

GEMMA ANALYTICS

PROPOSAL

AI Document Processing & Automation

Prepared for

hellomed Group GmbH

2026-06-15

Valid for 7 days

EXECUTIVE SUMMARY

This proposal covers five AI document processing agents (P1–P5) plus infrastructure set-up (PO) and project closing:

- **P0:** Infrastructure set-up
- **P1:** OCR agent – classification, splitting & matching
- **P2:** SEPA & consent – upload to OS
- **P3:** Co-payment exemption – upload to OS
- **P4:** Medication plan – duplicate detection & upload to OS
- **P5:** Health insurance card missing – forward to home care

This proposal includes our recommended architecture approach and a price scope for the full project.

MILESTONES

PO: INFRASTRUCTURE SET UP

- Set up development and staging environments
- Configure CI/CD pipeline and access to all required systems
- Establish the replayable test harness for ongoing measurement

P1: OCR AGENT – CLASSIFICATION, SPLITTING & MATCHING

Discovery & Planning

- Understand and co-design the process and underlying logic, including edge case handling, patient matching mechanism, HITL process, and test harness approach
- Focused workshop to identify the optimal set-up for the HITL process (interface design, routing logic, escalation paths)
- Agree on infrastructure details for development and production – relevant for all subsequent milestones
- Clarify hellomed OS API capabilities
- Benchmark current agent: measure classification accuracy per document type, catalog failure modes

Development & Iterations

- Extend OCR capabilities as planned
- Extend and build process steps as planned, including updated HITL UI and process
- Testing, review, and follow-up iteration

Budget: 2 iteration cycles*

P2: SEPA & CONSENT – UPLOAD TO OS

Discovery & Planning

- Understand and co-design the process and underlying logic (SEPA-specific constraints, signatures, check-boxes), including edge cases, HITL process, and test harness approach
- Review form design for better AI readability

Development & Iterations

- Build extraction modules as planned
- Build OS upload and HITL UI extension for P2-specific review
- Wire into P1 pipeline
- P1 mini-iteration: improve OCR classification and patient matching for SEPA & consent documents
- Testing, review, and follow-up iteration

Budget: 2 iteration cycles + 1 P1 mini-iteration*

P3: CO-PAYMENT EXEMPTION – UPLOAD TO OS

Discovery & Planning

- Understand and co-design the process and underlying logic (document variations, status categories, mapping rules, multi-patient handling), including edge cases, HITL process, and test harness approach

Development & Iterations

- Build extraction modules, splitting, and matching mechanisms as planned
- Build OS upload and HITL UI extension for P3-specific review
- Wire into P1 pipeline
- P1 mini-iteration: improve OCR classification and patient matching for co-payment exemption documents
- Testing, review, and follow-up iteration

Budget: 2 iteration cycles + 1 P1 mini-iteration*

P4: MEDICATION PLAN – DUPLICATE DETECTION & UPLOAD TO OS

Discovery & Planning

- Understand and co-design the process and underlying logic (duplicate definition, initial deduplication behavior), including edge cases, HITL process, and test harness approach

Development & Iterations

- Build duplicate detection logic against existing plans in OS and medication plan isolation from mixed documents
- Set up confidence-based routing, including HITL UI extension
- Wire into P1 pipeline and upload to OS
- P1 mini-iteration: improve OCR classification and patient matching for medication plan documents
- Testing, review, and follow-up iteration

Budget: 2 iteration cycles + 1 P1 mini-iteration*

P5: HEALTH INSURANCE CARD MISSING – FORWARD TO HOME CARE

Discovery & Planning

- Analyze revised form layout and define process steps

Development & Iterations

- Build detection module as planned
- Build extractor and routing, including HITL UI extension
- Wire into P1 pipeline
- P1 mini-iteration: improve OCR classification and patient matching for health insurance card documents
- Testing, review, and follow-up iteration

Budget: 2 iteration cycles + 1 P1 mini-iteration*

* Target error rate and confidence thresholds are defined together after evaluating iteration 1, then applied as success criteria for iteration 2. Each document-specific milestone (P2–P5) additionally includes a dedicated P1 mini-iteration to improve OCR classification and patient matching for that specific document type. This focused, per-document approach delivers more value than attempting to improve the agent across all document types in a single pass. Additional iteration cycles beyond budget are billed at the day rate.

