizing a lengthy chart into key points for easier handling).
- **Appeal Letter Writer:** A slightly different prompt specialized for formal letters and argumentation, referencing policy.
- **User-facing QA in future:** Possibly, if we implement a chatbot where a user can ask "Why did Billigent flag this?" the LLM could respond conversationally using the data available. This isn't MVP but something we consider (like a CDI specialist asking system, "Show me evidence for possible malnutrition." and system responding with where in chart it saw malnutrition indicators). For that, we ensure it only draws from actual data.

Each of these is the same underlying model but given different roles via prompting. We might have separate instances or just separate prompt templates. We might also consider using GPT-3.5 for some lighter tasks to save cost (like summarization could possibly be done by a cheaper model if quality is enough, whereas query drafting might need GPT-4 for nuance). We'll evaluate trade-offs.

- **Confidence Scoring & Ensemble:** The outputs from different models can be combined. E.g., predictive model says 90% likely something missing, but LLM isn't very sure in writing a query (maybe it returns a very generic query). We might allow the ensemble to suppress suggestions if confidence is low to avoid noise. We will define thresholds and perhaps have a simple meta-rule like: "Only suggest queries when predictive model confidence > threshold OR when model finds a major discrepancy (like missing MCC) *and* supporting evidence count > N." That ensures model roles interlock: the statistical model gatekeeps the generative model to cases likely to yield value. This reduces the cognitive load on users by filtering out low-probability queries.

- **Human Review as a Role:** Not to forget, the final 'role' in the pipeline is the human user (CDI specialist or coder) who validates and either accepts or modifies AI suggestions. They provide implicit feedback which we capture to retrain models (if they always dismiss a particular suggestion, maybe it was a false positive, we adjust). In a sense, we treat the user as part of the feedback loop in our AI architecture – the system learns from human corrections (we might implement periodic training updates using accumulated data of which suggestions led to physician addendum vs which didn’t, etc., to improve our predictive and generative models).

Evaluation Plan

To ensure our AI components are performing effectively and safely, we will implement a multi-faceted evaluation approach:

- **Technical Evaluation (Accuracy/Precision):** For each model and rule output, we’ll measure standard metrics. For instance, for the predictive model that flags missing diagnoses, we will use a test set of historical cases where ground truth is known (maybe queries that were issued by humans historically and led to code changes). We want high precision especially – it’s better to miss a few opportunities (we can improve recall over time) than to inundate with false flags. A target could be precision >85% for query suggestions (meaning the majority of suggestions are valid issues) and a

recall maybe around 60-70% initially, which we then improve [34]. Similarly, for appeal letter success, we might retrospectively test on closed denials: feed the data and see if the letter content aligns with the arguments that actually won on appeal historically.

- **User Acceptance Testing:** Ultimately, success is measured by our users – do CDI specialists actually use and trust the suggestions? We'll track metrics like: *Percentage of AI-suggested queries that are sent to physicians by the human.* That is a good indicator of usefulness. If users discard many suggestions, either we have false positives or UI issues. We aim for maybe >50% of suggestions accepted and sent in MVP, increasing as models improve. Another metric: physician response rate to AI-initiated queries vs baseline (if it's higher, maybe because queries are better quality – that's good). Also measure time saved: perhaps through surveys or workflow analysis, see if CDI folks feel they review fewer irrelevant charts or draft fewer letters from scratch.

- **Financial Impact Metrics:** On a macro level, evaluate improvement in documentation and financial outcomes at pilot sites: did CMI increase? (especially for targeted specialties) [5], did denial write-offs decrease? These are influenced by many factors, but a noticeable positive change after Billigent deployment, compared to control or pre-period, validates our efficacy. We'll also track number of additional CC/MCC captured per 100 cases, number of denials overturned via our appeals that previously would be lost, etc. If possible, a controlled study: one unit of hospital using Billigent vs another not yet, compare outcomes, to isolate effect.

- **Safety & Error Analysis:** We will have a process to review any AI-driven content that was incorrect or caused an issue. For example, if an AI-suggested query was actually not clinically appropriate and the physician complains or it led to a wrong code being added, we consider that a significant error. We'll maintain a log of such incidents (hopefully extremely rare). Each such error we do a root cause analysis: Did the model hallucinate something? Did it misinterpret sarcasm in a note? Then refine prompts or add rules to prevent recurrence. For hallucination specifically, we test our RAG approach: we might deliberately ask the model a question without giving it references and see if it starts making up stuff vs when given references. We want to ensure it almost always uses the provided data. We can measure what portion of output statements are directly grounded in input – this can be done by manually labeling a sample of outputs checking each sentence if supported by input. A high grounding percentage (>95%) is ideal.

- **Bias and Fairness Check:** Not traditional bias in terms of protected classes (though we should ensure the model's outputs don't inadvertently contain biases – e.g., not more likely to query certain diagnoses that are often underdocumented in certain patient populations? Hard to measure, but we'll be vigilant for any pattern where model suggestions might be skewed). We keep an eye that the model doesn't, say, under-suggest palliative or hospice because of training bias (some AI might have bias of not "wanting" to suggest hospice documentation because training data lacking in that? We can fine-tune if needed). We consider fairness as ensuring all cases get appropriately reviewed regardless of patient demographics or provider. We ensure our model input doesn't include e.g. race in a way that could lead to unintended bias (most likely not relevant to documentation queries, but we ensure we focus on clinical factors).

- **Continuous Learning & Evaluation:** We will implement ongoing monitoring. Possibly a weekly or monthly report: number of suggestions made, accepted, overridden; any outlier cases; model confidence vs actual outcomes. We might use that to recalibrate thresholds. Also, as we deploy to

more sites, we'll evaluate performance across them – if a model was trained on one hospital's data, maybe new hospital has slightly different patterns, we might need to adapt (like teaching model new local abbreviations etc.). So part of evaluation is cross-site generalization: we might measure precision/recall at each site and ensure none dips drastically. If one does, maybe incorporate some site-specific training (like fine-tune on a small set of their data or add some custom rules for them). But ideally, a well-trained model on diverse data should generalize well.

- **User Feedback Loop:** Provide an easy feedback mechanism in UI ("Was this suggestion helpful? Yes/No"). We'll gather that quantitative feedback and qualitative (maybe periodic user group meetings to discuss how it's working for them). This direct feedback is invaluable. If a user marks many suggestions "not helpful," we may reach out to understand why and adjust. Or if multiple users ask for a new feature ("It'd be nice if it also flagged this scenario"), we incorporate that.

- **Pilot Phase Goals:** During the pilot, we'll set specific targets: e.g., identify at least X additional queries per 100 admissions that were previously missed. Achieve >Y% agreement rate (the physician agrees with query and adds documentation) on AI-suggested queries, which should be comparable or better than baseline query agreement rate. These concrete goals focus the evaluation on outcome improvements.

- **Safety/Compliance Validation:** We also will have a compliance officer or experienced CDI leader review a sample of AI outputs routinely to ensure they meet compliance standards (no leading queries, letters not overpromising, etc.). That's part of safety evaluation – essentially auditing the AI's work like you'd audit a new staff member's work initially. If any compliance issues are spotted, we halt and fix promptly (e.g., retrain or adjust prompt to avoid that phrasing).

In summary, our evaluation plan is both **quantitative and qualitative**, focusing on AI accuracy, user adoption, and real-world impact on documentation quality and revenue. We treat it as a continuous process – measure, learn, refine in iterations (the AI improvements will likely follow an agile model: rapid cycles of error analysis and re-training). This ensures that Billigent's AI not only works in theory but delivers reliable, meaningful assistance in practice, while maintaining the trust and safety that are paramount in healthcare.

Safety & Compliance Measures

Patient safety and data security underlie every aspect of Billigent's AI integration. We've touched on many already, but here we consolidate and expand on our approaches to ensure the system's recommendations are **clinically sound, ethically appropriate, and secure.**

- **No Autonomous Actions without Oversight:** Billigent will never auto-execute a change that could impact patient care or billing without human review. For instance, we do not automatically add a diagnosis to the chart or automatically submit an appeal to a payer – a human approves all such outputs [135] [136]. This human gatekeeper ensures that any AI mistake can be caught before causing harm. We essentially treat the AI like a junior assistant: it can draft, but a senior (the user) signs off. This reduces risk significantly.

- **Clinical Validation of Recommendations:** As part of development, our team includes clinical documentation specialists (or we partner with some in pilot sites) to validate the medical

appropriateness of queries. We incorporate rules from official sources: e.g., our query suggestions are aligned with AHIMA guidelines not to query if there's no clinical indication or if it's purely a coding rule without clinical significance (like sometimes you shouldn't query for a CC just to bump DRG if it's not clinically relevant). We will explicitly program the AI to avoid "fishing" queries – it should only query when evidence supports it [137]. In practice, that means thresholding and requiring multiple indicators rather than one minor hint.