PROJECT CLOSING

- Comprehensive documentation of all agents, processes, and infrastructure
- Handover meeting with the hellomed team covering architecture, operations, and maintenance
- 2 follow-up Q&A sessions to address questions arising after handover

INFRASTRUCTURE: ARCHITECTURE OPTIONS

We present two architecture options for implementing the AI agents. Both are viable; they differ in flexibility, maintainability profile, and long-term extensibility.

OPTION A – N8N SELF-HOSTED

n8n instance with all agent logic built as workflows. Zoho Desk serves as the HITL interface.

Logging & Storage

- **Documents & images:** S3 bucket
- **AI requests/responses:** n8n execution history (native)
- **HITL interactions:** Zoho ticket (retrievable via API)

HITL Interface

Zoho CRM is the better fit here. A native web UI can be built on top of n8n, but the interdependencies between the two layers make maintenance difficult.

Advantages

- Familiar infrastructure and tooling
- Easy interface to inspect workflow runs and debug problems
- Low-code tool with good visual interface – potentially fewer technical skills needed for maintenance and new development

OPTION B – CUSTOM AI APPLICATION

Custom application with a Python/Django backend and Vue/Nuxt frontend, purpose-built for AI document processing.

Logging & Storage

- **Documents & images:** S3 bucket
- **AI requests/responses:** relational database with structured event log
- **HITL interactions:** relational database with structured event log

HITL Interface

Native web UI built in Vue/Nuxt. Offers full flexibility for process-specific views and UI design. Can include Zoho ticket creation with a link that stays connected to the UI and is updated accordingly.

Advantages

- Full audit trail and logging per document and per patient – queryable at any time
- Native support for the orchestrator pattern: enables a coordinator agent that decides at runtime which sub-agent to trigger, which tools to call, and how to handle ambiguous inputs
- More flexibility for programmatic evaluation and testing
- Strong AI ecosystem in Python, including open-source frameworks and libraries

- Native web application with full flexibility for different process approaches and integrations

COLLABORATION

PROJECT TEAM

- **David Bader** (Project Lead) – Managing Principal
- **Bijan Soltani** – Founder and Managing Director
- **Elena Alfonsi** – Analytics and AI Lead
- **Lui Pillmann** – Tech Lead

COMMUNICATION & AVAILABILITY

- Budget updates are provided at your preferred frequency
- Primary written communication language is German
- Team vacation plans are communicated before the project begins
- Unplanned absences are mitigated by having multiple senior team members who can take over at short notice

START DATE

Planned project start: **2026-06-29**

INVESTMENT SUMMARY

37 days × EUR 1,200/day	EUR 44,400
Volume discount (5%)	- EUR 2,220
<hr/>	
Total investment	EUR 42,180

PAYMENT TERMS

- 40% after 6 weeks
- Remaining 60% upon project completion

All prices net (excl. VAT). Fixed-price items are billed upon completion. Time & materials items are billed based on actual days worked. Travel expenses, if any, are billed at cost.

ABOUT GEMMA ANALYTICS

Gemma Analytics is a Berlin-based data and AI consultancy specializing in the design and implementation of modern data platforms and AI systems. With 70+ completed projects, our team brings deep, hands-on experience across the full modern data and AI stack. We operate code-first and leverage AI-assisted tooling to deliver at speed without compromising on quality or maintainability. Over these projects we have built extensive internal documentation, reusable patterns, and engineering best practices that carry over to every new engagement – accelerating development significantly from day one.

GEMMA ANALYTICS

Gemma Analytics GmbH – Chausseestraße 17, 10115 Berlin
gemmaanalytics.com