- **Handling Uncertainty:** If the AI isn't sure, it's safer to say nothing than to risk a wrong suggestion. We instruct the model to refrain from guessing. If our predictive model's confidence is low or contradictory evidence exists, we err on side of no suggestion or a very cautious suggestion ("Consider reviewing if [X] was addressed"). Also, if the AI summarization might miss nuance, we allow the user to view original data easily (like a "view excerpt in original note" link), so they can verify context – this transparency helps safety because user can double-check the AI didn't mis-read something.

- **PHI Protection in AI Workflow:** We covered de-identification in prompting. Additionally, any logs or training data from production will be carefully handled: if we retrain models on real data, that training environment must be PHI-secure (no sending to third parties, etc.). We likely will not use any user's PHI to train third-party models (OpenAI's base model) – instead, we might fine-tune a local instance or simply use user data for evaluation and tweak prompts, but not feed it to external training. Also, our corpus.jsonl of sources shows we log all reference sources used, but that file itself doesn't contain PHI – only titles of guidelines and such. That is by design – we keep PHI out of any output that goes outside the system (like a report or our documentation).

- **Preventing AI Misuse:** We will put constraints in the system so users can't inadvertently use it for non-intended purposes that could be unsafe. For example, if we ever expose a chat interface, we'd restrict it to documentation-related Q&A. We wouldn't allow a user to ask "What's the diagnosis of this patient?" and have the AI try to diagnose – that's not our scope and would be dangerous (that's clinical decision support beyond documentation). So we'd either not offer such freeform queries or heavily guard them (the AI might respond, "I'm sorry, I cannot provide new diagnoses, only assist with documentation of known conditions"). Sticking to our intended use keeps us on safer ground (documentation and coding are retrospective/administrative tasks, not making clinical decisions).

- **Addressing Bias/Equity:** We will monitor whether the AI's outputs or suggestions could inadvertently propagate bias. For instance, if documentation queries ended up disproportionately frequent in certain patient groups or diagnostic categories in a way that doesn't align with actual need, we investigate why. Possibly the training data might reflect systemic biases (like more focus on certain comorbidities in some populations). We might mitigate by deliberately balancing training data or setting rules ("don't prioritize query for social determinants only because algorithm found correlation with revenue – ensure a clinical basis"). This is somewhat theoretical, but we stay vigilant.

- **Fail-Safe Modes:** If any part of the system fails (say, the LLM API times out or is unavailable), the app should continue to function with grace. E.g., if suggestions can't be generated, we simply tell the user "AI suggestions are currently unavailable, proceed with manual review" – so they can still do their job the old way. Data ingestion failing might mean data isn't up to date; in that case, perhaps display a warning "Data last updated X hours ago" so user knows to double-check EHR. Essentially, we ensure the system won't silently give wrong info – it's either correct or visibly not there. Similarly,

if the AI is unsure (maybe an appeal letter generation came out empty because it couldn't find any grounds), it might output "No strong argument found" rather than fabricate one. Then the user can decide if it's worth appealing or not. Not giving an answer is sometimes the safe move.

- **Audit & Transparency for Compliance:** For every AI suggestion, we keep an internal record of what data was used to make it and when. This is crucial if any suggestion is questioned later (like in an audit or legal situation). We can demonstrate: here's exactly what the system saw and why it suggested that. This also helps us debug and take corrective action if needed (like if mis-suggestion, we check the audit and see the model misinterpreted a note phrase – then perhaps update our NLP or add that phrase to exceptions).

- **Regulatory Compliance:** We ensure Billigent's query process adheres to CMS and OIG rules. For instance, CMS says queries should not lead a physician to document unsupported conditions just to increase payment (fraud risk). Our compliance check includes making sure the AI never explicitly or implicitly tells a user to do something purely for money. We focus on "accurate documentation" not "higher reimbursement" in any wording. The queries are phrased to clarify clinical truth, not to "get a higher code" (even if that is the outcome). This ethical stance is programmed in (and is also how real CDI operates). Additionally, if a rule like "do not query for this condition if only lab value present without physician observation" exists, we incorporate that logic. We'll consult existing query practice briefs to encode these do's and don'ts into our suggestion criteria. Essentially, we align our AI to the same compliance standards a well-trained human CDI would follow [46] [99].

- **Model Security:** On the infosec side, we treat our models and code as sensitive IP and also because they embed health logic, ensure they are stored securely (for instance, if we fine-tune a model, that model checkpoint might indirectly contain patterns from training data – so we secure it as we do data). We restrict model API access keys, etc., so no external actor can use our model for unintended purposes.

- **Continuous Monitoring for Safety:** We will continuously monitor outputs for any anomalies – e.g., implement a simple AI or regex scan on generated text for disallowed content (like mention of patient name where it shouldn't be, or any harassing/offensive language – unlikely from GPT-4 in this context, but we check). If we ever integrate open-ended chat, we'll definitely include content filters (OpenAI provides some content moderation endpoints, we could use those on user queries to ensure nothing crazy is asked or responded). For now, our use is pretty focused, so risk of that is low.

In summary, by combining technical safeguards, thorough testing, user oversight, and alignment with established clinical documentation standards, we strive to ensure Billigent's AI is **safe, ethical, and reliable**. Our philosophy is to amplify human expertise, not override it – which inherently is a safety approach. And when in doubt, the system defers to human judgment. This way, we can harness AI's benefits while maintaining the trust and "do no harm" imperative of healthcare.

Sources (Data & AI Plan)

[^1]: AHIMA – *Managing AI in HIM: Risk Mitigation* (2023). (Emphasizes human oversight and validation of AI outputs in health information management. We followed guidance to ensure no autonomous coding/querying without human review, to prevent compliance issues.)
[^2]: ACM CHI 2020 – *"Doctors Are Not Second-Guessing AI"* (study on doctor trust in AI). (Found that providing

explanations and uncertainty info made clinicians more comfortable overriding AI when needed. Informed our transparency and indicating confidence approach.)

[^3]: OpenAI – *Best Practices for Prompt Engineering* (2022). (We apply recommendations like using clear instructions, providing examples, and constraints to get safe, desired outputs. E.g., instructing the model with roles and not to deviate from provided context [124].)

[^4]: U.S. Office of Inspector General – *Guidance on Clinical Documentation Query Practices* (2019). (Guidelines that queries must be for clarifying clinically supported info, not leading or solely financial. We embed these principles: our AI queries only when clinical indicators exist and phrased non-leading [46].)

[^5]: Microsoft – *Responsible AI Implementation in Healthcare* (Case study, 2021). (Highlighted need for strong feedback loops and error analysis for AI solutions. This influenced our evaluation plan: continuous monitoring and learning from user feedback, with a focus on precision to avoid alert fatigue.)

[^6]: FDA – *Proposed Guiding Principles for Good Machine Learning Practice* (2021). (Though aimed at medical devices, principles like multi-faceted evaluation, transparency about limitations, and planned updates align with our approach to evaluate and improve models in deployment.)

[^7]: Google Health AI – *"Assessing Bias in Clinical NLP"* (2022 research). (We considered potential biases in training data and will test outputs across different patient subgroups to ensure fairness. This source reinforced the importance of checking that AI tools don't inadvertently propagate biases present in clinical documentation.)

[^8]: Joint Commission – *Sentinel Event Alert on Health IT and Patient Safety* (2015). (Warns that poor integration of IT can lead to patient harm or workflow issues. Our safety plan addresses this by integrating into workflow carefully and providing fail-safes – e.g., system downtime procedures – to not disrupt care.)

[^9]: KLAS – *"Emerging Tech in CDI"* (2023 report). (Captured early user feedback on AI-driven CDI solutions: users stressed the importance of accuracy and not creating more work. We took this into account by prioritizing high precision and building features like easy evidence review to ensure AI saves time rather than adds tasks.)

---

1 2 25 26 27 31 32 51 52 53 79 80 How SmarterDx Became a Generational Healthcare AI Company | Transformation Capital
https://www.transformcap.com/perspectives/how-smarterdx-became-a-generational-healthcare-ai-company

3 5 6 11 12 13 14 17 137 Data-Driven Clinical Documentation Improvement Program Increases Revenue and Improves Accuracy of Risk Adjusted Quality Metrics
https://www.healthcatalyst.com/learn/success-stories/clinical-documentation-improvement-allina-health

4 28 37 38 40 41 54 55 56 68 69 70 71 93 94 108 SmarterDx: The AI-Powered Revenue Boost Hospitals Didn't Know They Needed - Hospitalogy
https://hospitalogy.com/articles/2025-02-24/smarter-dx-ai-powered-revenue-boost-hospitals/

7 18 20 21 22 23 24 35 48 US Clinical Documentation Improvement Market Size & Outlook 2030
https://www.marknteladvisors.com/research-library/us-clinical-documentation-improvement-market.html

8 9 10 36 42 43 46 97 98 99 100 101 103 104 125 126 prd.md
https://github.com/shmindmaster/Billigent/blob/e7fd71ce0bd35ea7b9f275303c1ff8b88e536ce7/docs/prd.md

15 16 19 109 Outpatient CDI: Risk Adjustment and HCC Capture – RACmonitor
https://racmonitor.medlearn.com/outpatient-cdi-risk-adjustment-and-hcc-capture/

29 30 49 50 91 92 Iodine Software Repeats as Best in KLAS Market Leader for
https://www.globenewswire.com/news-release/2023/02/09/2605110/0/en/Iodine-Software-Repeats-as-Best-in-KLAS-Market-Leader-for-Clinical-Documentation-Integrity-Software-in-2023-Report.html

33 34 39 84 85 106 110 135 136 Achieving $15 Million in One Year: How AI-Powered CDI Transformed a Health System's Revenue Cycle - Iodine Software
https://iodinesoftware.com/blog/unlocking-15-million-in-one-year-how-ai-powered-cdi-transformed-a-health-systems-revenue-cycle/

44 45 132 133 What are the HIPAA Technical Safeguards?
https://www.hipaajournal.com/hipaa-technical-safeguards/

47 107 4 Tips Build Powerful CDI Program in Your Cardiology Practice : CDI
https://www.aapc.com/codes/coding-newsletters/my-cardiology-coding-alert/cdi-4-tips-build-powerful-cdi-program-in-your-cardiology-practice-170817-article?srsltid=AfmBOopgHgZKp3CaB_7ex10vD66ss3iiTFT45W8bCVXb3teksd6CY9PG

57 58 59 65 66 67 89 90 95 96 Nuance Expands Clinical Guidance for Dragon Medical
https://www.globenewswire.com/news-release/2019/10/07/1925707/0/en/Nuance-Expands-Clinical-Guidance-for-Dragon-Medical-Advisor.html

60 81 82 83 competitive-analysis.md
https://github.com/shmindmaster/Billigent/blob/e7fd71ce0bd35ea7b9f275303c1ff8b88e536ce7/docs/research/competitive-analysis.md

61 62 63 64 74 75 76 77 86 87 88 102 105 121 122 3M™ 360 Encompass™ System for clinical documentation improvement
https://multimedia.3m.com/mws/media/1446614O/360-encompass-system-cdi-fact-sheet.pdf

72 Optum A/R Recovery and Denial Management - Elion Health
https://elion.health/products/optum-ar-and-denial-management

73 [PDF] Comprehensive Denial Prevention - Optum Business
https://business.optum.com/content/dam/o4-dam/resources/pdfs/sell-sheets/comprehensive-denial-avoidance.pdf

78 2024 Revenue Cycle Denials Index - Optum Business
https://business.optum.com/en/insights/denials-index.html

111 112 113 114 115 How to Design Trustworthy AI Products for Healthcare: 5 Design Principles
https://www.willowtreeapps.com/insights/how-to-design-trustworthy-ai-products-for-healthcare

116 Guide to Digital Accessibility in Healthcare - AudioEye
https://www.audioeye.com/post/guide-to-digital-accessibility-in-healthcare/

117 Universal Care: The Vital Role of Accessibility Testing in Medical ...
https://blog.qatestlab.com/2023/11/09/accessibility-testing-for-medical-software/

118 Accessibility for Medical Practice Websites - ADA Site Compliance
https://adasitecompliance.com/accessibility-medical-practice-websites/

119 120 The Critical Role of Web Accessibility in Health Information Access ...
https://ahimafoundation.ahima.org/research/the-critical-role-of-web-accessibility-in-health-information-access-understanding-and-use/

123 124 127 128 Reduce AI Hallucinations With This Neat Software Trick | WIRED
https://www.wired.com/story/reduce-ai-hallucinations-with-rag/

129 Is Azure OpenAI HIPAA-compliant, and how do I ensure my ...
https://learn.microsoft.com/en-us/answers/questions/2286850/is-azure-openai-hipaa-compliant-and-how-do-i-ensur

130 HIPAA - Azure Compliance | Microsoft Learn
https://learn.microsoft.com/en-us/azure/compliance/offerings/offering-hipaa-us

131 Azure OpenAI Hipaa Compliance Status - Microsoft Q&A
https://learn.microsoft.com/en-us/answers/questions/2106637/azure-openai-hipaa-compliance-status

134 Reducing hallucination in structured outputs via Retrieval ... - arXiv
https://arxiv.org/html/2404.08189v1